PRINCIPAL COMPONENT ANALYSIS

- $A = U\Lambda U^T$ diagonalizzazione
e le colonne $U^{(k)}$ sono gli autovalori
e i valori in Λ sono gli autovalori

- se U è ortonormale ($U^T U = I$)
otteniamo ~~autovalori~~ autovettori
ortogonali (e possiamo normalizzarli)

- se $B = PCP^{-1}$ allora B e C
sono SIMILI (se P è ortogonale
allora $P^{-1} = P^T$) quindi anche A e
Λ sono simili se U è ortogonale

STEP DELLA PCA

1) CENTRARE IL DATASET dato $X' \in \mathbb{R}^{n \times p}$

$$\underline{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \left[\frac{1}{n}\sum_{i=1}^{n} x_{i1}, \dots, \frac{1}{n}\sum_{i=1}^{n} x_{ip} \right]$$

$$\hookrightarrow \quad X = \begin{bmatrix} x_1' - \mu \\ \vdots \\ x_n' - \mu \end{bmatrix} \rightarrow$$

le features vengono centrate

> OBIETTIVO

Vogliamo ottenere una trasformazione

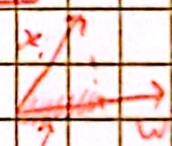
NUOVO DATASET $\rightarrow T = XW$

$\uparrow$ DATASET

Dove W è una matrice in cui ogni colonna è una "direzione" principale dei dati $\rightarrow$

$$w^{(1)} = \underset{w \in \mathbb{R}^p}{\arg\max} \left\{ \frac{1}{n-1} \sum_{i=1}^{n} \left(\underbrace{\sum_{j=1}^{p} x_{ij} w_j}_{x_i \cdot w} \right)^2 \right\}$$

con $\|w\| = 1$



$Var(\cancel{x_i} \cdot w)^2$ (?)

$$= \underset{w \in \mathbb{R}^p}{\arg\max} \|Xw\|^2 = \underset{w \in \mathbb{R}^p}{\arg\max}\ X^T w^T X w$$

con $\|w^{(1)}\| = 1$ con $\|w^{(1)}\| = 1$

$$= \boxed{\underset{w \in \mathbb{R}^p}{\arg\max} \frac{w^T X^T \cancel{X} w}{w^T w}} =$$

Autovettore con l'autovalore più grande di $X^T X$

Quindi scegliendo la direzione w spieghiamo una certa porzione della varianza.
Togliamo quest'ultima per rimanere con la varianza rimasta:

$$X_k = X - \underbrace{X \sum_{i=1}^{k-1} w^{(i)} (w^{(i)})^T}_{}$$

proiezione di X sulle direzioni trovate

Dove k è il numero di dimensioni trovate

X_k è la proiezione dei dati ~~[illegible]~~ sullo spazio perpendicolare alle direzioni trovate

Se ripetiamo la stessa procedura su X_1 otteniamo $w^{(2)}$, se ripetiamo su X_2 otteniamo $w^{(3)}$, ecc...

INTERPRETAZIONE

Possiamo calcolare la matrice di covarianza del dataset come:

$$C = \frac{1}{n-1} \sum_{i=1}^{n} x_i^T x_i = \frac{1}{n-1} X^T X$$

1) Dal teorema spettrale $Au^{(i)} = \lambda_i u^{(i)}$

$$X^T X \, w^{(1)} = \lambda_1 w^{(1)}$$

$$(w^1)^T X^T X \, w^1 = \lambda_1 w^1 \cdot (w^1)^T$$

$$(w^1)^T X^T X \, w^1 = \lambda_1 \|w^1\|^2 \qquad \|w^1\| = 1$$

$$\|X w^1\|^2 = \lambda_1$$

2) Dai ragionamenti precedenti

$$Var(X w^1) = \frac{1}{n-1} \|X w^1\|^2 = \frac{\lambda_1}{n-1}$$

Se facciamo la trasformazione ~~con~~ con la W ottenuta

$$T = XW$$

Otteniamo la nuova rappresentazione del dataset. Calcoliamo la matrice di covarianza

$$C' = \frac{T^T T}{n-1} = \frac{W^T X^T X W}{n-1} =$$

autovalori

autovettori

per il teorema spettrale $\underline{A = V \Lambda V^T}$

$$= \frac{W^T W \Lambda W^T W}{n-1} = \frac{\Lambda}{n-1} = \begin{pmatrix} Var_1 & 0 & 0 \\ 0 & Var_2 & 0 \\ & & \ddots \\ 0 & 0 & Var_p \end{pmatrix}$$

Quindi nella nuova rappresentazione le features sono indipendenti!

▶ RIDUZIONE DELLA DIMENSIONALITÀ

Possiamo scegliere solo alcune delle "direzioni principali" e spiegare solo una percentuale della varianza, ad esempio solo il 95%

⟼

Per scegliere usiamo un indice chiamato proporzione di varianza spiegata

$$PVE = \frac{\sum_{k=1}^{m} \left[\sum_{i=1}^{n} \frac{1}{n-1} \sum_{j=1}^{p} x_{ij} w_{jk} \right]}{\sum_{j=1}^{p} \frac{1}{n-1} \sum_{i=1}^{n} x_{ij}^2}$$

$$= \frac{\sum_{k=1}^{m} \lambda_k}{\sum_{j=1}^{p} \lambda_j}$$

NB: il numero di autovalori che possiamo scegliere come gli m principali dipendono dal rango di $X^T X$

$$m \leq \text{rango}\left[X^T X\right]$$

Singular Value Decomposition

Usare la diagonalizzazione per ~~fare~~ trovare gli autovalori funziona solo se la matrice é quadrata $\rightarrow A = U \Lambda U^T$

Se invece é rettangolare abbiamo che

$$M_{n \times p} = U_{n \times n} \cdot \Sigma_{n \times p} \cdot V^T_{p \times p}$$

SONO MATRICI ORTOGONALI

oppure, se rango$[M] < \min\{n, p\}$

$$M_{n \times p} = U_{n \times r} \cdot \Sigma_{r \times r} \cdot V^T_{r \times p}$$

Quindi $M = U \Sigma V^T$

Usiamo questa proprietà su $X^T X$

$$X^T X = V \Sigma^T U^T U \Sigma V^T = V \Sigma^T \Sigma V^T$$

Da cui possiamo dire che

$X^T X$ e $\Sigma^T \Sigma$ sono simili, cioé hanno gli stessi autovalori

Sappiamo peró che gli autovalori

di $X^T X$ sono proporzionali alla varianza di ogni feature e quindi gli autovettori associati sono le direzioni principali, lo stesso però possiamo dire di V!

Quindi

$$T = XV$$

Usando la SVD possiamo anche riscrivere

$$T = XV = U\Sigma V^T V = U\Sigma$$

Quindi possiamo ottenere il nuovo dataset da entrambi i singular vectors (U e V)