Dalla corda tesa al machine learning

un punto di vista sul curricolo di matematica
nella scuola secondaria

Gabriele Anzellotti
professore senior

Laboratorio DiCoMat
Dipartimento di Matematica - Università di Trento

Indice

Sommario.

Vorrei presentare e discutere un possibile punto di vista sugli obiettivi di apprendimento per la scuola media, i licei e gli istituti tecnici, insieme a diversi esempi di attività didattiche.

Nel discorso vorrei in particolare far emergere

i fili che portano dai concetti più elementari della matematica, rappresentati dalla corda tesa con cui si costruiscono le rette e i cerchi degli assiomi di Euclide, al machine learning, che è diventato negli ultimi dieci anni uno strumento potente e pervasivo per descrivere il mondo, prevedere i fenomeni e prendere decisioni.

1. introduzione

Vorrei parlare con voi del **curricolo di matematica per la scuola secondaria** di primo e di secondo grado (scuole medie, licei, istituti tecnici)

- obiettivi (concetti, linguaggi, tecniche di calcolo, prestazioni)
- attività, materiali e strumenti di lavoro per gli studenti
- collegamenti, percorsi e intrecci
- ambiente e relazioni per l'apprendimento
- un sistema di valutazione

---

il curricolo non è solo un elenco di argomenti!

il curricolo viene stabilito dagli istituti scolastici e quindi dai docenti, con decisioni collegiali e individuali, nell'ambito delle **indicazioni nazionali**

vorrei oggi presentare un **punto di vista che penso sia utile nel prendere tali decisioni**

---

non ritengo che questo punto di vista sia l'unico possibile, né pretendo che sia il migliore, ma credo che valga la pena di considerarlo e vorrei discuterlo con voi

Osservo che il presente documento è molto lungo e non è inteso per essere letto interamente nel seminario. Sarà la base del discorso, ma molte parti non saranno presentate. Spero però che il testo, che contiene diverse osservazioni, dettagli e riferimenti, sia utile per chi eventualmente vorrà guardarlo, anche senza aver partecipato al seminario.

2. un punto di vista su modelli matematici e curricolo

sintesi

- perché si studia la matematica a scuola?
- abbiamo teorie *antiche* di grande successo
- modelli basati su equaz. differenziali a derivate parziali
- soluzioni approssimate e metodi computazionali
- il Machine Learning e le Reti Neurali Artificiali
- esempi di applicazioni
- le idee e i concetti di base sono semplici
- il mio punto di vista

perché si studia la matematica a scuola?

la matematica costruisce **teorie, sistemi di simboli, linguaggi, modelli**, che si usano per rappresentare il mondo, per fare calcoli e previsioni, e per prendere decisioni

l'apprendimento della matematica richiede e stimola lo **sviluppo di strumenti e di capacità basilari del pensiero**

la matematica è **bella**

questi aspetti si fondono **sempre** uno con l'altro e concorrono a rendere **significativo** l'apprendimento;

non sono in contrasto uno con l'altro!

abbiamo teorie *antiche* di grande successo
che si studiano a scuola per le ragioni suddette

geometria euclidea (300 a.C. – 100 d.C.)

algebra (equazioni) e geometria analitica (XVI e XVII sec.)

calcolo infinitesimale (equaz. differenziali) (XVII sec. –   )

calcolo delle probabilità (XVII sec. –   )

sono una parte importante del curricolo
di tutte le scuole e di molti corsi di laurea

con le teorie antiche abbiamo costruito

modelli basati su equazioni differenziali a derivate parziali

con i quali formuliamo le leggi della fisica
e descriviamo il comportamento di dispositivi
e sistemi complessi naturali e artificiali

quasi mai ci sono formule che esprimono le funzioni che risolvono le equazioni differenziali

(ma succede anche per le equazioni algebriche di quinto grado o per semplici equazioni come $2^x = 3x$)

si cerca però di mostrare che **le soluzioni esistono**

e di **trovare successioni che convergono alle soluzioni**

per le equazioni differenziali abbiamo quindi costruito

metodi computazionali

[A. Quarteroni - Enciclopedia Treccani]

che consentono di affrontare problemi molto complessi riducendoli alla soluzione di problemi concettualmente semplici: **sistemi di equazioni algebriche lineari di "molte" equazioni con "molte" incognite**. I sistemi lineari sono concettualmente semplici, ma "molte" vuol dire dalle migliaia ai milioni e occorre avere algoritmi efficienti e strumenti potenti di calcolo numerico.

ci sono **importanti problemi per i quali i modelli differenziali sono troppo complicati, oppure non esistono!**

Un sistema di tecniche recenti (esplose negli ultimi 20 anni)

il Machine Learning e le Reti Neurali Artificiali

che non si basa su modelli differenziali, ma su

algoritmi di calcolo numerico, potenza di calcolo dei processori, **memoria, architetture parallele dei computer**

produce **soluzioni a problemi in ogni campo**

ad esempio

- riconoscimento di immagini: facce, scene, testi manoscritti
- guida automatica di veicoli: automobili, veicoli spaziali
- controllo di impianti e dispositivi: centrali nucleari, lavatrici
- generazione di testi e immagini: chatGPT, https://openai.com/dall-e-2/, https://thispersondoesnotexist.com/
- processi decisionali per le imprese e le amministrazioni
- diagnosi, terapie personalizzate, sviluppo di farmaci
- motori di ricerca, marketing

Le applicazioni

del **Machine Learning** (**ML**)

e delle **Artificial Neural Networks (ANN)**

possono essere complicate

ma **le idee essenziali sono semplici**

Quali sono i concetti e i linguaggi di base per ML e ANN?

li raggruppo in temi:

- ➔ **funzioni** di più variabili reali, derivate e gradiente
- ➔ **geometria analitica** in N dimensioni: sistemi lineari, vettori, matrici, spazi vettoriali, operatori, trasformazioni
- ➔ **linguaggio per i fenomeni aleatori** (probabilità)
- ➔ **algoritmi per il calcolo numerico**
- ➔ **linguaggi di programmazione** (Python, Matlab)

il mio punto di vista:

è possibile e opportuno cominciare già presto a scuola (e all'università) **a vedere questi concetti intessuti nelle "solite" competenze matematiche e informatiche di base**

---

ciò si può fare considerando opportuni **problemi ed esempi,** proponendo opportune **attività per gli studenti**, e **ripensando gli argomenti "classici"**

---

in questo modo si avrebbe anche **un apprendimento più significativo e più solido delle competenze di base "usuali"**

3. verso il machine learning

sintesi

un Gradus ad Parnassum di modelli, dalla proporzionalità alle reti neurali

Quanto costano le mele

Quanto costa una cena al ristorante

Prevedere il successo all’università

i classificatori lineari in più variabili e il problema del riconoscimento dei caratteri manoscritti

Cos'è il machine learning?

è un sistema di tecniche per **imparare una funzione**

le tecniche si attuano mediante **macchine**
(computer, algoritmi, codici)

noi consideriamo il caso in cui sono disponibili degli
esempi da cui imparare (*supervised machine learning*)

è chiaro che da questa slide non è chiaro cos'è il machine learning, ma diventerà chiaro!

Quanto costano le mele?

Sotto casa mia c'è un fruttivendolo che non espone i prezzi della merce in vendita. Dovrebbe farlo, ma non mi va di dirglielo e neanche di chiedere i prezzi.

Tre giorni fa ho comprato due mele e ho speso **85 centesimi**. A casa le ho pesate con la mia bilancia digitale (che dà pesi multipli di dieci grammi) e il peso era 0.37 kg.

Allora ho pensato che il prezzo al chilo delle mele, in euro, doveva essere **0,80 diviso 0,37**. Ma la divisione dà come risultato **2,1622…** che non è un prezzo possibile. Ho pensato allora che il prezzo doveva essere **2,2 euro al kg**.

Un’ora dopo ho comprato altre tre mele, che pesavano **550 grammi**. Il fruttivendolo mi ha salutato cordialmente. Avrei dovuto spendere **550 x 2,2 = 1,21 euro**, ma ho pagato invece **1 euro e 25 centesimi**.

Tutto sommato, la differenza non mi ha sorpreso, perché le monetine da un centesimo non si usano più e so che le bilance non sono precisissime. Ma sono rimasto un po’ curioso di quale sia il prezzo delle mele.

Così sono tornato al negozio e ho comprato altre 4 mele. Il fruttivendolo mi ha guardato incuriosito. Le ho pagate **1 euro e 50** e le ho poi pesate a casa: **720** grammi; e questo corrisponde a un prezzo al chilo di **1,5 diviso 0,72** che è uguale a **2,083…**

Riassumiamo in una tabella i risultati dei tre acquisti

peso in kg	costo in euro	posso pensare che
0,37	0,80	**la tabella rappresenta una funzione, che chiamo** f
0,55	1,25	
0,72	1,50	

ad ogni valore della variabile **peso** la funzione associa il valore della variabile **costo**. In simboli si scrive

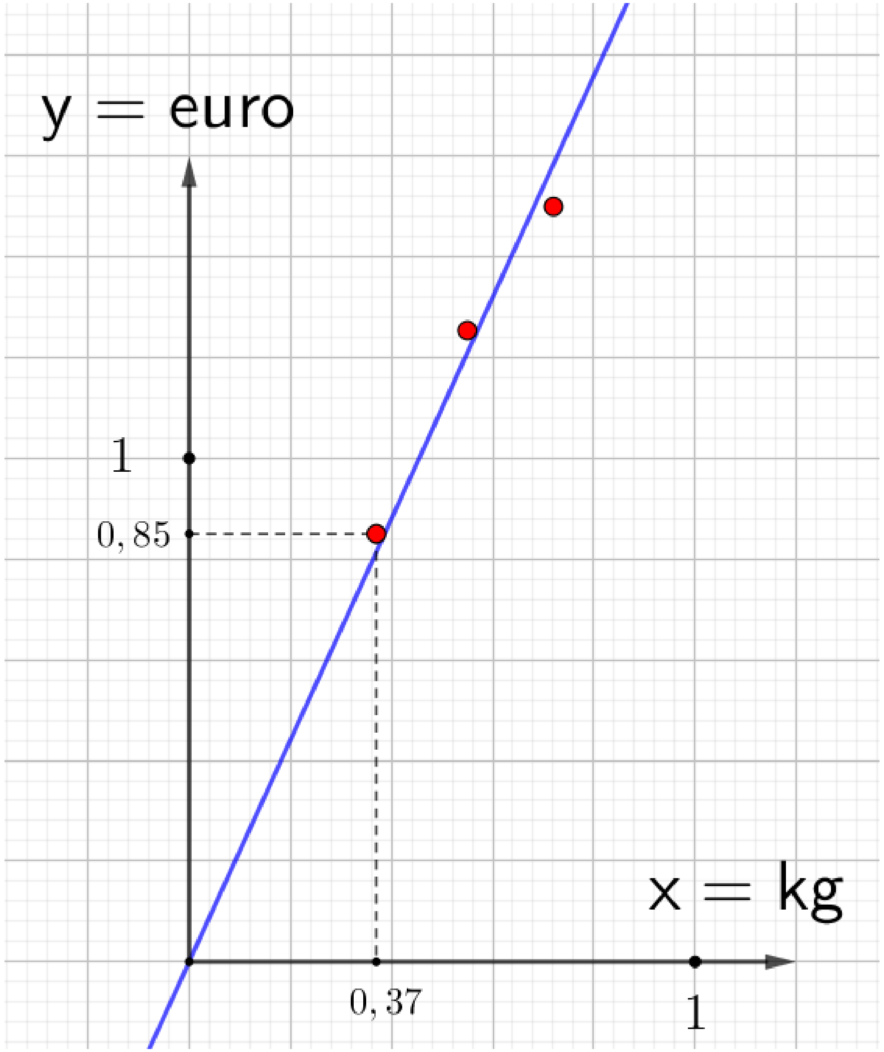
$$costo \ = \ f(peso)$$

e mi aspetto che valga la formula

$$f(peso) = k \cdot peso$$

dove k è il prezzo al kg, che non conosco.

Posso fare un grafico di questa situazione (uso GeoGebra)



ogni **puntino rosso** rappresenta un acquisto di mele:
la coordinata **x è il peso in kg**;
la coordinata **y è il costo in euro**;

l'equazione della **retta blu** è

$y = kx$ dove ho preso $k = 2,2$

La retta passa vicino a ogni punto rosso, ma non esattamente per nessuno.

La funzione $f : x \to kx$ è un *modello* per il costo delle mele

e il parametro k rappresenta il prezzo al kg

L'idea del machine learning è proprio questa che abbiamo visto:

- ➔ voglio **trovare una funzione** da usare come **modello** di un fenomeno

  nel nostro caso: come dipende il costo dal peso

- ➔ decido **una classe di funzioni** tra cui cercare

  nel nostro caso: le funzioni del tipo $y = kx$ che sono determinate da **un** solo **parametro** k

- ➔ cerco in questa classe la funzione che **meglio sta vicina ai dati** che ho

- ➔ guardo come questa funzione **prevede anche altri nuovi dati** ed eventualmente cerco di trovarne una migliore

I casi di interesse sono più complessi; si usano classi di funzioni più ricche: le **Reti Neurali**, che vedremo più avanti. La rete che usa Chat GPT-3 ha **175 miliardi di parametri** e bisogna trovare i modi per maneggiare queste reti.

Torniamo alle mele: come facciamo a stabilire qual è il valore "giusto" del prezzo k al chilo?

Nella figura di sopra abbiamo preso k=2,20, ma è stata una scelta "a occhio".

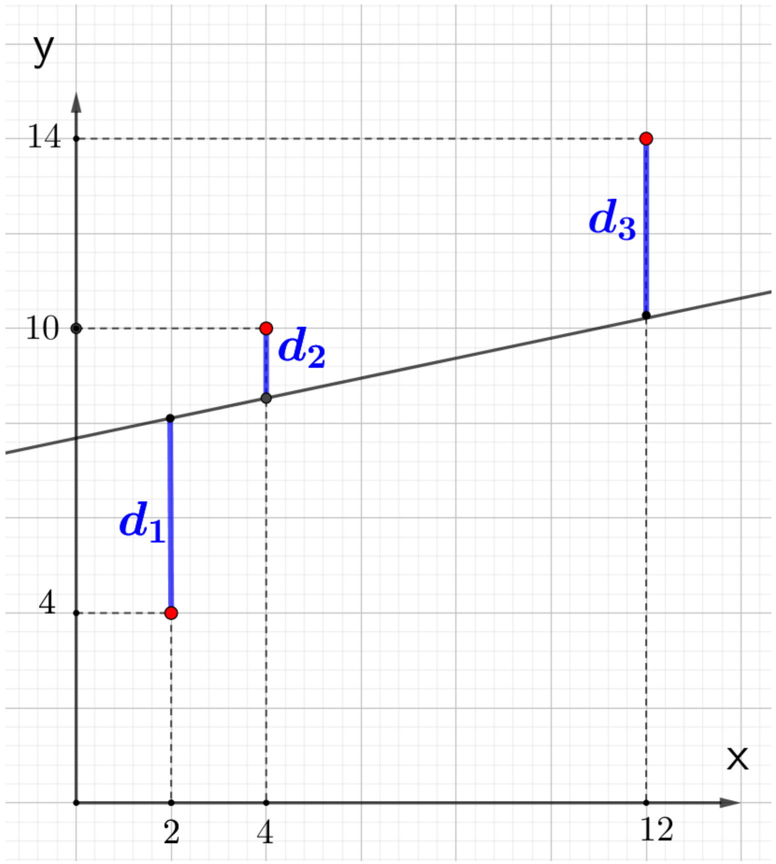
Possiamo fare la media dei prezzi nei tre acquisti

peso in kg	**costo in euro**	**costo in euro/ peso in kg**
0,80	0,37	2,162
1,25	0,55	2,273
1,50	0,72	2,083

e otteniamo 2,165, che, arrotondato, è proprio 2,20, ma in generale non è detto che fare la media dei prezzi sia una buona idea. Un modo migliore è quello di cercare la retta che nel complesso passa più vicina di tutte ai tre punti dati.

Vediamo un problema **simile**. Prendiamo i punti $P_1 = (2, 4)$,

$P_2 = (4, 10)$; $P_3 = (12, 14)$ e una retta generica r.



Indichiamo con d_1 la distanza, in verticale, tra il punto P_1 e la retta r.

Facciamo la cosa analoga per d_2 e d_3

Ciascuno dei numeri

$$d_1 + d_2 + d_3 \quad \text{e} \quad d_1^2 + d_2^2 + d_3^2$$

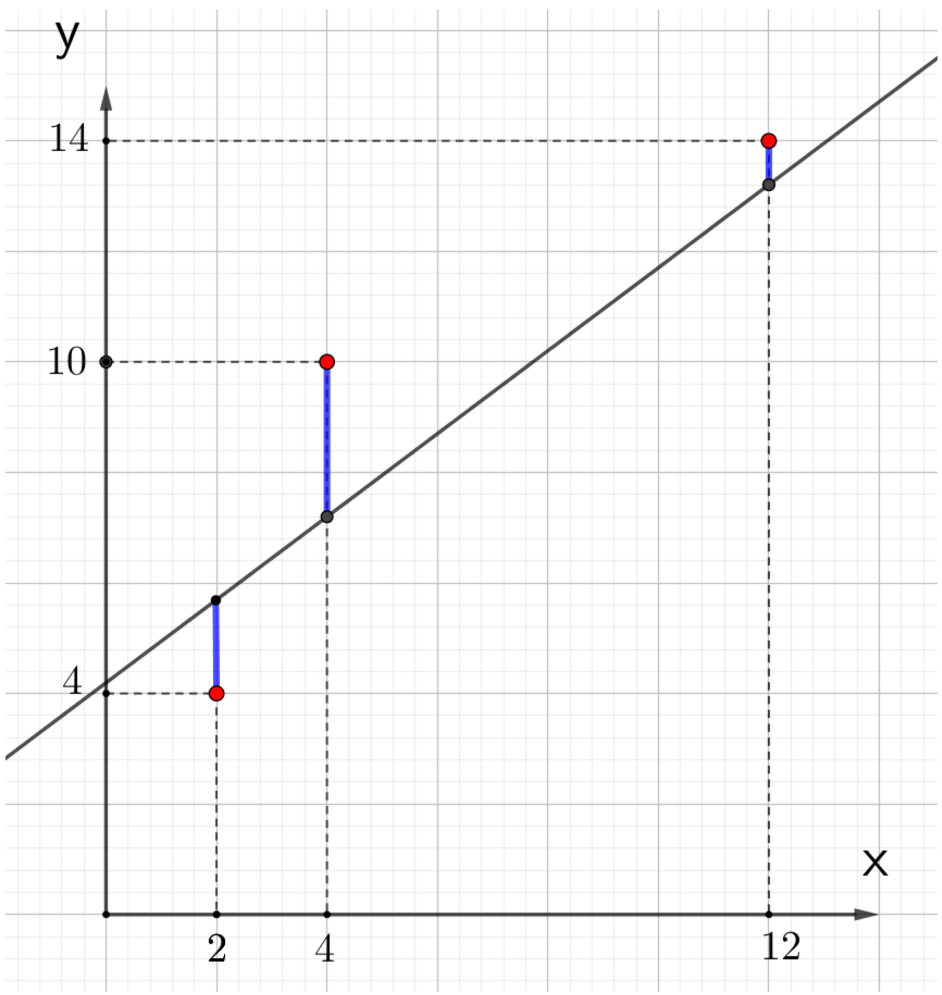
è una misura di quanto la retta è complessivamente distante dai tre punti. Hanno significati diversi. Ce ne sono molte altre possibili.
Noi useremo la somma dei quadrati.

Ci chiediamo: **qual è la retta che rende minima la somma dei quadrati delle distanze (in verticale) dai tre punti dati?**
La risposta si può calcolare a mano oppure si può chiedere a qualche software (ce sono migliaia).
Ad esempio, **GeoGebra**, nella vista **foglio di calcolo**, ha il comando **Regressione**, che ha l'opzione "**lineare**".
Con questo comando otteniamo la retta di equazione

$$y = 0.751x + 4.195$$

(i coefficienti sono arrotondati)

Questa si chiama **retta di regressione** calcolata col **metodo dei minimi quadrati.**

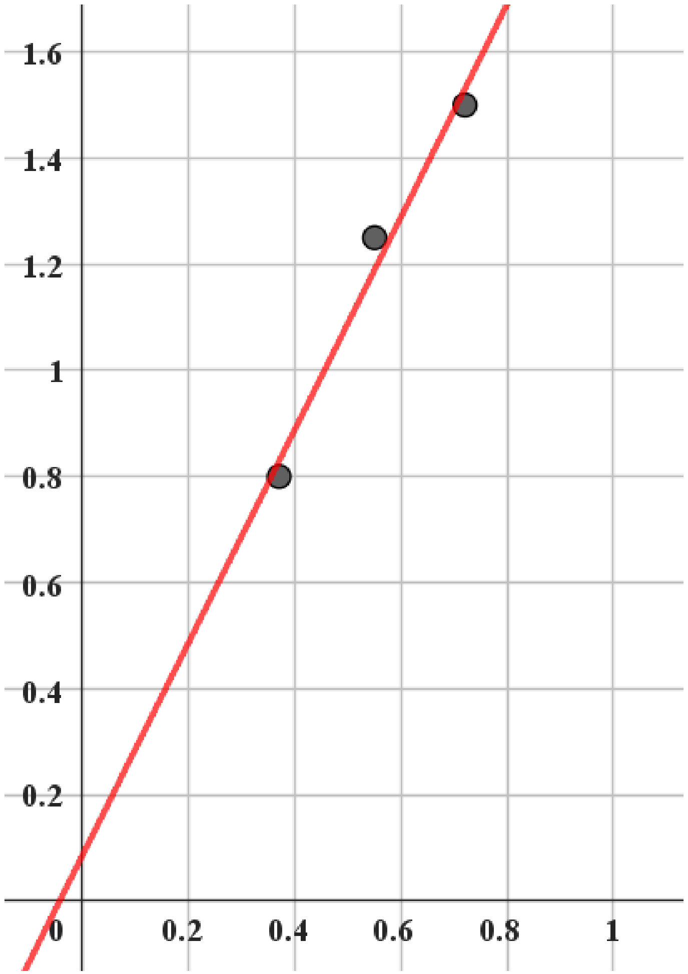


Nella figura si vedono i punti $P_1 = (2, 4)$, $P_2 = (4, 10)$;

$P_3 = (12, 14)$

e si vede la **retta di regressione** calcolata col metodo dei minimi quadrati, ossia quella che rende minima la somma dei quadrati delle lunghezze dei segmenti di colore blu.

Torniamo al modello per il costo delle mele



La retta di regressione che approssima il costo delle mele in funzione del peso è

y= 2,005 x + 0,087

Come si interpreta?

Il prezzo per chilo potrebbe essere 2 euro (cioè circa la pendenza della retta), ma perché la retta non passa per l'origine degli assi? Cosa potrei dire al mio fruttivendolo?

Come si calcolano i coefficienti della retta di regressione?

Vediamo un caso molto semplice, ma che contiene le idee. **Cerchiamo solo tra le rette che passano per l'origine**, ossia cerchiamo un modello del tipo

$$f(x) = kx$$

che si chiama di "proporzionalità diretta".

La distanza "in verticale" tra la retta e il punto $P_i = (x_i, y_i)$ è

$$|y_i - f(x_i)| = |y_i - kx_i|$$

e la quantità da minimizzare è la somma dei quadrati

$$L = \sum_i (y_i - kx_i)^2$$

che è una funzione del parametro k e dei dati (x_i, y_i).

Pensando che i dati sono fissati, L è funzione solo di k;
L è un polinomio di secondo grado in k e il minimo si trova completando il quadrato, ma per trovare il punto di minimo di L si può calcolare la sua derivata

$$\frac{dL}{dk} = -2\sum_i (y_i - kx_i)x_i$$

e la si pone uguale a zero. Così si ottiene subito

$$\sum_i (y_i x_i - kx_i^2) = 0 \qquad \text{e} \qquad k = \frac{\sum_i x_i y_i}{\sum_i x_i^2}$$

In modo analogo si possono trattare tutti i modelli lineari nei parametri, ma le formule diventano rapidamente assai complicate da scrivere e difficili da interpretare. C'è però una procedura molto più comoda, efficace e significativa che usa l'algebra lineare e l'interpretazione geometrica degli operatori lineari. La vedremo più avanti.

Secondo passo verso il machine learning

Quanto costa una cena al ristorante?

Vado spesso a cena con gli amici in un piccolo ristorante, che non ha un menù e quindi non sappiamo quanto costano i piatti.
Al termine della cena ci divertiamo a prevedere quanto sarà il conto, ma non indoviniamo mai. Se ci fosse il menù sarebbe facile sapere quanto dovremo pagare. Basterebbe

- numerare gli n piatti che si possono ordinare,
- chiamare w_j il prezzo del piatto j-esimo
- e fare la somma $x_1w_1 + x_2w_2 + \ldots x_jw_j + \ldots x_nw_n$
- dove x_j è il numero di ordini del piatto j-esimo.

Dunque il costo della cena è dato dalla funzione

$$f : (X, W) \mapsto x_1 w_1 + \cdots + x_n w_n$$

$$f : \mathbf{R}^n \to \mathbf{R}$$

dove abbiamo indicato brevemente con X e W i vettori

$$X = (x_1, \ldots x_j, \ldots x_n) \qquad W = (w_1, \ldots w_j, \ldots w_n)$$

La funzione f ha $2n$ variabili. Se abbiamo il menù, allora le variabili w_j sono fissate (diciamo che sono dei parametri) e restano le variabili x_j.

Ogni volta che andiamo a cena **conosciamo**

- il vettore X dei numero di ordinazioni per ogni piatto
- il costo della cena $y = f(X, W) = x_1 w_1 + \cdots + x_n w_n$

e questa è **un'equazione che lega tra loro i parametri** $W = w_1, \ldots w_n$, **che non conosciamo**

L'equazione è lineare nei parametri incogniti.

Se andiamo a cena un'altra volta, otteniamo ancora un'equazione.

Se andiamo a cena q volte otteniamo nel complesso q equazioni che devono essere verificate dai parametri, ossia **un sistema lineare di q equazioni in n incognite**. Se riusciamo a risolverlo, troviamo i parametri, ossia i prezzi dei piatti.

Quante volte dovremo andare a cena, prima di essere in grado di risolvere il sistema?
è una domanda alla quale si possono dare significative risposte e su cui vale la pena pensare

E se mi accontentassi di riuscire a **prevedere il costo della cena con una certa approssimazione?** Forse non è necessario trovare tutti i parametri e quindi forse bastano meno cene… **questa domanda porta anch'essa al metodo dei minimi quadrati.**

Terzo passo verso il machine learning

Prevedere il successo all'università

Consideriamo gli studenti che si iscrivono a un certo corso di laurea. Numeriamo gli studenti con un indice i. Lo studente i-esimo ha un punteggio x_i nel test di ingresso.

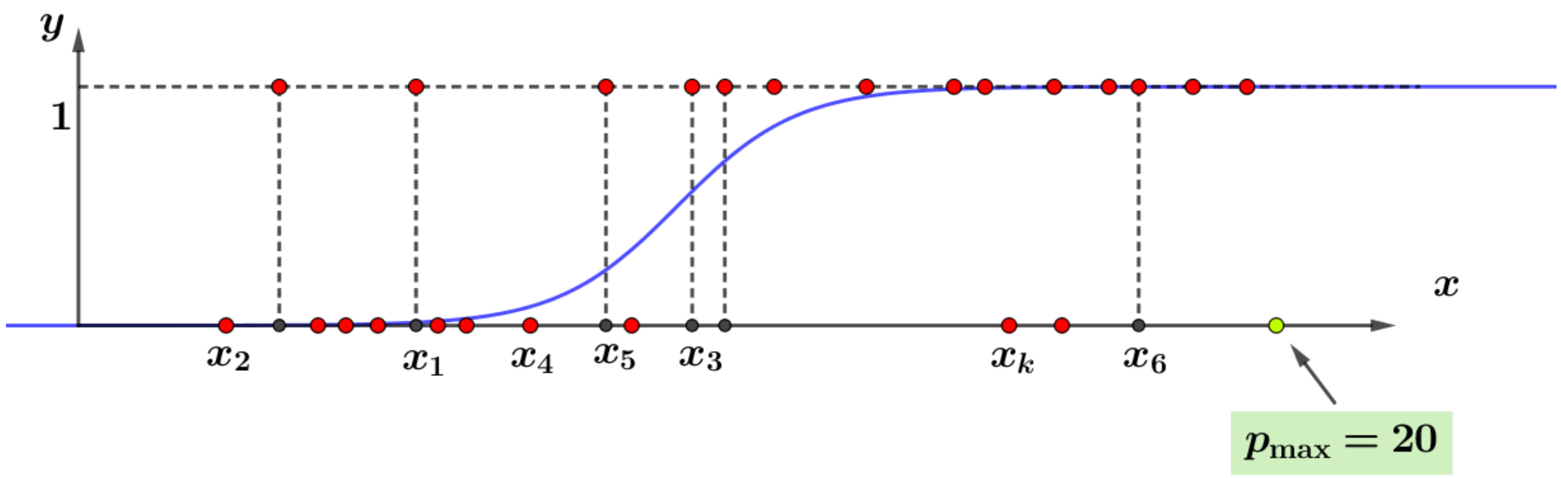
Per ogni i, definiamo il numero y_i, ponendo che

$y_i = 1$ se alla fine del primo anno lo studente ha conseguito 40 crediti o più;

$y_i = 0$ se lo studente ha conseguito meno di 40 crediti;

Se costruiamo una tabella delle coppie (x_i, y_i) e le rappresentiamo in un piano cartesiano, otteniamo una figura del tipo seguente.

x_1	x_2	x_3	x_4	x_5	x_6	...	x_k	...	...
1	**0**	**1**	**0**	**1**	**1**	...	**0**	...	...



dove abbiamo rappresentato in blu il grafico di una funzione $y = g(x)$ che tenta di prevedere quale sarà il valore di y sapendo il valore di x.

Si dice che funzione g è un modello predittivo.

Il valore g(x), che è compreso fra zero e uno, si può pensare come **la probabilità**, che il modello prevede, **che uno studente con punteggio x nel test superi 40 CFU** nel primo anno di studio

Di che tipo è la funzione g?

come si determina precisamente?

Si possono usare diversi tipi di funzioni, che devono andare da zero a uno ed essere crescenti. Qui abbiamo utilizzato la funzione logistica

$$g(x) = \frac{e^{k(x-a)}}{1 + e^{k(x-a)}}$$

Nella figura abbiamo scelto i parametri $k = 1$ e $a = 10$, solo per dare un'idea. I parametri devono essere determinati a partire dai dati (x_i, y_i) in modo che sia minimo l'errore complessivo, in inglese ***loss***, che il modello commette nella previsione. Come loss si può prendere la somma degli errori che vengono fatti nelle previsioni per ciascuno studente.

Ciascuno di questi errori si può misurare in diversi modi, ad esempio col numero $(y_i - g(x_i))^2$ o mediante la cosiddetta *entropy loss*, che scriveremo più avanti; per ora non occorre approfondire.

Il parametro a nella funzione logistica dice qual'è il punteggio del test per cui la **probabilità di successo è 1/2**.

Il parametro k dice quanto è ripida la funzione e quindi quant'è la **capacità di discriminazione** del test.

Possiamo usare il modello per ripartire gli studenti in due **classi**: quelli che hanno punteggio x minore di un certo valore t_0 e quelli che hanno punteggio $x \geq t_0$, e possiamo decidere che agli studenti della prima classe consigliamo di migliorare la loro preparazione. Un modello analogo si può fare usando il voto del diploma invece che il punteggio nel test. Le diverse funzioni logistiche che si ottengono nei due casi si possono usare per comparare tra loro le capacità predittive del voto di diploma e del punteggio del test.

I classificatori lineari in più variabili e il problema del riconoscimento dei caratteri manoscritti

Il riconoscimento automatico del testo è parte del più generale problema del **riconoscimento delle immagini**, che è una componente presente in molti sistemi di intelligenza artificiale.

Un testo può essere

- una pagina stampata,
- un documento di identità,
- un manoscritto medievale,
- un esercizio di algebra,
- uno spartito musicale,

…...

l’obiettivo del riconoscimento può essere di

- codificare il testo in qualche formato al fine di archiviarlo digitalmente
- presentarlo in una forma scritta o vocale più facile da capire, ad esempio per per aiutare una persona con difficoltà
- tradurlo in un’altra lingua
- …….

da molto tempo (1914) si sono costruiti
sistemi OCR (Optical Character recognition)
basati su diverse tecnologie
https://en.wikipedia.org/wiki/Optical_character_recognition#

recentemente sono stati realizzati sistemi accurati e molto veloci basati sul Machine Learning

ad esempio **Google Lens** https://lens.google/
che si può scaricare sul cellulare:
si fa una foto di una pagina o di una strada
e il sistema individua nell'immagine i testi, li riscrive in un file,
se volete li traduce o li legge ad alta voce

un altro esempio è il software **Mathpix** https://mathpix.com/

se gli date un'immagine con formule scritte a mano, matrici, integrali, figure, Mathpix vi produce il codice Latex e il pdf

1. $$\frac{e^{-\sqrt{\frac{3/4}{5/2}}-\log_3\left(1+\frac{1}{n}\right)^n}}{3!}$$

2.

1	2	3
4	5	6

$$\frac{e^{-\sqrt{\frac{3/4}{5/2}}-\log_3\left(1+\frac{1}{n}\right)^n}}{3!}$$

1	2	3
4	5	6

\(\Lambda \).

\[\frac{e^{-\sqrt{\frac{3 / 4}{5 / 2}}-\log _{3}\left(1+\frac{1}{n}\right)^{n}}}{3 !} \]

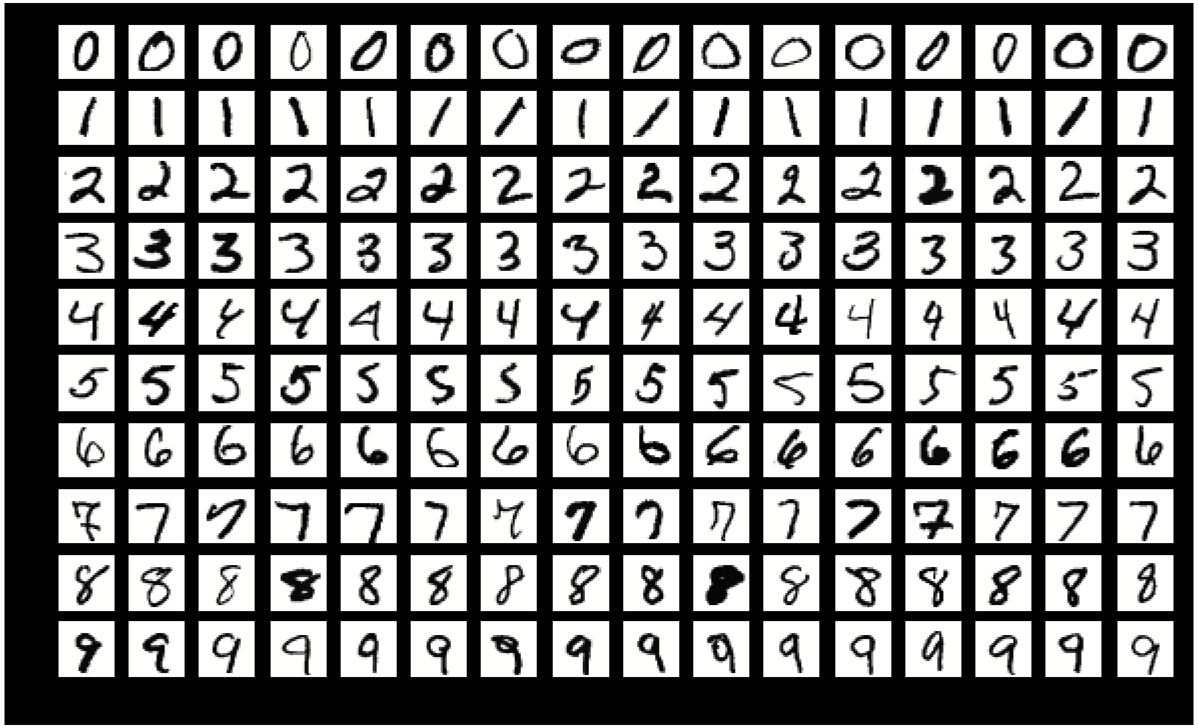
2. \begin{tabular}{|l|l|l|} \hline 1 & 2 & 3 \\ \hline 4 & 5 & 6 \\ \hline \end{tabular}

Non conosco questi software né precisamente gli algoritmi che essi usano, ma immagino che siano abbastanza complicati, anche se non sono tra le cose più alla frontiera di ciò che si sa fare oggi.

Per cominciare a capire il tipo di problema conviene prendere un caso più semplice.

Il riconoscimento delle cifre manoscritte

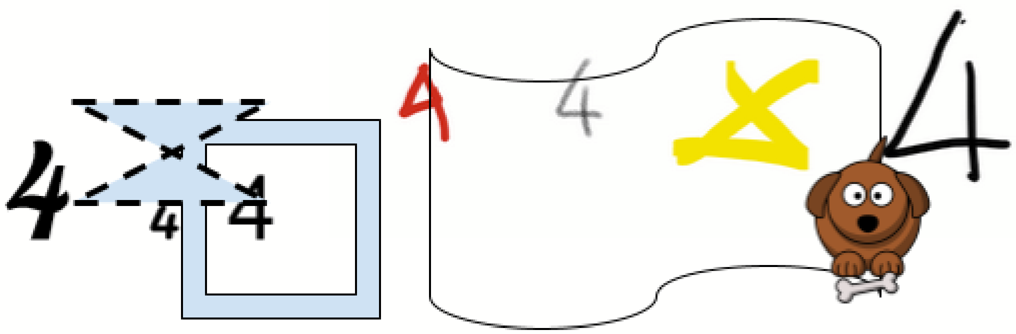
Questo problema è importante e ha una storia lunga. Esiste un database standard, che si chiama **MNIST** sul quale si misurano le prestazioni dei sistemi di riconoscimento https://en.wikipedia.org/wiki/MNIST_database



Ci limitiamo a un caso particolarissimo:

Come riconoscere un'immagine di una cifra, ad es. 4

La cifra 4 si può presentare in molti modi



e con sfondi variamente capaci di confondere

Cosa intendiamo per “immagine”?

Un’immagine è una matrice di numeri, che si ottiene facendo la scansione di una pagina o una fotografia.
In entrambi i casi un sistema ottico proietta e mette a fuoco l’immagine su un **sensore** rettangolare, suddiviso in un reticolo di celle.
Per ogni cella, che chiamiamo pixel, il sensore produce un numero, che dipende dall’intensità luminosa in quella cella.

Le fotografie fatte dai cellulari hanno milioni di pixel, ma le immagini nel database MNIST sono 28X28 pixel.

Se ci limitiamo a considerare **immagini in bianco e nero**, in ogni cella avremo il numero 1 (bianco) o il numero 0 (nero).
Anche se un'immagine si presenta naturalmente come una matrice di numeri, conviene pensarla come una **lista**, ossia una n-upla ordinata, ossia come un **vettore**. Dunque,

un'immagine è un vettore $\mathbf{x}$ in uno spazio $\mathbf{R}^n$

dove n è il numero di pixels; n=784 nel database MNIST;

le immagini 28X28 in b/n sono 2^{784}

gli elementi x_j del vettore sono

- 1 oppure 0, se l'immagine è in bianco e nero
- interi tra 0 e 255, se ha 8 bit di profondità di grigio
- terne di numeri, se è a colori

Il problema del riconoscimento è quindi:
determinare la funzione

$$f : \mathbf{R}^n \to \{0, 1\}$$

che assegna il valore 1 a ogni vettore in cui si vede la cifra 4, e assegna il valore zero agli altri.

Una funzione di questo tipo si chiama ***classificatore***.

Non è necessario, né praticamente possibile,
determinare *esattamente* la funzione f
è sufficiente trovare una sua **buona approssimazione**.

Come possiamo fare?

Si può provare a **scegliere una qualche classe** $\mathcal{H}$ **di funzioni** e tra queste **trovare una funzione** g **che**, **su un certo insieme di esempi di immagini sbaglia meno di altre** e quindi **è una possibile approssimazione della funzione** f **che vorremmo**.

Perché questo metodo dovrebbe funzionare?

Se la funzione f che fa il riconoscimento fosse totalmente pazza, **non** avremmo speranze .

Ma l'idea è che la funzione f non sia del tutto irregolare e che se la classe di funzioni $\mathcal{H}$ è abbastanza ampia e flessibile, lì dentro ci siano funzioni che approssimano ragionevolmente f.

Questa idea, in modo molto generale, sta dietro anche allo sviluppo in serie di Fourier o al teorema di Stone-Weierstrass sulla densità dei polinomi nello spazio delle funzioni continue su un compatto.

Come scegliere la classe di funzioni $\mathcal{H}$?
Come trovare una buona funzione approssimante?

Una prima idea: **prendere la classe delle funzioni del tipo**

$$h(x) = \varphi(t) = \varphi\left(\sum_j w_j x_j\right)$$

dove $\mathbf{x}$ è un generico vettore di $\mathbf{R}^\mathbf{n}$, w_j sono coefficienti reali e φ è una funzione ***gradino*** da $\mathbf{R}$ in $\mathbf{R}$ che ha valore zero se l'argomento

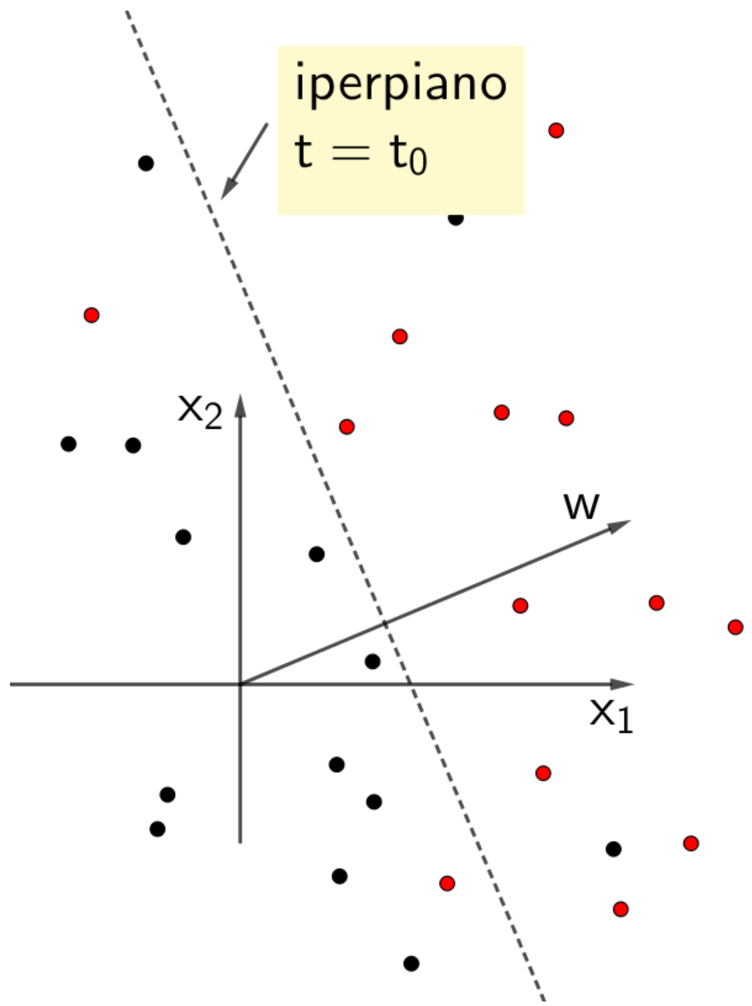
$$t = \sum_j w_j x_j$$

è minore di un certo valore soglia, e invece ha valore 1 se l'argomento è maggiore o uguale della soglia. Una tale funzione si dice **percettrone** ed è un tipo di **classificatore lineare** https://it.wikipedia.org/wiki/Classificatore_lineare

Il grafico della funzione φ è quindi

e la funzione h ha valore 1 oppure zero da una parte o dall’altra dell’iperpiano

$$\sum_j w_j x_j = t_0$$



Rappresentazione di un percettrone in due dimensioni; $w = (w_1, w_2)$ è il vettore dei coefficienti; l'iperpiano (retta) $t = w_1 x_1 + w_2 x_2 = t_0$ è ortogonale a w;
da una parte dell'iperpiano il classificatore ha valore 1, dall'altra ha valore zero.
Si cerca w in modo che l'iperpiano separi meglio possibile le due classi, ma qualche errore ci sarà.

Insomma: un **percettrone** è una funzione h che si ottiene dalla **composizione di una funzione lineare**

$$L : x \mapsto t = \sum_j w_j x_j$$

con una funzione gradino $\varphi : t \mapsto y$.

Per ogni scelta del vettore dei coefficienti $\mathbf{w}$ e del numero t_0 otteniamo un classificatore h. C'è dunque una certa libertà di scelta e la speranza è che scegliendo bene i coefficienti si riesca a trovare un classificatore che non sbagli troppo.

La funzione gradino però ha il difetto di non essere derivabile e al suo posto useremo d'ora in poi la funzione logistica

$$\psi(t) = \frac{1}{1+e^{-k(t-t_0)}} = \frac{e^{k(t-t_0)}}{1+e^{k(t-t_0)}}$$

che è una specie di "gradino smussato" e ha 2 parametri
https://en.wikipedia.org/wiki/Sigmoid_function

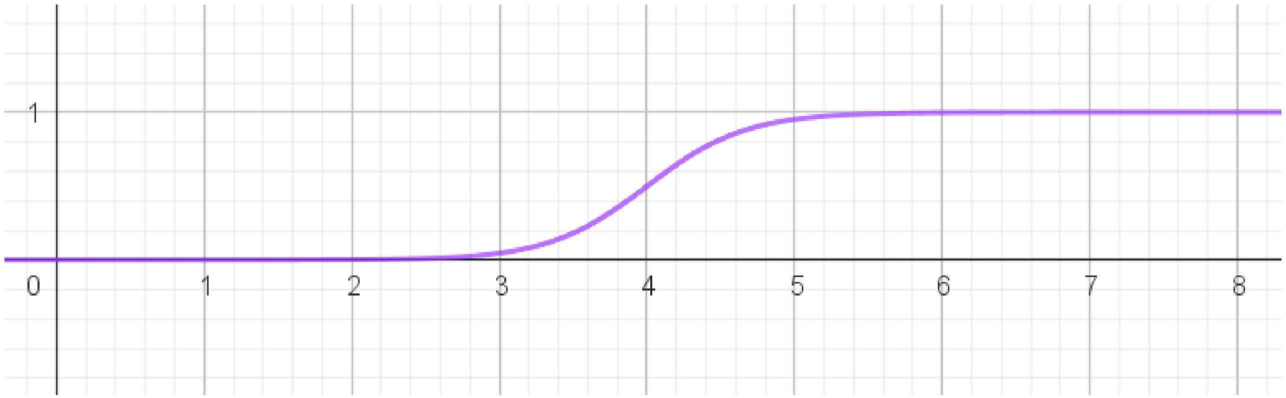


grafico della funzione logistica con k=3 e $t_0 = 4$

Il funzionamento di un percettrone
o di un classificatore lineare
è *in qualche modo simile* a quello di
un neurone del cervello:

riceve impulsi da altri neuroni e,

se la somma degli impulsi supera una certa soglia,

“spara” a sua volta un impulso

Osservazione di passaggio

La funzione logistica https://www.geogebra.org/m/rNGeYdmV ha proprietà interessanti ed è anche soluzione dell'**equazione differenziale logistica**, che ha interpretazioni in molti campi. Un buon riferimento è https://en.wikipedia.org/wiki/Logistic_function; consiglio meno la voce in italiano.

In generale osservo che, piuttosto che studiare funzioni prive di significato, conviene studiare funzioni che sono soluzioni di equazioni differenziali interessanti, e guardare come cambia il comportamento della funzione soluzione, al variare dei parametri dell'equazione differenziale, i quali dicono come sono le condizioni del fenomeno e i dati iniziali.

Ma come si fa determinare $\mathbf{w}$, k e t_0 in modo da ottenere il migliore o, almeno, un buon classificatore nella classe?

Si prende un insieme di immagini per le quali si sa qual è la risposta corretta che il classificatore dovrebbe dare (*insieme di allenamento*, o *training set*).
Ricordiamo che la i-esima immagine è un vettore

$$\mathbf{x}^i = (x_1^i, x_2^i, \ldots, x_j^i, \ldots, x_n^i)$$

Per ogni indice i definiamo il numero y^i ponendo che sia uguale a 1 se nella i-esima immagine si vede la cifra 4, e sia invece zero altrimenti.

Avremo così che il generico classificatore individuato dai coefficienti $\mathbf{w}$, k e t_0, manda l'immagine $\mathbf{x^1}$ nel numero

$$z^1 = \psi\left(\sum_j w_j x_j^1\right)$$

che è una funzione derivabile a piacere dei coefficienti.

Noi vorremmo che z^1 fosse uguale a y^1 e ci serve un modo per valutare di quanto invece il classificatore si sbaglia.

Un modo è di prendere come errore $E(\mathbf{w}, k, t_0, \mathbf{x^1}, y^1)$ la quantità $(z^1 - y^1)^2$.

Un altro modo molto usato è di prendere la quantità

$$-y^1 \ln(z^1) - (1 - y^1) \ln(1 - z^1)$$

(*entropy loss*).

Osserviamo che in ogni caso l'errore $E(\mathbf{w}, k, t_0, \mathbf{x^1}, y^1)$ sulla prima immagine è una funzione derivabile quanto si vuole delle variabili w_j e t_0, che in questo contesto si chiamano di solito ***pesi***.
Definiamo ora l'errore totale del classificatore sul training set come la somma degli errori su tutte le immagini di esempio

$$\mathcal{L} = \sum_i E(\mathbf{w}, k, t_0, \mathbf{x^i}, y^i)$$

Questa funzione si chiama loss, o perdita, o errore complessivo. Dipende dai dati di esempio ed è anch'essa derivabile quanto si vuole rispetto ai pesi w_j, k, t_0.

Si tratta ora di vedere se si riesce a trovare un classificatore per cui la funzione loss è minima, o anche soltanto piccola, in modo che il classificatore sbagli poco.
Non entro nei dettagli, ma l'idea è che questa ricerca si fa con algoritmi numerici di ottimizzazione, partendo da valori dei coefficienti scelti arbitrariamente e poi seguendo il gradiente discendente della funzione loss, cercando di non imbucarsi in un minimo locale.
Una volta trovato un classificatore che sbaglia poco sul training set, bisognerà verificare come se la cava con altri esempi…

Nel database MNIST ci sono appunto un set di 60 mila immagini per trovare il classificatore, e anche un altro set di 10 mila immagini, per fare i test di verifica.
Secondo il sito http://yann.lecun.com/exdb/mnist/ (che è una specie di archivio sacro, non più aggiornato), usando classificatori lineari del tipo descritto, nel 1998 si è ottenuto un sistema di riconoscimento che aveva un errore del 12%.
Oggi si ottengono accuratezze significativamente migliori, minori dello 0,5% e paragonabili con quelle degli umani, utilizzando altre classi di funzioni, in particolare le cosiddette Deep Neural Networks DNN, che hanno molti strati, e usando essenzialmente lo stesso metodo di allenamento, o learning, descritto sopra.

Ma questo è niente rispetto a ciò che riescono a fare le DNN nel riconoscimento di immagini con rumore.

https://www.researchgate.net/figure/Classification-error-for-all-models-on-recognition-in-MNIST-with-and-without-Gaussian_fig5_345157989

Vediamo allora nella prossima sezione cosa sono le reti neurali

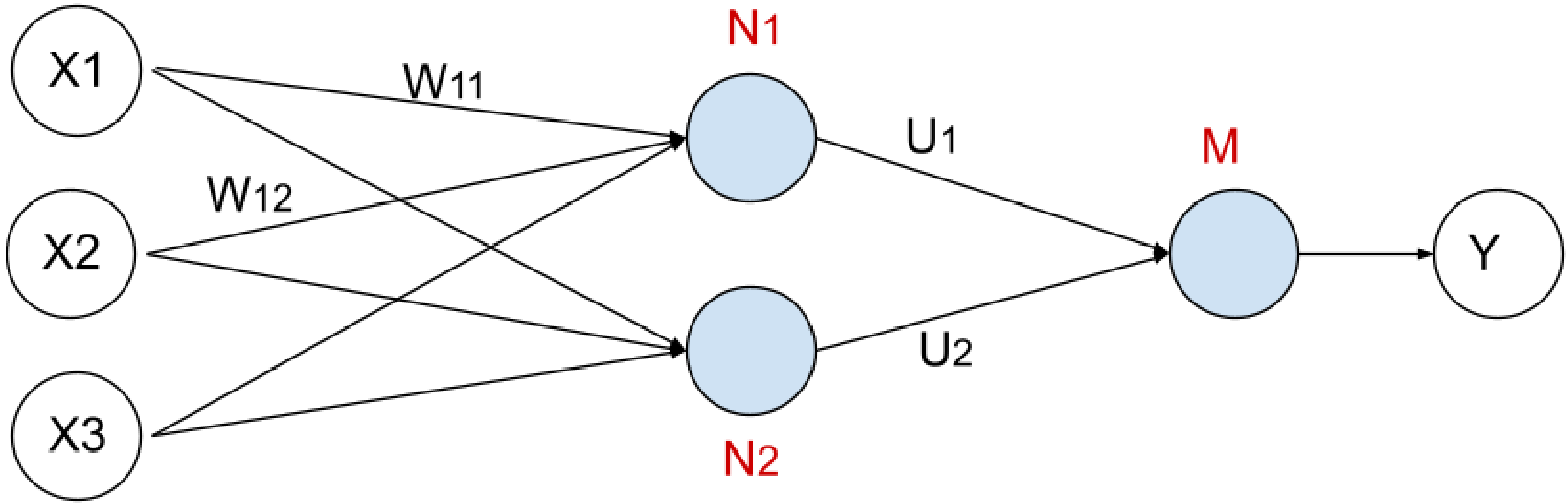
4. reti neurali artificiali

Una **rete neurale artificiale**, brevemente **rete neurale**, è una **classe di funzioni**, che si ottengono componendo opportunamente funzioni lineari di molte variabili e funzioni soglia di una variabile. Come funzioni soglia si usano ad esempio una funzione logistica fissata o la funzione

$$\eta(t) = \max\{0, t\}$$

In rete si trovano molte presentazioni dell'argomento, più o meno precise e approfondite. Una buona via di mezzo mi sembra quella che si trova su Wikipedia in italiano https://it.wikipedia.org/wiki/Rete_neurale_artificiale

Vediamo una semplice rete neurale rappresentata nella prossima figura



La rete ha tre ingressi x_1, x_2, x_3 in **R** e un'uscita y, anche in **R**
Ha **tre nodi** N_1, N_2, M, che si pensano su **due strati**:
uno strato N, che ha due nodi
e uno strato M che ha un nodo solo.

Ricordiamo che le reti neurali che si usano comunemente sono molto più complesse e possono avere centinaia di strati, ciascuno con migliaia di nodi, e un'architettura assai più complessa.
Nel nodo N_1 convergono tre collegamenti che provengono dagli ingressi x_j e che hanno i coefficienti w_{1j}. Dobbiamo pensare il nodo come una funzione che somma i tre prodotti $x_j w_{1j}$ e applica a tale somma una funzione non lineare, ad esempio una logistica φ (o la funzione η detta sopra).
Dai nodi N_1 e N_2 escono quindi i numeri

$$t_1 = \varphi\left(\sum_j x_{1j} w_{1j}\right) \qquad t_2 = \varphi\left(\sum_j x_{2j} w_{2j}\right)$$

Questi numeri t_1, t_2 vanno al nodo M passando per i collegamenti che hanno coefficienti u_1, u_2. Il nodo M li somma e applica di nuovo la funzione φ. Ecco il risultato:

$$y = \varphi\left(t_1 u_1, t_2 u_2\right) = \varphi\left(\varphi\left(\sum_j x_j w_{1j}\right) u_1, \varphi\left(\sum_j x_j w_{2j}\right) u_2\right)$$

Ogni nodo è dunque un classificatore lineare del tipo che abbiamo visto in precedenza: **riceve impulsi dai neuroni dello strato precedente e manda in uscita una funzione non lineare della somma degli ingressi.**

Ogni nodo si comporta in modo un po' simile a un neurone del cervello e da qui viene il nome "rete neurale" o "rete neuronale".

Insomma, la nostra rete neurale non è altro che una classe di funzioni, determinate dai valori dei parametri w_{ij}, u_i che chiamiamo "pesi". Ciascuna di tali funzioni prende come ingresso un vettore $\mathbf{X} = (x_1, x_2, x_3)$ di dimensione 3 e ha in uscita un numero y.

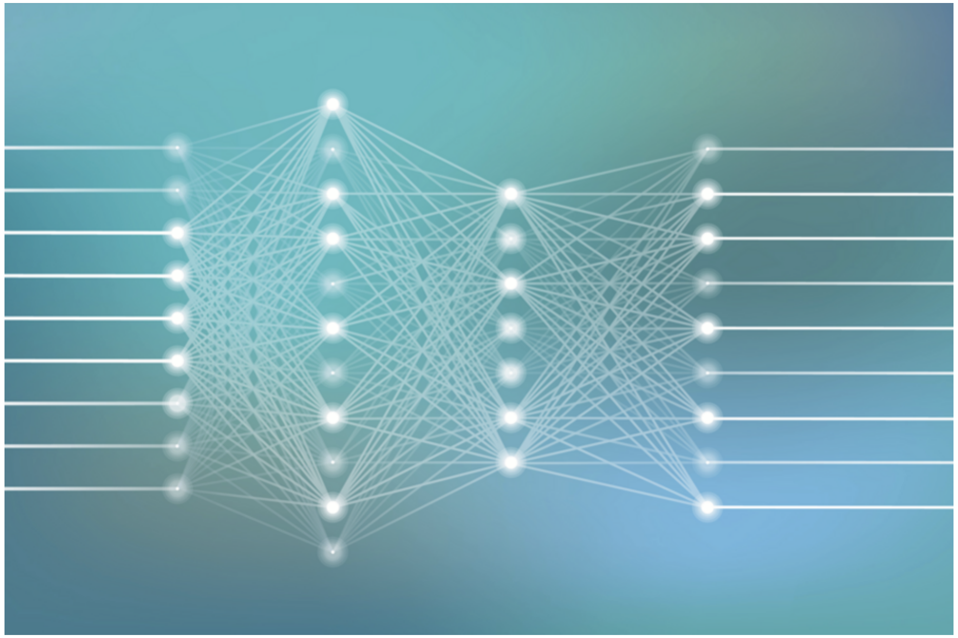
Supponiamo ora di avere **un insieme di dati di ingresso**

$\mathbf{X}^i = (x_1^i, x_2^i, x_3^i)$ dove i varia da 1 a un certo n,

per ciascuno dei quali sappiamo quale deve essere l'uscita y^i.

Possiamo allora cercare i valori dei pesi w_{ij}, u_i **in modo che sia minima la loss** su tale insieme di dati. Questo processo si chiama **training** della rete.

Esempi di reti neurali più complesse si trovano facilmente in internet. Ad esempio, una figura elegante, ma ancora piuttosto semplice, si trova qui
https://news.mit.edu/2020/looking-black-box-deep-learning-neural-networks-0727



nel sito si legge la frase:

“**Deep learning was in some ways an accidental discovery**,” explains **Tommy Poggio**, investigator at the McGovern Institute for Brain Research, director of the Center for Brains, Minds, and Machines (CBMM), and the Eugene McDermott Professor in Brain and Cognitive Sciences.

“**We still do not understand why it works.** A theoretical framework is taking form, and I believe that we are now close to a satisfactory theory. It is time to stand back and review recent insights.”

Mi fa piacere ricordare che Tomaso Poggio ha soggiornato a Trento per diversi periodi di molti mesi alla fine degli anni ‘80, contribuendo alla formazione di gruppi di ricerca ancora oggi attivi in FBK. Tenemmo allora un seminario su ciò che si sapeva a quel tempo su questi temi. Parteciparono Luciano Tubaro, Cesare Furlanello, Bruno Caprile e Federico Girosi. Con quest’ultimo scrissi un articolo: Convergence rates of approximation by translates PdF da dtic.mil

Una bella esposizione
delle reti neurali,

e di come si applicano proprio al riconoscimento delle cifre manoscritte, viene fatta in una serie di video che si trovano in questo sito

https://www.3blue1brown.com/topics/neural-networks

(ricchissimo anche di altri argomenti)

Questi video non sono facilissimi, ma meriterebbero di essere studiati in tutte le scuole e aiuterebbero anche a imparare l'inglese.

5. il curricolo

In questa sezione farò alcuni esempi di problemi e attività che si possono inserire nel curricolo a diversi livelli e in diversi momenti, con particolare riferimento ai temi:

funzioni, algebra, geometria analitica

Non ho avuto molto tempo di preparare esempi dettagliati e comunque qui non avrei il tempo di raccontarli. Dunque cercherò soprattutto di dare delle idee e di indicare direzioni di lavoro.

Sono disponibile ad approfondire questi temi in ulteriori incontri con le persone eventualmente interessate.

5.1. Le funzioni

In molte situazioni è comodo pensare le cose in termini di funzioni e insiemi. Quando si presenta il caso, si dovrebbe cogliere l'occasione per farlo, mettendo in evidenza i concetti e usando i simboli e il linguaggio opportuni, senza esagerare nella formalizzazione.

Questo lavoro è importante che cominci nella scuola media e prosegua con continuità fin dall'inizio delle scuole superiori.

estratto dalle Indicazioni nazionali 2012 per la fine della terza media

– Interpretare, costruire e trasformare formule che contengono lettere per esprimere in forma generale relazioni e proprietà.

– Usare il piano cartesiano per rappresentare relazioni e funzioni. Conoscere in particolare le funzioni del tipo

$$y=ax, \qquad y=a/x, \qquad y=ax^2, \qquad y=2^n$$

e i loro grafici e collegare le prime due al concetto di proporzionalità.

– Esplorare e risolvere problemi utilizzando equazioni di primo grado.

Le funzioni lineari, le funzioni quadratiche e la funzione esponenziale, in diversi contesti e con complessità crescente sono importanti in tutto il curriculum delle superiori e all'università (anche nei test di ingresso).

Sono la base anche del linguaggio che occorre per parlare di Machine Learning e di Reti Neurali.

A scuola conviene cominciare presto a usare il linguaggio delle funzioni per descrivere il comportamento di fenomeni e di sistemi. Si noti che già nel primo ciclo è indicata la funzione "esponenziale discreta", che manda il numero intero n nel numero 2^n.

Osservazione sull' "operazione" potenza

La potenza x^n viene chiamata di solito "operazione", ma questa parola è fuorviante perché suggerisce un'analogia con le operazioni binarie di somma e prodotto, che sono invece molto diverse.

È molto meglio pensare che abbiamo due funzioni:

fissato n , abbiamo una **funzione** che manda x in x^n

fissato x, abbiamo una **funzione** che manda n in x^n

La prima si chiama **funzione potenza**. La seconda si chiama **funzione esponenziale**. Di queste funzioni si costruiscono tabelle e grafici, si trovano le proprietà, si definisce la funzione inversa. La funzione esponenziale si estende a $\mathbf{Z}, \mathbf{Q}, \mathbf{R}$.

Osservazione sul grafico di una funzione

Gli studenti tendono a

- identificare una funzione col suo grafico
- a identificare il grafico col disegno del grafico.

Sono invece cose ben distinte.

Questo è un serio problema che impedisce di progredire nella comprensione del concetto di funzione e di usare efficacemente il linguaggio delle funzioni.

Invito a considerare la questione.

Osservazione sulla necessità di introdurre presto le funzioni di più variabili.

Non si capisce perché a scuola si parla così poco di funzioni di più variabili oppure a valori vettoriali.

Io credo che sia per due ragioni:

- una volta si parlava di funzioni solo quando si arrivava al calcolo differenziale e integrale, e questo calcolo, in più variabili, è un po' difficile e delicato; perciò si evita;
- c'è una tendenza generale a fare confusione tra grafico e funzione; dunque quando il grafico è difficile o impossibile da vedere sembra che venga meno la possibilità di considerare la funzione.

Ma ci sono molte cose utili, interessanti, belle, importanti e non difficili che si possono fare in più variabili, anche se non si disegna il grafico. Ad esempio all’interno di questi argomenti

- le funzioni lineari in n variabili,
- le forme quadratiche in due variabili,
- le funzioni a simmetria circolare,
- la funzione di stato dei gas perfetti e quella dell’acqua,
- le curve nello spazio.

Esempio 1

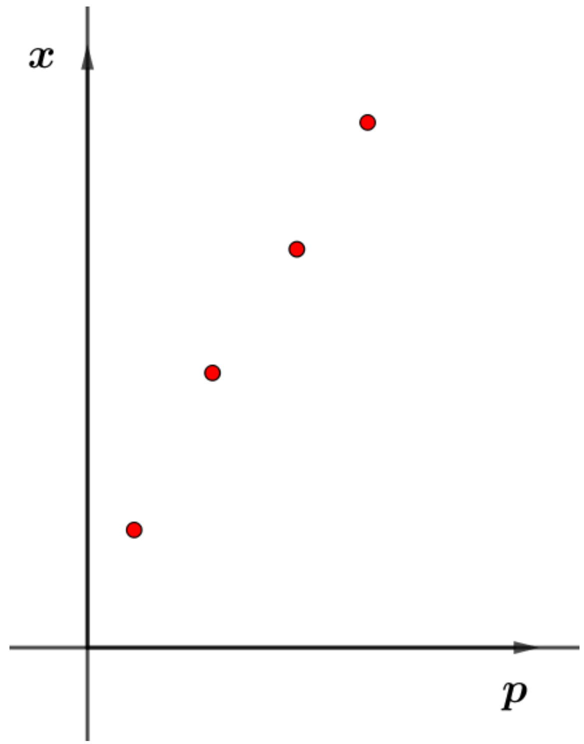
Un esempio interessante e istruttivo di fenomeno fisico da modellizzare è la **deformazione di un corpo elastico soggetto a una forza**.

Il caso più semplice è quello di **una molla alla quale si appende un oggetto.**

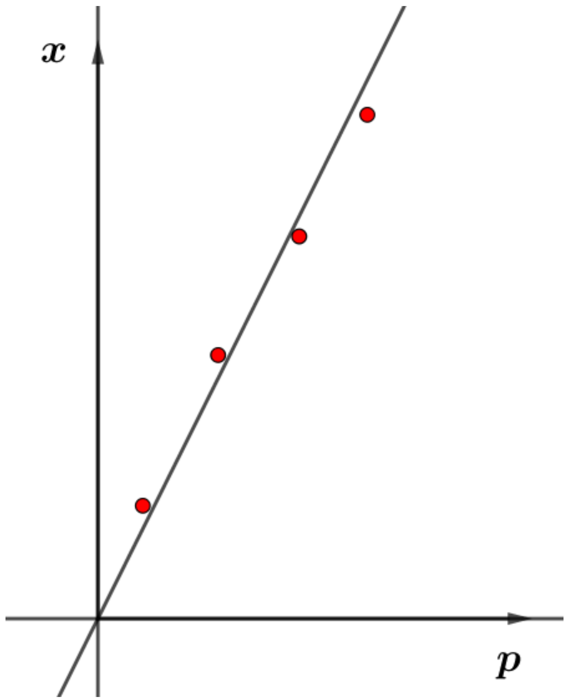
Per diversi valori del peso x si può misurare l'allungamento y della molla, e le coppie di valori (x, y) si possono scrivere in una tabella. Poi su un piano cartesiano si possono disegnare i punti che hanno tali coordinate.

Si ottiene allora una figura del tipo seguente

peso	allungam.
0	0
x_1	y_1
x	y_2
…	…
x_i	y_i



e si vede che i punti sono approssimativamente su una retta.



La retta che si vede nella figura qui a fianco è solo una delle tante rette possibili.
Se chiamiamo k la sua pendenza, allora l'equazione della retta è
$y = kx$.
Possiamo prendere la funzione
$f\colon x \to kx$ come modello del comportamento della molla.
La funzione f è lineare.

Ma qual è allora la retta "giusta"?

(questa slide e la successiva ripetono concetti e calcoli già visti in precedenza)

Non c’è una retta “giusta”.

Si possono scegliere dei criteri che ci sembrano significativi e poi si può cercare se c’è una retta “migliore” di tutte le altre rispetto a tali criteri.

Una possibilità è di prendere la retta di equazione

$$y = kx$$

per la quale è minima “la distanza” dai punti di coordinate (x_i, y_i).

Ma come si definisce tale “distanza”? Ci sono molti modi.

Un modo comodo è di prendere come errore la quantità

$$L = \sum_i \left(y_i - kx_i\right)^2$$

che è una funzione del parametro k e dei dati (x_i, y_i)
Pensando che i dati sono fissati, L è funzione solo di k.
Per minimizzarla si può calcolare la derivata di L rispetto a k

$$\frac{dL}{dk} = -2\sum_i (y_i - kx_i)x_i$$

e la si può porre uguale a zero. Così si ottiene subito

$$\sum_i (y_i x_i - kx_i^2) = 0 \qquad \text{e} \qquad k = \frac{\sum_i x_i y_i}{\sum_i x_i^2}$$

Il metodo usato sopra per trovare la retta “migliore” si chiama metodo dei **minimi quadrati**.

Può capitare che per approssimare i dati sia meglio prendere una retta che non passa per l’origine. In tal caso si cerca la retta tra quelle che sono grafici di funzioni del tipo

$$y = f(x) = ax + b$$

Oppure può essere meglio usare funzioni polinomiali, ad esempio di secondo grado

$$y = f(x) = ax^2 + bx + c$$

In tutti questi casi funziona ancora bene il metodo dei minimi quadrati, perché questi modelli sono “lineari nei parametri” a,b,c,...

In effetti, tecniche simili, anche se più complesse, si possono usare per **fenomeni o sistemi che sono descritti da modelli che non sono lineari nei parametri**. In qualche caso i modelli si linearizzano con qualche trucco. Ad esempio, un fenomeno istruttivo e facile da sperimentare è l'andamento nel tempo della temperatura di un corpo di piccole dimensioni che si sta raffreddando. Si vede che un buon modello è la funzione esponenziale 2^{-at} con a opportuno. La funzione 2^{-at} non è lineare nel parametro a, ma se ci si mette in scala logaritmica il modello diventa lineare.
Questo modello si chiama *Legge di raffreddamento di Newton*, e di solito viene formulato come "*la derivata della temperatura è proporzionale alla differenza di temperatura tra l'ambiente e il corpo*". Tale formulazione richiede una qualche conoscenza del concetto di derivata, ma si può sostituire con la pendenza media in un piccolo intervallo di tempo. Per la prima formulazione non è richiesto nulla.

Commento

Il modello lineare per la relazione tra forza e deformazione si chiama comunemente “Legge di Hooke”. A scuola si propone talvolta di “verificare sperimentalmente” la Legge di Hooke. Meglio è non nominare alcuna legge ed esplorare invece la relazione tra le due grandezze, mettendo pesi anche grandi e scoprendo che oltre un certo valore di p la relazione $f\colon\ p \ \to x$ non è più lineare.

E osservando sperimentalmente che: a variazioni u piccole, che partono da un valore fissato P del peso, corrispondono variazioni v dell’allungamento che sono approssimativamente uguali a ku, dove la costante k dipende dal valore P.

Dunque il modello lineare vale ancora per le **variazioni** in piccolo! la funzione $u \ \to \ kv$ è il differenziale della funzione f nel punto P e k è la derivata di f nel punto P).

Ci sono molti fenomeni che, entro certi limiti, si possono descrivere bene con funzioni lineari e che si studiano a scuola. Ad esempio la relazione tra la **corrente** che passa in un resistore e la **differenza di potenziale** ai suoi estremi, oppure la relazione tra il **volume** e la **temperatura** di una certa quantità di gas a pressione costante.

Tutti questi fenomeni vanno bene per riflettere sulla linearità e dovrebbero essere utilizzati.

Un vantaggio della deformazione elastica è che le grandezze in gioco sono facilmente misurabili "a mani nude" con strumenti elementari: metro e bilancia. E infatti Hooke fa i suoi esperimenti intorno al 1660, molto prima che si studiassero i comportamenti dei gas e della corrente elettrica.

Osservazione sul linguaggio della proporzionalità, sulle sue origini storiche e sull’opportunità di superarlo

A scuola si usa molto il linguaggio della proporzionalità per esprimere relazioni tra grandezze e modelli di fenomeni. Inoltre abbiamo anche la proporzionalità inversa e quadratica. Questi termini possono essere utili, soprattutto per chi li usa da sempre, ma ci si dovrebbe interrogare sull’opportunità di usarli per superarli.

Il concetto di proporzionalità dovrebbe essere usato per portare rapidamente, già in una variabile **e soprattutto in più variabili,** al principio di “sovrapposizione degli effetti” e poi al concetto di **funzione lineare**, che è quello di gran lunga più generale e utile.

La proporzionalità inversa e quella “quadratica” sono solo due casi, per quanto importanti, del modello potenza x^{α} e del modello polinomiale. Va bene cominciare da quelli, ma lo studente rischia di pensare che il mondo finisce lì.
Anche sulla prassi comune di studiare abbastanza dettagliatamente le proporzioni nella scuola media si dovrebbe riflettere. A mio parere il linguaggio delle proporzioni va comunque imparato, poiché consente di esprimere a parole, e pensare, concetti e relazioni importanti. **L’uso delle proporzioni andrebbe invece abbandonato del tutto per calcolare**, poiché l’algebra ci permette di fare questo con maggiore rapidità e sicurezza. Naturalmente, mi farebbe piacere discutere di questa mia convinzione.

L’importanza storica delle proporzioni. Credo sia interessante chiedersi quali sono le ragioni dell’importanza storica delle proporzioni.
La teoria delle proporzioni si trova nel V Libro degli Elementi di **Euclide** ed è notevolissima. Per più di mille anni è stata sostanzialmente l’unico linguaggio per parlare delle relazioni quantitative fra grandezze e per fare calcoli.
Galileo descrive la cinematica della caduta dei gravi dicendo essenzialmente che “gli spazi percorsi in tempi uguali stanno fra loro in proporzione come la sequenza dei numeri dispari”. Noi diciamo che lo spazio va col quadrato del tempo.
Newton descrive la legge del raffreddamento di un corpo dicendo che “la velocità di raffreddamento è proporzionale alla differenza di temperatura tra il corpo e l’ambiente”, che equivale a dire che la legge è esponenziale.

Le **leggi dei gas ideali e le leggi delle reazioni chimiche** vengono formulate tra XVIII e XIX secolo utilizzando il linguaggio della proporzionalità e della proporzionalità inversa.

Le proporzioni come strumento di calcolo erano necessarie quando non c'era una notazione comoda per i numeri e quando non c'era il calcolo letterale (la notazione decimale, inventata in oriente, si diffonde in Europa nel Medioevo, portata da Fibonacci; il calcolo letterale si diffonde nel XVI secolo).

Ma, appena le cose si fanno più complesse, il sistema per scrivere i numeri e l'algebra sono strumenti di calcolo di gran lunga più efficienti delle proporzioni.

Le proporzioni rimangono comunque nell’educazione elementare a lungo, probabilmente perché l’algebra si faceva solo al ginnasio ed era invece necessario che artigiani e commercianti, che non andavano al ginnasio, avessero già con la formazione elementare qualche strumento per risolvere problemi.

Oggi però il modo in cui vengono trattate le proporzioni a scuola dovrebbe essere radicalmente ripensato.

5.2. Dati, liste, operazioni, sistemi lineari

il magazzino di un negozio

i clienti di un professionista

le cartelle cliniche di un ospedale

l’anagrafe di un’azienda sanitaria

i miei vecchi dischi di vinile

le squadre di calcio di serie A, i calciatori e i gol segnati

i miei studenti

sono situazioni in cui è necessario archiviare e gestire dati

i dati si presentano in forma di **liste,** ossia n-uple ordinate

Ad esempio, un negozio deve avere

elenco dei prodotti che il negozio vende

prezzi di acquisto e di vendita

quantità in magazzino

quantità vendute giornaliere

elenco dei fornitori

elenco degli ordini

…………..

si possono poi fare liste di liste
e liste di (liste di liste)
eccetera…

una lista di numeri si chiama anche **vettore**
e una lista di liste si può pensare come una **matrice**

tra le liste di dati ci sono usualmente delle **relazioni**,
per descriverle con precisione occorre
il linguaggio degli insiemi e delle funzioni

in questo modo si costruiscono degli **archivi** o **database**

un foglio excel o un Google sheet è già un archivio,

ma si possono mettere relazioni tra più fogli ed esistono dei

software specifici per situazioni più complesse tra due liste di numeri ci sono **operazioni naturali di base**,

mediante le quali si fanno operazioni più complesse, ad esempio

somma   +

prodotto *

prodotto scalare

prodotto di matrici

Molto comodo è scrivere i sistemi di equazioni lineari utilizzando il linguaggio dei vettori e delle matrici

Scriviamo sotto a destra **in forma vettoriale** il sistema a sinistra

$$\begin{cases} 3x + 2y = 1 \\ x - 5y = 4 \\ -x + 3y = 0 \end{cases} \qquad x \begin{bmatrix} 3 \\ 1 \\ -1 \end{bmatrix} + y \begin{bmatrix} 2 \\ -5 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \\ 0 \end{bmatrix}$$

in breve il sistema si scrive x**U**+y**V = W** dove U è la colonna dei coefficienti di x, V è la colonna dei coefficienti di x, W è la colonna dei termini noti.

Risolvere il sistema vuol dire scrivere il vettore W dei termini noti **come combinazione lineare dei vettori U, V dei coefficienti.** I parametri della combinazione lineare sono le incognite x e y . Vedremo l'interpretazione geometrica di questo.

Per archiviare i dati e operare su di essi
ci sono **linguaggi** specifici e linguaggi generali
in particolare: **Matlab** e **Python**
Python si può usare in un notebook jupiter https://jupyter.org/ ad esempio utilizzando lo strumento **Google Colab**

che è gratuito.

Al seguente link si trova un tutorial specifico sull'uso di Google Colab per l'insegnamento e l'apprendimento della matematica https://colab.research.google.com/drive/1fzPtJLKKdZuEzcohtXRkdUcO6XfGNbdV

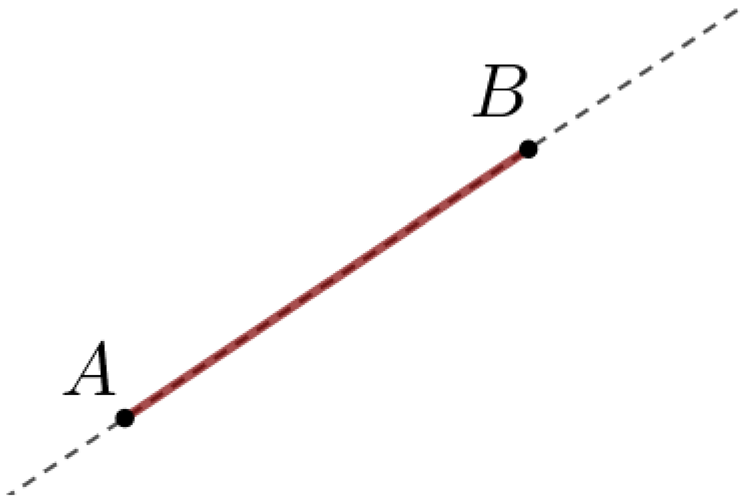
5.3. I vettori in geometria

Il concetto di vettore nel piano (e nello spazio) si definisce nell'ambito della geometria euclidea e consente di introdurre strutture algebriche sugli insiemi di oggetti geometrici, senza bisogno di utilizzare le coordinate. In questo modo

- si introduce un linguaggio che consente di descrivere le proprietà degli oggetti e delle trasformazioni in modo sintetico ed efficace, indipendente dalla dimensione dello spazio;
- le dimostrazioni sono indipendenti dalla scelta di una base e alcune dimostrazioni diventano immediate;
- il linguaggio funziona anche in dimensione infinita (Hilbert, Banach).

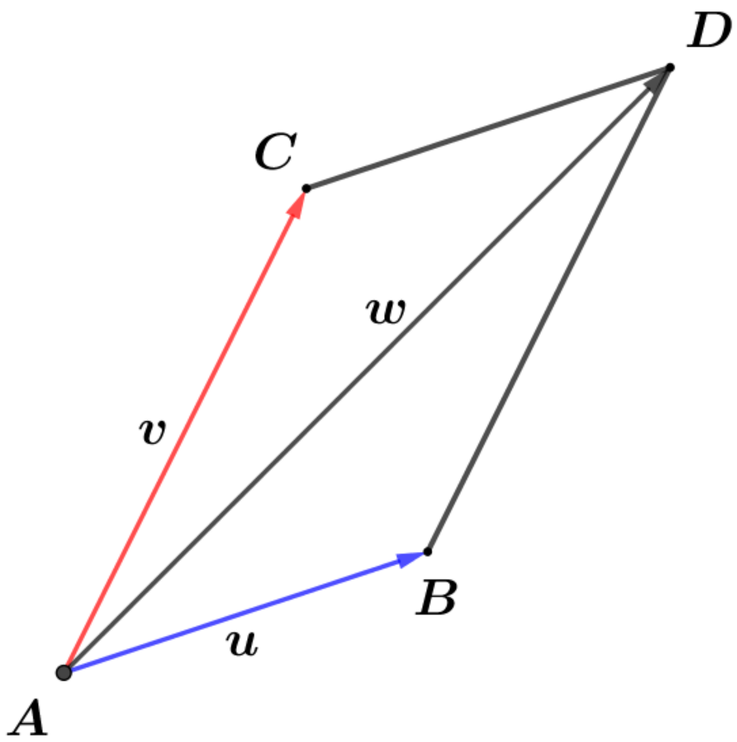
La parola “vettore” si può cominciare a usare a scuola il giorno dopo che si è usata la parola “segmento”.

Un **segmento è una coppia** (non ordinata) **di punti** (A, B) del piano o dello spazio, che si dicono estremi del segmento. Il segmento si può denotare $\overline{AB}$ e in un disegno si rappresenta come la parte della retta per i punti A e B, che sta fra i punti stessi.



Il **vettore è una coppia ordinata di punti**, ossia un **segmento del quale si è deciso di scegliere uno degli estremi come primo estremo**. Il vettore si rappresenta come una freccia che va dal primo estremo all'altro e si denota $\overrightarrow{AB}$ oppure assegnandogli un nome, ad esempio v.

Definiremo ora delle **operazioni** sull'insieme dei vettori che hanno primo estremo in uno stesso punto A dato. Si dice che tali vettori sono **vettori applicati** in A.



Dati i vettori $u = \overrightarrow{AB}$ e $v = \overrightarrow{AC}$, che sono entrambi applicati in A,

si dice somma di u e v, e si denota

$u + v$, il vettore $w = \overrightarrow{AD}$,

dove il punto D è il vertice opposto ad A del parallelogramma $ABDC$, che ha i segmenti AB e AC come lati.

Per trovare la somma $u + v$ occorre costruire il parallelogramma.

Il parallelogramma si può costruire in diversi modi. Ad esempio si può costruire il vettore $\overrightarrow{CD}$, congruente a u, parallelo a u e applicato in C. In questo modo si è trovato il punto D e si può dimostrare che il vettore $\overrightarrow{BD}$ è congruente e parallelo a v.

Il problema sostanziale è la **costruzione della retta parallela a una retta data**, per un punto esterno a quest'ultima.

La cosa didatticamente interessante è **come fare la costruzione** o, meglio, **come farla scoprire agli studenti**.

Tali costruzioni si possono fare con la riga (non graduata!) e il compasso, che riassume in sé due assiomi di Euclide:

- ➔ per due punti si può tirare un unico segmento di retta, che si può estendere quanto si vuole da entrambe le parti
- ➔ dato un punto e un raggio c'è un cerchio che ha quel raggio e centro nel punto dato

Ma le costruzioni con la riga e il compasso risultano spesso ostiche per gli studenti, i quali:

- fanno fatica a usare il compasso, che richiede una certa abilità manuale, alla quale non sono abituati
- spesso non hanno la possibilità di usare un compasso decente;
- soprattutto non ne comprendono il senso.

MOLTO più efficace è fare le costruzioni in cortile o in atrio o in palestra, utilizzando una corda tesa e creando gruppetti in competizione, che discutono se le operazioni che stanno facendo producono o no i risultati desiderati. Gli studenti apprezzano la situazione insolita, fanno un po' di confusione, si coinvolgono, si divertono e imparano pure qualcosa. POI si può usare il compasso.

Si possono inoltre costruire i **multipli**

$$k \cdot u$$

e i **sottomultipli**

$$\frac{1}{k}u$$

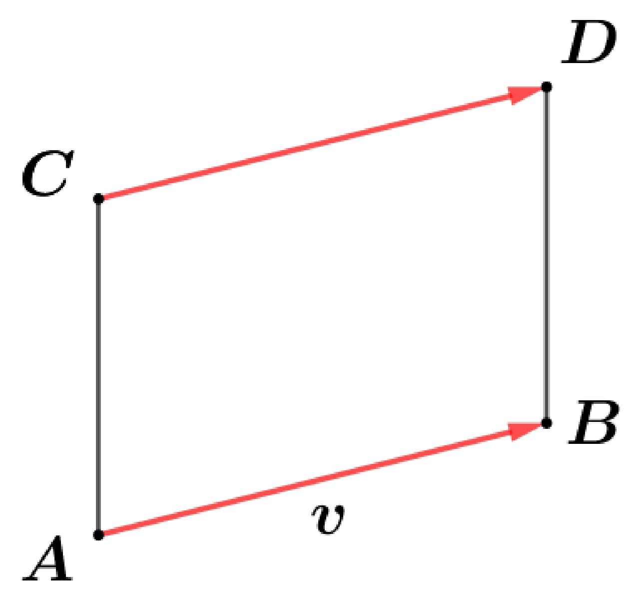
di un vettore u, dove k è un numero naturale;
e quindi si possono costruire i vettori del tipo

$$a \cdot u$$

dove a è una frazione.

Il prodotto di un vettore per un numero si può infine estendere ai numeri reali.

Mediante il concetto di vettore si può definire il concetto di traslazione. Infatti, dato un vettore $v = \overrightarrow{AB}$ e dato un punto qualsiasi C possiamo considerare il vettore $\overrightarrow{CD}$, tale che il quadrilatero $ABDC$ è un parallelogramma.



Il vettore $\overrightarrow{CD}$ si chiama **vettore parallelo a v con primo estremo in** C. In questo modo, a ogni punto C del piano viene associato uno e un solo punto D, che si chiama **traslato di C, mediante v** e si può denotare $T_v(C)$.

Così, per ogni fissato v si definisce la **traslazione di vettore** v

$$T_v : C \to T_v(C)$$

che è una funzione del piano in sé stesso. Ovviamente, due vettori paralleli definiscono la stessa traslazione.

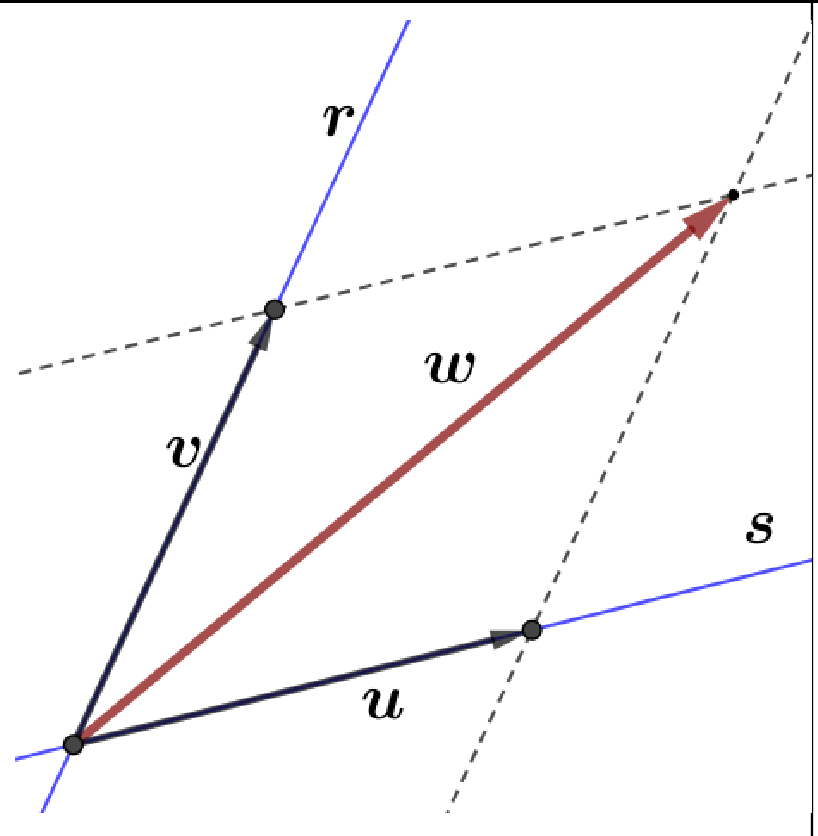
Comunque presi due vettori u e v si ha

$$T_u(T_v(A)) = T_{u+v}(A) = T_v(T_u(A))$$

per ogni punto A del piano.

Come si estendono tutte le costruzioni e definizioni precedenti allo spazio?

Un problema da proporre presto agli studenti in geometria euclidea è quello della **decomposizione di un vettore come somma di vettori in direzioni assegnate.**



Il problema è: date due rette r ed s, e dato un vettore w, trovare un vettore u sulla retta s e un vettore v sulla retta r, tali che $w = u + v$

Il problema si risolve costruendo le parallele ad r e ad s che passano per il secondo estremo del vettore w.

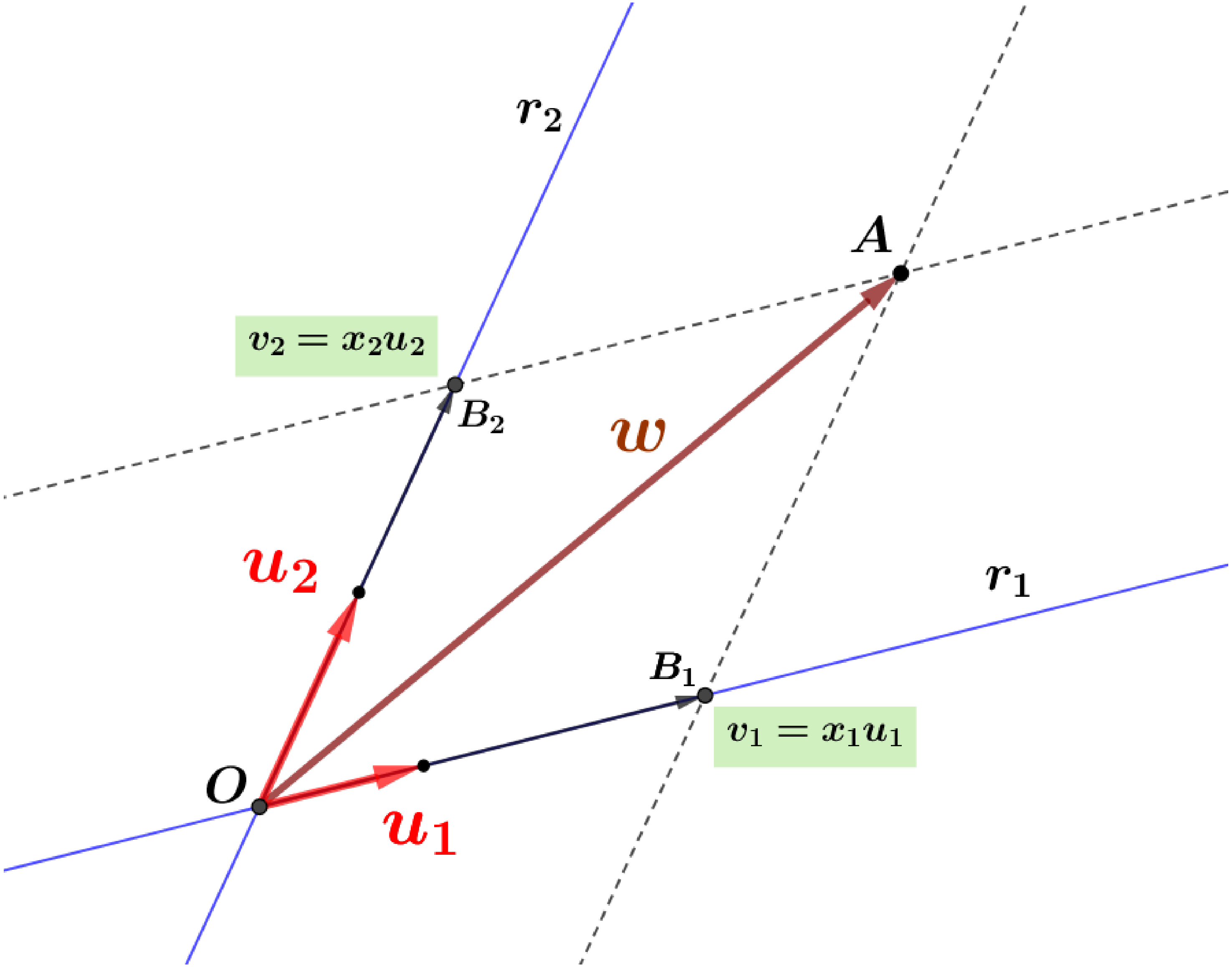
La decomposizione è importante per lo sviluppo del linguaggio vettoriale, per l'interpretazione algebrica, e per le applicazioni fisiche (ad esempio alle forze e alla velocità).
Anche per questo problema è necessaria la costruzione di una retta passante per un punto e parallela a una retta data.
A mio parere, inizialmente, per comprendere i concetti, conviene fare tale costruzione con una corda tesa o col compasso anche se ciò richiede un po' di lavoro.
In un secondo momento si possono usare una riga e una squadra, oppure Geogebra. Se invece si usano subito questi strumenti c'è il rischio che si producano delle misconcezioni (ad esempio che i concetti di *perpendicolarità* e di *distanza* siano vicini al concetto di *parallelismo*, cosa invece falsa).

Basandosi sulla decomposizione di un vettore lungo due direzioni, si può affrontare il problema di **scrivere un vettore come combinazione lineare di due vettori dati.**

Nella figura si vedono i vettori $w = \overrightarrow{OA}$ u_1 u_2
Si vuole ottenere w come combinazione lineare di u_1 e u_2, ossia si vuole trovare una coppia ordinata di numeri (x_1, x_2) tali che

$$w = x_1 u_1 + x_2 u_2$$

La costruzione geometrica che risolve il problema si vede in figura:
si costruisce la retta per A parallela a r_1 e si trova il punto B_1;
si costruisce la retta per A parallela a r_2 e si trova il punto B_2;


r_2
A
$v_2 = x_2 u_2$
B_2
w
u_2
r_1
B_1
$v_1 = x_1 u_1$
O
u_1

si prendono poi i vettori $v_1 = \overrightarrow{OB_1}$ e $v_2 = \overrightarrow{OB_2}$ e si ha

$$w = v_1 + v_2$$

A questo punto si trova il rapporto x_1 di v_1 con u_1 , ossia quel numero x_1 tale che $v_1 = x_1 u_1$; la stessa cosa si fa per trovare x_2 tale che $v_2 = x_2 u_2$. In conclusione si è ottenuto

$$w = v_1 + v_2 = x_1 u_1 + x_2 u_2$$

Il rapporto tra i segmenti si fa con l'algoritmo di Euclide, che è molto importante comprendere in questo contesto, prima che nei numeri interi, ma per cominciare gli studenti possono misurare le lunghezze dei segmenti con la riga.

Si è visto così che,

SE u_1 e u_2 sono **vettori nel piano**, sono applicati in un punto O, e **non sono allineati,**

ALLORA ogni vettore w del piano si può scrivere in uno e in un solo modo come combinazione lineare dei primi due, ossia esiste una e una sola coppia di numeri (x_1, x_2) tali che

$$w = x_1 u_1 + x_2 u_2$$

In questa situazione si dice che

la coppia di vettori u_1 , u_2 è un **sistema di riferimento** nel piano, applicato in O;
i numeri x_1 e x_2 si chiamano **coordinate del punto A rispetto al sistema di riferimento** u_1 e u_2.

È molto importante che i sistemi di riferimento e le coordinate siano introdotti mediante il parallelismo e senza usare il concetto di ortogonalità, che non c’entra nulla.

Oltre a evitare misconcezioni, questo consente di capire l’assonometria, che si usa nel disegno architettonico o meccanico per rappresentare sul piano oggetti tridimensionali. A questo punto si aprirebbe un’interessante discussione sull’intreccio tra l’apprendimento della matematica e quello del disegno.

5.4. La Geometria analitica

la **geometria analitica**

nasce nel seicento col **metodo delle coordinate** **(Cartesio)**

e consente di

- ➔ **usare l'algebra per descrivere le figure e le trasformazioni geometriche**
- ➔ **usare la geometria per dare significati alle equazioni e alle strutture algebriche**

Lo studio del **piano cartesiano** fa **parte del curricolo standard** di tutte le scuole secondarie di primo e secondo grado.

Si studia invece raramente la geometria analitica dello spazio

In effetti si capisce che nello spazio ci sono delle difficoltà maggiori, a cominciare da quelle di rappresentazione. Eppure, ce ne saebbe bisogno, sia per il disegno, sia per la fisica del moto. E anzi, sarebbe naturale estendere il **metodo delle coordinate allo spazio n dimensionale**; ad esempio per descrivere i fenomeni naturali e sociali, o i sistemi con diversi gradi di libertà (6 gradi per una matita, 3N per un sistema di N punti, quanti per una bicicletta? quanti per una corda? quanti per i clienti di Amazon? quanti per descrivere come si ripiegano le proteine?)

Quindi è necessaria una **“geometria analitica”, che mantenga l’obiettivo di intrecciare l’algebra e la geometria, negli spazi di dimensione n > 2, anche di dimensione infinita.**

Per fare questo **occorre abbandonare alcune delle tecniche che si usano e delle abitudini che si hanno nella geometria analitica piana, perchè non funzionano più, o sono maledettamente complicate, nello spazio n dimensionale,** già per n=3.

in effetti conviene pensare che

la Geometria analitica è il linguaggio geometrico dell'Algebra lineare, ovvero il linguaggio algebrico della Geometria

e si fonda sul concetti di

vettore, visto sia geometricamente, sia come lista, o n-upla ordinata;

spazio di vettori e sottospazi (rette, piani, ecc…);

operatore lineare e sua rappresentazione come **matrice**

interessante e piacevole è la storia del concetto di vettore e del calcolo differenziale vettoriale che viene sintetizzata in questa conferenza

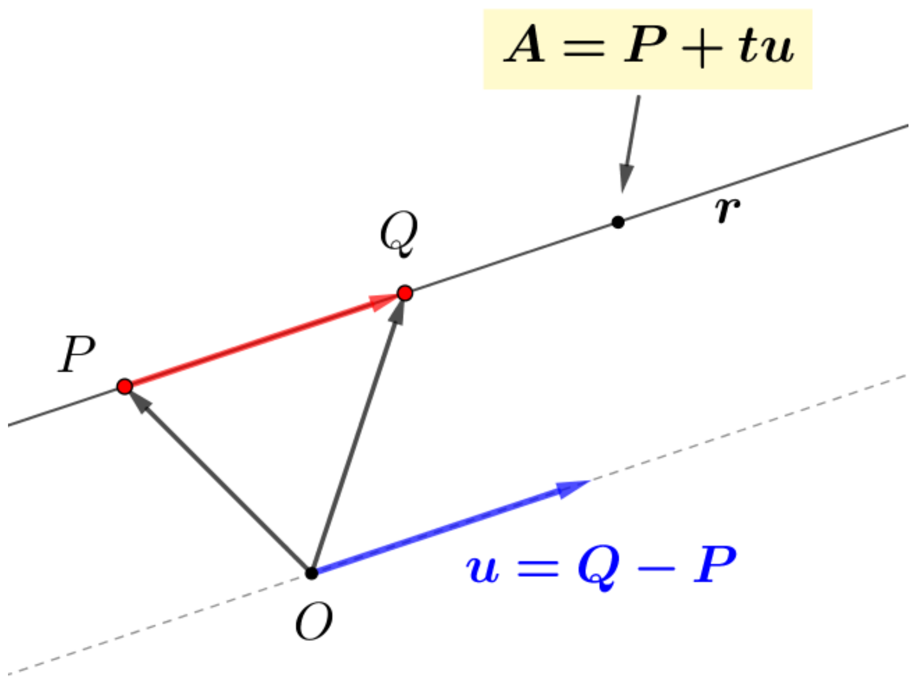
Michael J. Crowe, A History of Vector Analysis (talk at University of Louisville, 2002)

I concetti dell'algebra lineare e la loro interpretazione geometrica sono comodissimi e utili già in dimensione bassa, n=2 e n=3.

Vale quindi la pena di introdurli presto a scuola, almeno a partire dalla prima superiore, come strumenti per la descrizione e la modellizzazione delle rette, dei piani e dello spazio, delle figure geometriche e delle trasformazioni, del movimento, della statica. Temo però che gli insegnanti siano restii a farlo, forse perché negli studi universitari hanno visto questi concetti trattati dai matematici in modo troppo astratto e dai fisici o dagli ingegneri in modo troppo trasandato.

Eppure sono davvero notevoli. Facciamo qualche esempio.

Equazione di una retta



Dati due punti P e Q scriviamo un'**equazione della retta r per P e Q**. Fissato un punto O, ogni punto P si può identificare col vettore $\overrightarrow{OP}$. Grazie a questo si definiscono la *somma di punti* e il *prodotto di un punto per un numero*.

I punti della retta sono dunque tutti e soli

i punti del tipo $P + tu$

dove $u = Q - P$

al variare di t **nei numeri reali.**

Esattamente la stessa equazione si può scrivere in dimensione n.

Parametrizzazione di un segmento

Il segmento S di estremi P, Q è l'insieme dei punti

$$\Big\{ P + t(P - Q) \mid t \in [0,1] \Big\}$$

la funzione $\gamma : t \to P + t(Q - P)$, che manda l'intervallo [0,1] in S, si chiama parametrizzazione del segmento. Si ha

$\gamma(0) = P;$ $\qquad \gamma(1) = Q \qquad$ $\gamma(\frac{1}{2}) =$ punto medio di S.

Equazione di un piano.

Tre punti P_0, P_1, P_2 nello spazio 3-dimensionale, che non sono allineati, determinano un piano.
I punti di questo piano sono tutti e soli i punti del tipo

$$P_0 + t(P_1 - P_0) + s(P_2 - P_0)$$

al variare dei coefficienti t, s in $\mathbf{R}$.

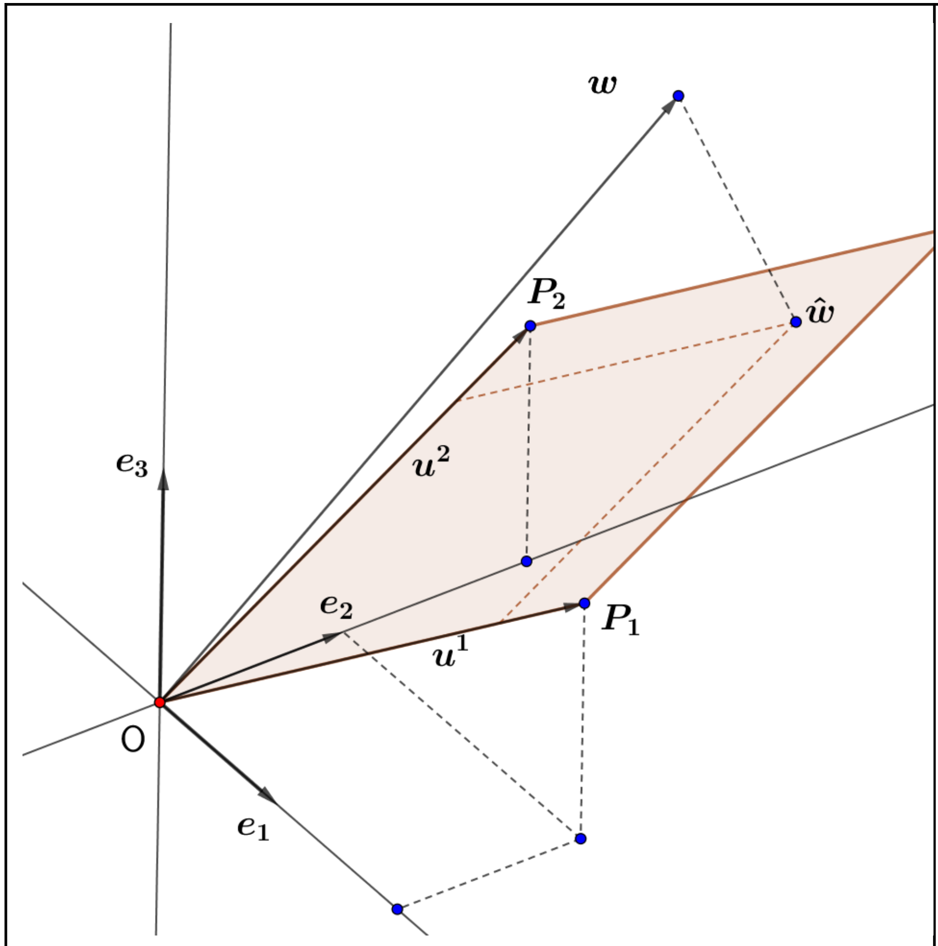
La stessa equazione vale se i tre punti sono nello spazio n-dimensionale, e anche in dimensione infinita.

Interpretazione geometrica di un sistema lineare di tre equazioni in due incognite.

Consideriamo il sistema sotto a sinistra

$$\begin{cases} 2x_1 + 0 \cdot x_2 = 1,35 \\ x_1 + 2x_2 = 1,9 \\ x_1 + x_2 = 2,6 \end{cases} \qquad x_1 \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1,35 \\ 1,9 \\ 2,6 \end{bmatrix}$$

Sono tre equazioni nelle due incognite x_1, x_2.
Si vede immediatamente che non ha soluzioni.
Ma scriviamolo come a destra ossia nella forma $x_1 u_1 + x_2 u_2 = w$
e guardiamo cosa significa geometricamente.



$$u_1 = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} \quad u_2 = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} \quad w = \begin{bmatrix} 1,35 \\ 1,9 \\ 2,6 \end{bmatrix}$$

il sistema chiede se è possibile trovare una coppia di numeri x_1, x_2 tali che il vettore w dei termini noti sia combinazione lineare dei vettori u_1, u_2 con coefficienti x_1, x_2

non è possibile perché w non sta sul piano M generato dai vettori u_1, u_2

però si può trovare un punto $\hat{w}$ appartenente al piano, che ha distanza minima da w.

Questo $\hat{w}$ è la migliore approssimazione possibile di una soluzione che non esiste.

Questa “soluzione” di un sistema lineare che non ha soluzione si chiama soluzione ai minimi quadrati https://en.wikipedia.org/wiki/Least_squares

Il problema della regressione, ossia di far passare una retta per più di due punti, porta a un sistema lineare del tipo di quello visto sopra.

Far passare un piano o una parabola per quattro punti porta a un sistema di quattro equazioni in tre incognite, che si interpreta come il problema di scrivere un certo vettore di R^4 come combinazione lineare di tre vettori.

Naturalmente in quattro dimensioni non ci vediamo più, ma le rappresentazioni grafiche e la visualizzazione mentale che impariamo ad avere nei casi semplici in due e tre dimensioni ci aiuta a comprendere come muoverci e che tipo di comportamenti aspettarci nei calcoli algebrici in dimensione alta.

Ci sarebbero moltissime altre bellissime idee di cui si potrebbero a scuola sviluppare i primi elementi, utilizzando la Geometria analitica i vettori e l'algebra lineare, con e senza coordinate:

come fare con Geogebra l'ultima figura qui sopra (che si può muovere e guardare da qualsiasi direzione)

baricentri e mediane nel triangolo e nel tetraedro

il prodotto scalare: distanze, angoli, aree e volumi k dimensionali negli spazi a n-dimensioni

isometrie, simmetrie, proiezioni ortogonali e altre trasformazioni

ellissi, ellissoidi, autovalori

analisi delle componenti principali

Vedremo se ci saranno occasioni di parlarne.

Grazie per l'attenzione