

Dimensione Cognitiva

4. Come le macchine comprendono il linguaggio

Embeddings e Spazi Semantici

Giovanni Della Lunga
giovanni.dellalunga@unibo.it

A lezione di Intelligenza Artificiale

Siena - Giugno 2025

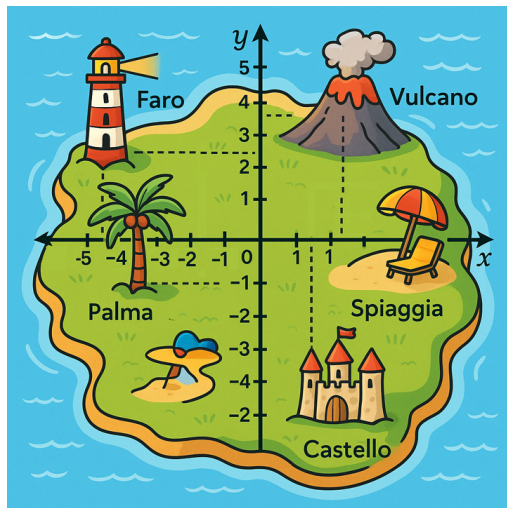
- 1 Parliamo (ancora) di Geometria
- 2 Passare dalle parole ai ... Numeri!
- 3 Come si generano gli embeddings?

Parliamo (ancora) di Geometria

Che cos'è un Vettore

- Immaginiamo di trovarci su un'isola immaginaria.
- Su quest'isola ci sono varie attrazioni turistiche: un faro, un vulcano, una palma, una spiaggia, un castello.
- Sono luoghi ben precisi, visibili e facilmente riconoscibili.
- Ora proviamo a fare un piccolo sforzo mentale: vogliamo **descrivere la posizione** di ciascuna attrazione in un modo chiaro, condivisibile e, soprattutto, **calcolabile**.

Che cos'è un Vettore



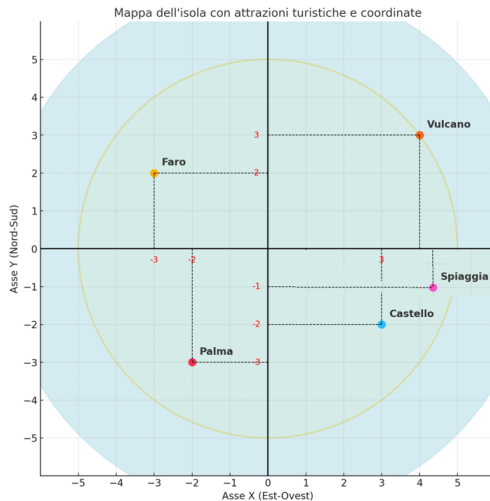
Dalle immagini alle coordinate

Nel primo disegno, notiamo che l'isola è attraversata da due linee perpendicolari: una orizzontale (asse X, Est-Ovest) e una verticale (asse Y, Nord-Sud). Questo sistema di assi divide l'isola in quattro quadranti, e permette di assegnare a ogni luogo una **coppia di numeri**: le sue **coordinate**.

Ad esempio:

- Il **Faro** si trova a sinistra e un po' in alto, in posizione $(-3, 2)$
- La **Spiaggia** è sulla destra e un po' in basso, in $(4, -1)$
- Il **Castello** è a destra e in basso, in $(3, -2)$

Che cos'è un Vettore



Dalle coordinate al concetto di vettore

- Finora abbiamo parlato di "punti" su una mappa. Ma se ora immaginiamo di voler **muoverci** da un punto all'altro — ad esempio dal faro al vulcano — abbiamo bisogno di qualcosa che non dica solo **dove siamo**, ma anche **come spostarci**.
- Qui entra in gioco il **vettore**.
- Un **vettore** è una **freccia che unisce due punti**, indicando:
 1. **direzione** (verso dove andare)
 2. **verso** (da dove a dove)
 3. **lunghezza** (quanto spostarsi)

Ad esempio, il vettore che va dal Faro al Vulcano dice: "spostati di 7 unità verso Est (da -4 a 3) e di 1 unità verso Nord (da 2 a 3)". Questo lo possiamo scrivere come il vettore **(7, 1)**.

Il salto concettuale: spazio vettoriale

Quando iniziamo a lavorare con vettori, ci accorgiamo che hanno delle proprietà molto interessanti:

- Possiamo **sommarli** (fare un tragitto in due tappe)
- Possiamo **scalarli** (fare lo stesso tragitto ma due volte più lungo)
- Possiamo **confrontare** vettori che hanno la stessa direzione, verso e lunghezza, anche se partono da punti diversi

Il salto concettuale: spazio vettoriale

- Tutto questo avviene in quello che in matematica si chiama **spazio vettoriale**: un insieme di vettori su cui è possibile fare queste operazioni in modo coerente.
- Nel nostro caso, lo spazio è il piano delle coordinate (X, Y) , dove ogni vettore è rappresentato da una coppia di numeri.
- È come un linguaggio universale per descrivere spostamenti e posizioni.

- Ogni attrazione turistica dell'isola ha una **posizione**, data da una **coppia di coordinate** (x, y)
- Ogni **spostamento** da un punto a un altro può essere rappresentato come un **vettore**
- I vettori hanno direzione, verso e lunghezza
- Insieme, i vettori formano uno **spazio vettoriale**, una struttura matematica che ci permette di analizzare e combinare movimenti e relazioni spaziali

Misurare caratteristiche non spaziali

Abbiamo compreso che un vettore nello spazio tridimensionale è rappresentato da tre numeri (le coordinate cartesiane x , y , z). Ad esempio:

- La posizione di una città può essere identificata da tre coordinate geografiche.
- Un punto in una stanza può essere identificato da tre coordinate (lunghezza, larghezza, altezza).

In questi casi le coordinate rappresentano letteralmente una posizione nello spazio fisico, quello a cui siamo abituati.

Un altro passo verso l'astrazione

Misurare caratteristiche non spaziali

Adesso facciamo un piccolo passo verso l'astrazione, usando sempre dei vettori con coordinate numeriche, ma stavolta non riferite a posizioni nello spazio. Per esempio, immaginiamo di voler rappresentare delle persone attraverso alcune caratteristiche misurabili numericamente:

Persona \rightarrow (età, peso, altezza)

In questo modo, una persona può essere vista come un "punto" in uno spazio astratto definito da queste tre dimensioni: età, peso e altezza. Ad esempio:

- Mario: (30 anni, 75 kg, 180 cm)
- Lucia: (28 anni, 62 kg, 165 cm)

Misurare caratteristiche non spaziali

- In questo "spazio delle persone", due individui con caratteristiche simili (ad esempio età simile, altezza simile) saranno due "punti" vicini tra loro.
- Notiamo che, anche se parliamo ancora di numeri semplici, qui abbiamo fatto un primo passo verso l'astrazione: non siamo più nello spazio fisico, ma in uno spazio di **caratteristiche numeriche**.

Includere caratteristiche non numeriche (qualitative)

- Ora facciamo un altro piccolo passo avanti: aggiungiamo caratteristiche non immediatamente numeriche ma che possiamo rappresentare numericamente.
- Immaginiamo ad esempio di voler rappresentare dei film tramite vettori numerici, usando alcune caratteristiche come:
 - Quanto è comico (da 0 a 10)
 - Quanto è drammatico (da 0 a 10)
 - Quanto è romantico (da 0 a 10)

Un ulteriore passo verso l'astrazione

Includere caratteristiche non numeriche (qualitative)

Ogni film potrebbe essere descritto da un vettore di tre numeri che rappresentano intensità di caratteristiche qualitative:

- Film A (molto comico, poco drammatico, medio romantico): (8, 2, 5)
- Film B (poco comico, molto drammatico, molto romantico): (1, 9, 8)

In questo **spazio astratto**, due film simili si troveranno "vicini" tra loro, proprio come due città vicine in una mappa, anche se lo spazio non ha più nulla a che vedere con posizioni fisiche.

Rappresentare concetti astratti come le parole

Finalmente, siamo pronti per il salto finale, quello più astratto di tutti:

*E se provassimo a rappresentare **il significato delle parole** con numeri?*

L'idea del word embedding nasce proprio qui: rappresentare una parola come una lista di numeri (un vettore), ciascuno dei quali esprime quanto la parola è associata a concetti o contesti particolari.

Rappresentare concetti astratti come le parole

Ad esempio, per le parole "gatto", "cane" e "automobile", un modello intelligente potrebbe assegnare coordinate numeriche in modo che:

- "gatto" = (8, 9, 1, ...)
- "cane" = (7.5, 9.2, 1.5, ...)
- "automobile" = (0.5, 1, 9.5, ...)

Rappresentare concetti astratti come le parole

Anche se queste coordinate non sono più interpretabili facilmente una ad una (sono prodotte automaticamente dai modelli di AI), si mantengono due proprietà importanti

- Parole con significati simili hanno vettori vicini nello spazio astratto.
- Parole con significati molto diversi hanno vettori lontani.

Quindi, come nello spazio ordinario:

- Vicinanza geografica → città vicine.
- Vicinanza astratta → significati vicini.

Riassunto del percorso fatto

Livello	Esempio	Componenti del vettore
1 Ordinario	Posizione geografica	Coordinate spaziali (x,y,z)
2 Prima astrazione	Persone	Caratteristiche numeriche (età, peso, altezza)
3 Seconda astrazione	Film	Caratteristiche qualitative numeriche (comico, romantico, drammatico)
4 Totale astrazione	Parole (Word embedding)	Coordinate semantiche automatiche

Passare dalle parole ai ... Numeri!

Il modello Bag of Words

- Immaginiamo che la nostra macchina capisca solo un piccolo dizionario di poche parole. Per costruire i vettori che rappresentano le frasi date utilizzando il metodo **Bag of Words**, inizieremo creando un dizionario con questi termini.
- Nella **Bag of Words** ogni frase è trasformata in un vettore, lungo esattamente come il vocabolario, nel quale ad ogni elemento corrisponde una parola del dizionario.

VETTORE

Indice:

1

2

3

4

5

6

7

8

9

10

amo	casa	il	ma	non	sole	stare	viaggiare	volare	vorrei

Il modello Bag of Words

- Il valore di ciascun elemento del vettore verrà calcolato contando la frequenza di ciascuna parola. Se una parola del dizionario non appare nella frase il suo conteggio sarà 0. Se appare una o più volte, il valore sarà pari a 1 o al numero totale di volte che la parola appare nella frase.

FRASE: amo viaggiare ma non amo volare

VETTORE

Indice:

1 2 3 4 5 6 7 8 9 10

amo	casa	il	ma	non	sole	stare	viaggiare	volare	vorrei
2	0	0	1	1	0	0	1	1	0

FRASE: vorrei stare a casa

VETTORE

Indice:

1 2 3 4 5 6 7 8 9 10

amo	casa	il	ma	non	sole	stare	viaggiare	volare	vorrei
0	1	0	0	0	0	1	0	0	1

Il modello Bag of Words

Sembra funzionare, ogni frase ha il suo vettore. Ma quando proviamo frasi più complesse come le seguenti

FRASE: amo viaggiare, vorrei non stare a casa

VETTORE

Indice:

1 2 3 4 5 6 7 8 9 10

amo	casa	il	ma	non	sole	stare	viaggiare	volare	vorrei
1	1	0	0	1	0	1	1	0	1

FRASE: non amo viaggiare, vorrei stare a casa

VETTORE

Indice:

1 2 3 4 5 6 7 8 9 10

amo	casa	il	ma	non	sole	stare	viaggiare	volare	vorrei
1	1	0	0	1	0	1	1	0	1

Il modello Bag of Words

- Ci accorgiamo che entrambe le frasi hanno la stessa rappresentazione vettoriale.
- Questo significa che, secondo questo modello, le frasi avranno il medesimo significato!
- Tuttavia le due frasi hanno significati opposti e questo dimostra chiaramente uno dei principali limiti di questo metodo nell'analisi del linguaggio naturale.

- Coscienti di questi limiti i ricercatori hanno sviluppato modelli che partono dall'assunto che le parole che appaiono frequentemente vicine le une alle altre hanno significati più strettamente collegati.
- Quindi l'ipotesi è che analizzando le parole all'interno di una certa finestra di contesto attorno ad una parola target il modello possa apprendere rappresentazioni più precise.

"Era una buia e tempestosa"

Cos'è un word embedding?

- Un **word embedding** è un modo per rappresentare una parola come una **sequenza di numeri** (un vettore), così che i computer possano lavorare con le parole in modo simile a come fanno con i numeri.
- Ma non sono numeri qualsiasi.
- A differenza del modello BoW, ogni numero nel vettore **porta con sé un significato**, perché è stato appreso analizzando milioni di frasi e testi.
- In altre parole, i vettori **catturano il significato** delle parole basandosi su come queste vengono usate nel linguaggio.

Cosa vuol dire “dimensione” in un word embedding?

- Immaginate che ogni parola venga trasformata in un vettore con, ad esempio, **300 numeri**.
- Ogni numero è una coordinata lungo una particolare **dimensione** dell'embedding.

Ora: queste **dimensioni non sono etichette esplicite**, ma rappresentano **sfumature di significato** che l'algoritmo ha imparato da solo. Alcune di queste dimensioni possono (in modo implicito) rappresentare:

- il concetto di **maschile vs femminile**
- il grado di **astrazione** di una parola
- la sua **carica emotiva** (positiva/negativa)
- il legame con **luoghi** o **tempi**
- la **categoria grammaticale** (sostantivo, verbo...)

Cosa si intende per "dimensione semantica"?

Una **dimensione semantica** è quindi **una direzione nello spazio dei significati**, lungo la quale possiamo cogliere un cambiamento semantico specifico.

Un esempio molto noto:

- Prendiamo i vettori delle parole "re" (king), "regina" (queen), "uomo" (man) e "donna" (woman)
- **re - uomo + donna = regina**
- In questo caso, la **differenza tra "re" e "uomo"** può essere interpretata come una **dimensione semantica di regalità**, e la **differenza tra "uomo" e "donna"** come una **dimensione semantica di genere**.

Un'analogia visiva

Immaginate uno spazio tridimensionale:

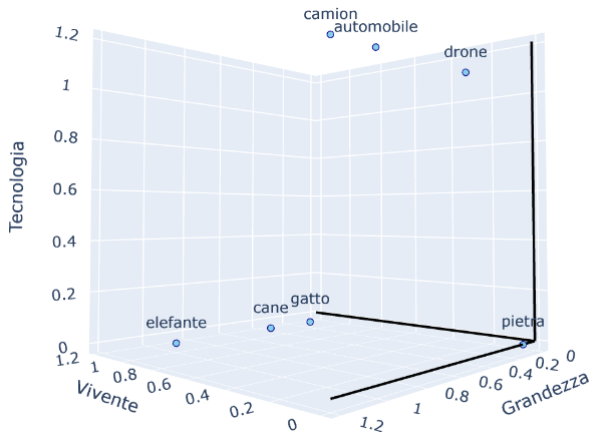
- L'asse X potrebbe rappresentare **il grado di positività** della parola
- L'asse Y il **campo semantico** (es. "cibo", "emozione", "luogo")
- L'asse Z il **livello di concretezza** (oggetti tangibili vs concetti astratti)

Una parola come **"cioccolato"** potrebbe avere coordinate (8, 2, 9), mentre **"libertà"** potrebbe stare a (5, 8, 1).

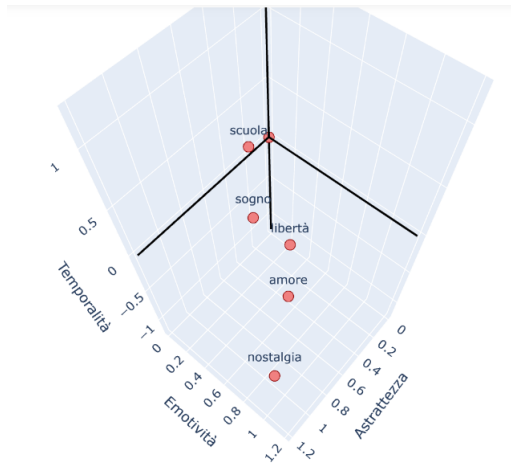
In un vero word embedding ci sono **molte più dimensioni** (50, 100, 300...), ma il concetto è lo stesso: ogni direzione rappresenta **una possibile variazione di significato**.

- Una **dimensione semantica** in un word embedding è una direzione nello spazio matematico che riflette **una caratteristica latente del significato** delle parole, appresa dai dati.
- Non sempre sappiamo **cosa** rappresenta esattamente ogni dimensione, ma possiamo studiarle osservando come le parole si posizionano e si muovono nello spazio.

Un'analogia visiva



Un'analogia visiva



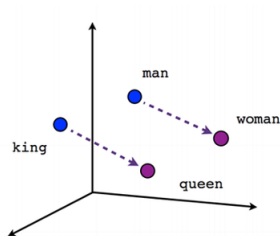
Cosa rappresenta questa visualizzazione?

- Le **parole sono punti** in uno spazio dove la distanza e la posizione **riflettono caratteristiche semantiche**.
- Se due parole sono vicine, significa che **hanno significati simili** lungo le dimensioni scelte.
- Se sono lontane, rappresentano **concetti diversi**.

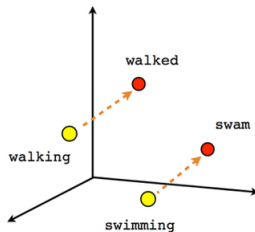
Esempi interpretativi

- "gatto" e "cane" sono vicini → entrambi viventi, taglia simile
- "camion" e "automobile" sono vicini → non viventi, grandi, tecnologici
- "pietra" è isolata → non è vivente, non è tecnologica, è piccola
- "drone" è vicino alle auto → tecnologico, ma più piccolo

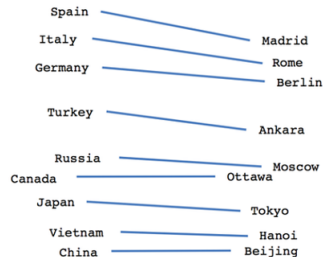
Addizione e Sottrazione di Vettori



Male-Female



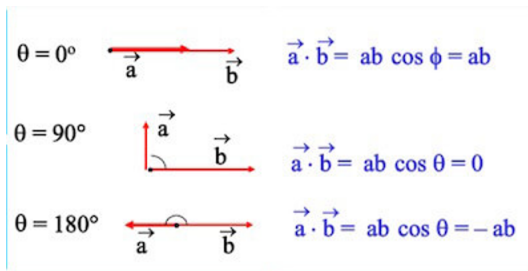
Verb tense



Country-Capital

Prodotto di Vettori: $a \cdot b = a_1b_1 + a_2b_2 + \dots + a_nb_n$

- Il prodotto fra vettori può essere in qualche modo relazionato al concetto di similarità.
- Geometricamente il prodotto scalare misura quanto due vettori sono **allineati**.
- Un prodotto scalare pari a zero indica che i vettori sono perpendicolari mentre un valore maggiore di zero indica un certo grado di allineamento nella stessa direzione e un numero minore di zero indica che sono più o meno allineati ma puntano in direzioni opposte.



Come si generano gli embeddings?

- Un modello come **word2vec** impara gli embeddings (rappresentazioni numeriche delle parole) basandosi su un principio semplice:

parole simili appaiono in contesti simili.

- Per esempio, le parole "gatto" e "cane" spesso compaiono vicine a parole come "animale", "cibo" o "giocare", mentre "auto" e "bicicletta" appaiono vicine a parole come "guidare", "strada", "velocità".

Ecco, passo per passo, come funziona in maniera molto semplificata:

1. **Scansione di un testo:** Il modello legge tantissime frasi e osserva quali parole si trovano spesso vicine tra loro.
2. **Addestramento:** Il modello cerca di prevedere, data una parola, quali parole tendono ad apparire vicine. Ad esempio, data la parola "gatto", il modello cerca di indovinare parole come "miagolare" o "croccantini".
3. **Conversione in numeri (embeddings):** Durante questo allenamento, il modello assegna automaticamente a ogni parola dei valori numerici (i vettori di embedding appunto) che cercano di catturare queste relazioni facendo in modo che parole con significati o usi simili abbiano embeddings numerici molto simili tra loro.

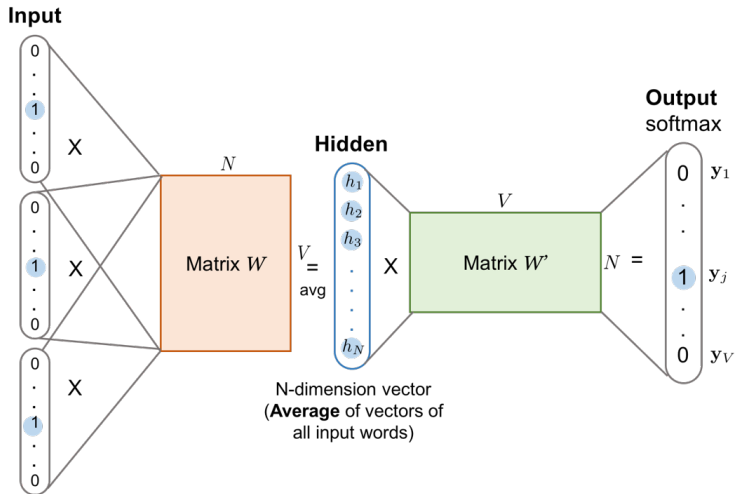
Il modello **word2vec** assegna a ciascuna parola dei valori numerici (embeddings) attraverso una procedura molto semplice, che si basa su tentativi ed errori e piccoli aggiustamenti continui:

- 1. **Inizia in modo casuale**. All'inizio, ogni parola viene associata a una sequenza di numeri casuali.
- 2. **Gioco di "indovina la parola"**. Il modello prende una parola dal testo, ad esempio "gatto", e prova a prevedere quali altre parole spesso compaiono vicino ad essa (ad esempio "mangia", "gioca", "miagola"). Se il modello riesce a prevedere correttamente le parole vicine, allora i numeri (embeddings) assegnati sono buoni. Se sbaglia, vuol dire che i numeri devono essere modificati.

- 3. **Impara dagli errori.** Quando il modello sbaglia, cambia leggermente i valori numerici delle parole coinvolte per migliorare la previsione futura. Le parole che compaiono spesso insieme nel testo, progressivamente, avranno valori numerici sempre più simili tra loro.
- 4. **Ripeti molte volte.** Il modello ripete questo processo moltissime volte, su milioni di frasi e parole. A ogni ripetizione, i valori numerici si aggiustano leggermente fino a diventare sempre più precisi. Alla fine del processo, parole che appaiono in contesti simili finiscono per avere valori numerici molto vicini tra loro, parole con significati molto diversi avranno valori numerici più distanti.

Questo meccanismo è ciò che permette a **word2vec** di catturare e rappresentare numericamente i significati e le relazioni tra le parole.

Word2Vec



Come vengono utilizzati gli embeddings nel mondo dell'IA Generativa?

- ChatGPT usa gli embeddings per capire il significato delle parole.
- Parola per parola, frase per frase per poi generare del testo.
- Noi abbiamo parlato di embeddings semplici a poche dimensioni, abbiamo parlato di Word2Vec che usa 300 dimensioni, ma giusto per darvi un'idea gli **embeddings** usati da ChatGPT usano **12888 dimensioni!**