**Primjenjivost PCA na AE**

PCA je statistička metoda bez nadzora koja se koristi za reduciranje dimenzionalnosti skupa podataka. PCA transformira ulazne značajke u jednak broj novih značajki koje nazivamo principalne komponente. Prva principalna komponenta sadrži najviše varijance prisutne u originalnim značajkama. Nadalje svaka slijedeća principalna komponenta sadrži sve manje varijance iz originalnih značajki.

Pretpostavke primjene PCA:

1. Mora postojati linearnost u skupu podataka. Sve značajke moraju imati međusobno linearni odnos.

**Linearni odnos varijabli:** [2]

* Prikaz grafički kao pravac
* Jednadžba koja opisuje odnos može se sastojiti od samo 2 varijable: f(x) = ax + b
  + x i f(x) varijable u linearnom odnosu
  + a – koeficijent nagiba pravca, b – odsječak na osi
* Varijable su na prvu potenciju
* Korelacija dobro opisuje linearni odnos varijabli – što je udaljenija od nule je veći linearni odnos
* Linearna regresija je metoda koja otkriva linearne odnose između varijabli

**Ovaj uvijet nije rigorozan. Samo upozorava da će značajke s većom korelacijom povezati u jednu principalnu komponentu, dok će značajke s manje korelacije ostaviti zasebno tj. neće dobro razvrstati.**

1. Podrazumijeva da varijable s najviše varijance su najbitnije. Isto tako varijable s najviše korelacije s ostalim varijablama. Međutim jednako su bitne sve principalne komponente.
2. Imati puno podataka za svaku značajku.

**Proces dobivanja značajki:**

**Analiza matrica ekviliziranih značajki:**

[time series - Feature Selection: Correlation and Redundancy - Cross Validated (stackexchange.com)](https://stats.stackexchange.com/questions/242792/feature-selection-correlation-and-redundancy)

Prateći gore navedeni link zaključujemo da samo značajke koje koreliraju +1 ili -1 unose redundanciju. Ostale značajke su bitne za ispravno klasificiranje primjera. Moguće je izabrati niži prag korelacije za izbacivanje redundancije međutim potrebno ga je argumentirati.

**Primjena PCA:** Koristiti samo za potvrdu korelacijskih odnosa između značajki.

1. Potrebna normalizacija varijabli – predlaže se oduzimanje sa srednjom vrijednosti određene značajke: x -mean(x)
2. -

**Interpretacija rezultata** - [Interpret the key results for Principal Components Analysis - Minitab](https://support.minitab.com/en-us/minitab/18/help-and-how-to/modeling-statistics/multivariate/how-to/principal-components/interpret-the-results/key-results/)

[1] i [3]

[1] - [PCA in Machine Learning: Assumptions, Steps to Apply & Applications | upGrad blog](https://www.upgrad.com/blog/pca-in-machine-learning/)

[2] - [Linear Relationship Definition (investopedia.com)](https://www.investopedia.com/terms/l/linearrelationship.asp)

[3] - [How to perform a principal components analysis (PCA) in SPSS Statistics | Laerd Statistics](https://statistics.laerd.com/spss-tutorials/principal-components-analysis-pca-using-spss-statistics.php)

**Ostale metode selekcije značajki**

Slika na kojoj se prikazuje stol

Opis je automatski generiran

Vrste selekcija značajki ne označenog skupa podataka:

1. Filter metode – preprocessing
2. Wraper metode – izabere se skup značajki te se trenira na skup podataka (SU) i analizira za sve skupove značajki
   1. Komputacijski skupo
   2. Gledaš kako je dodavanje nove značajke utjecalo na rezultat clusteringa
3. Embedded metode

**DBSCAN**

[4] - [Sci-Hub | Feature Selection of Acoustic Emission Signal for the Slow-Speed and Heavy-Load Equipment. Applied Mechanics and Materials, 110-116, 3199–3203 | 10.4028/www.scientific.net/amm.110-116.3199](https://sci-hub.se/https:/www.scientific.net/AMM.110-116.3199)

[5] - [Feature Selection Techniques in Machine Learning (analyticsvidhya.com)](https://www.analyticsvidhya.com/blog/2020/10/feature-selection-techniques-in-machine-learning/)