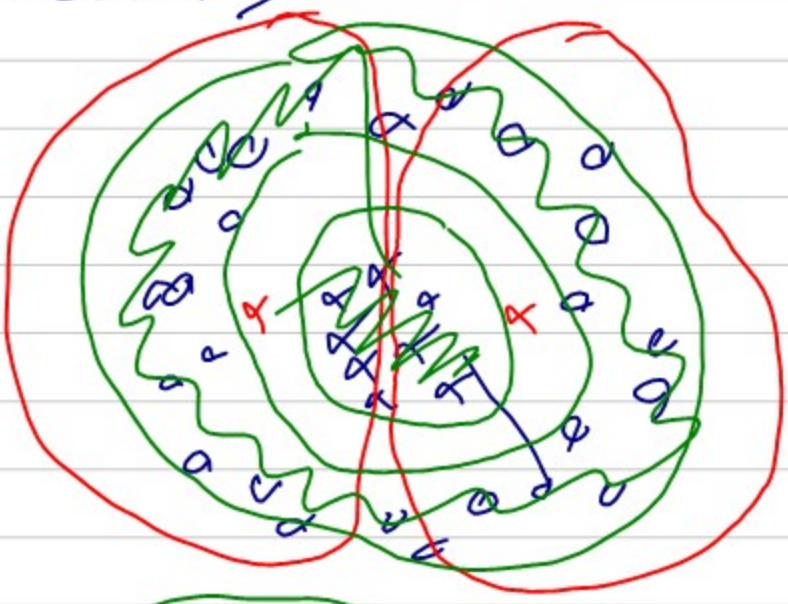
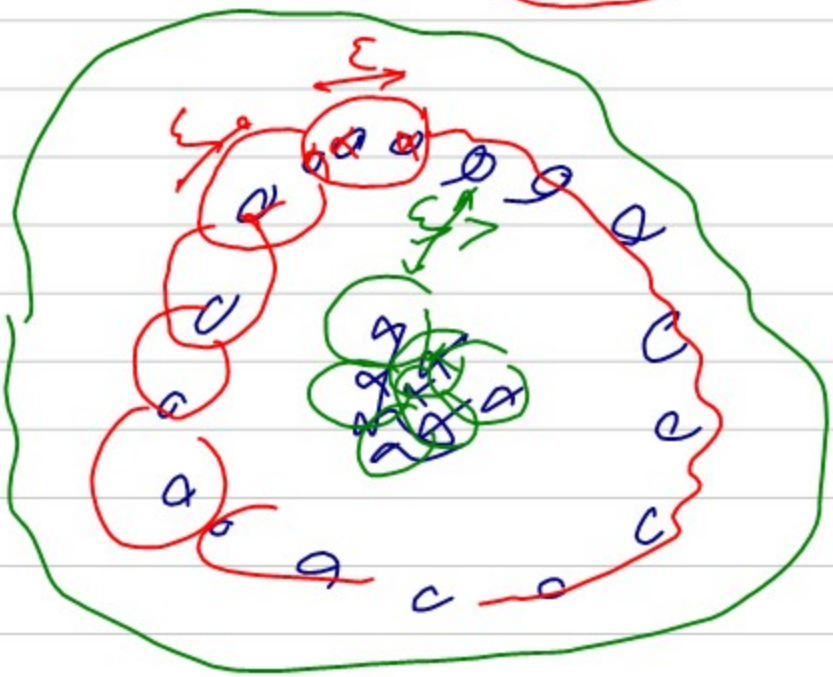
Session 6

جلسه ششم



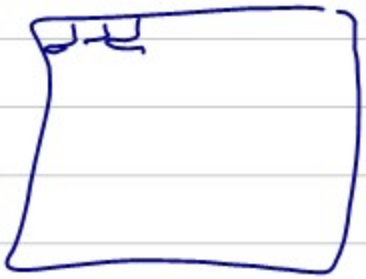


DBSCAN

ε Hyperparameter

Dimentionality Reduction

(DR)



30,000

Book Book 3

Visualization (1)

2D, 3D

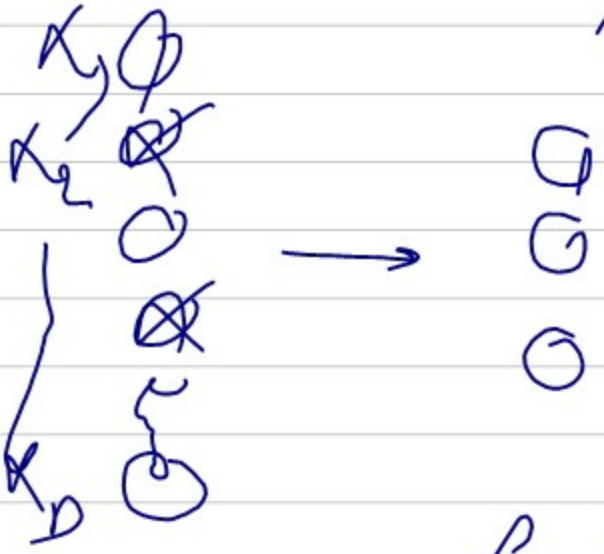
overfitting ← HD (2)

DR ← ↑ timing training (3)

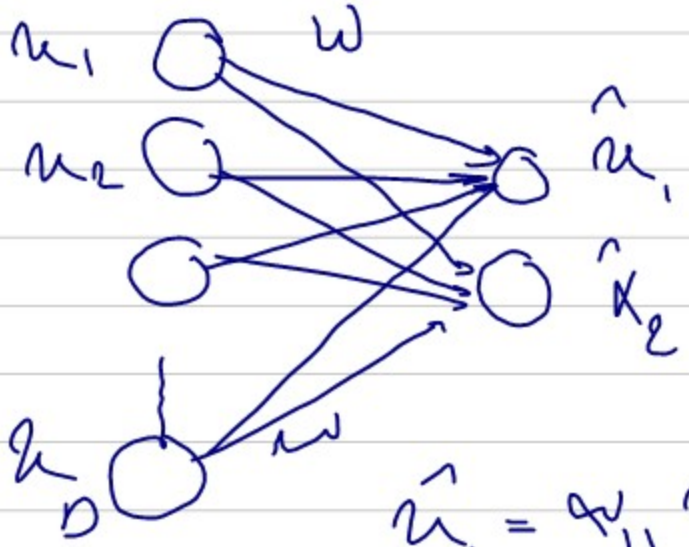
↓ ← DR ← ↑ Cost

DR Methods

feature selection (1)

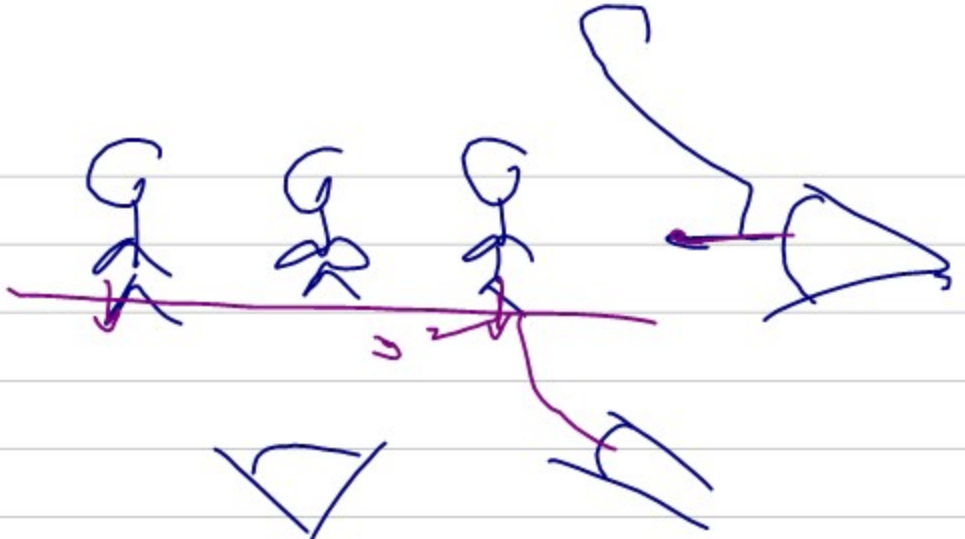


feature extraction (2)

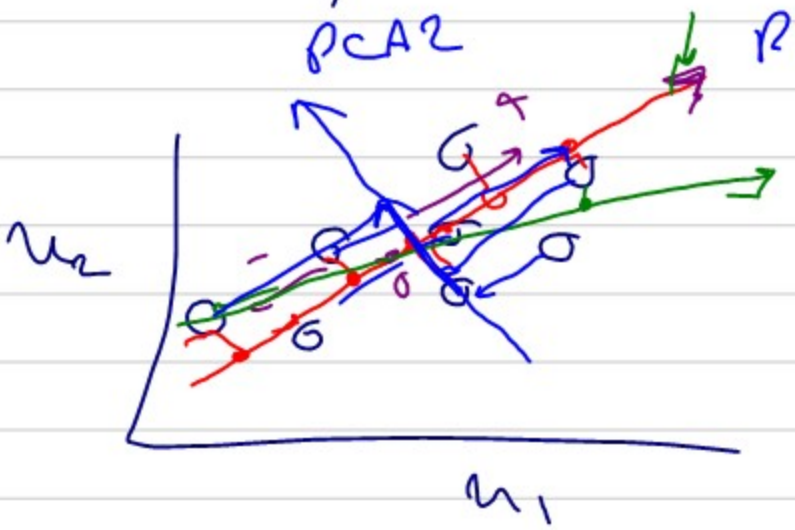


$$\hat{x}_1 = w_{11} x_1 + w_{12} x_2 + \cdots + w_{1D} x_D$$

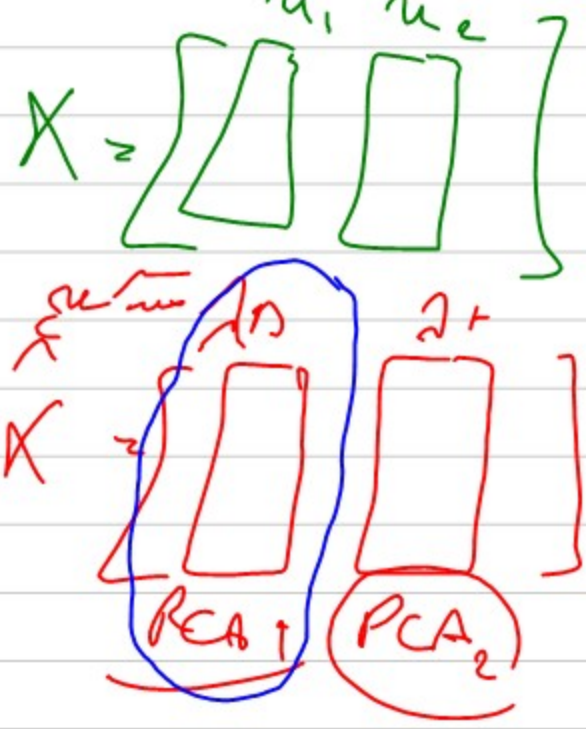
$$\hat{x} = W x$$

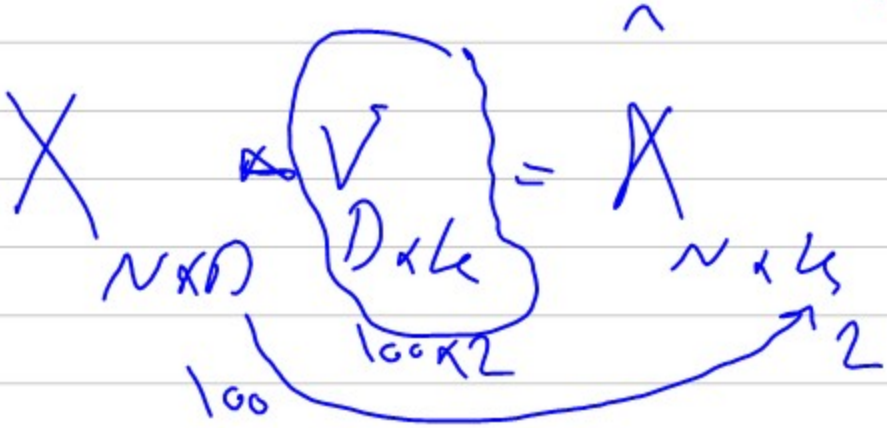


LW

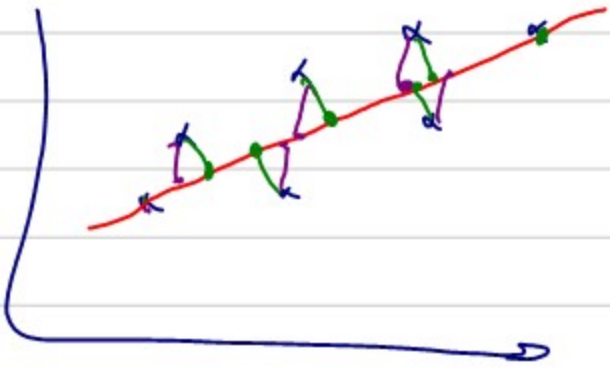


PCA



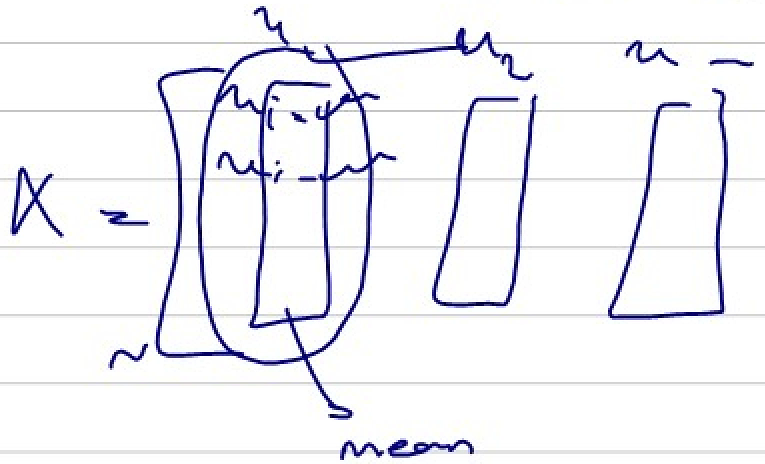


Reg , PCA



PCA: الگوریتم

① mean center



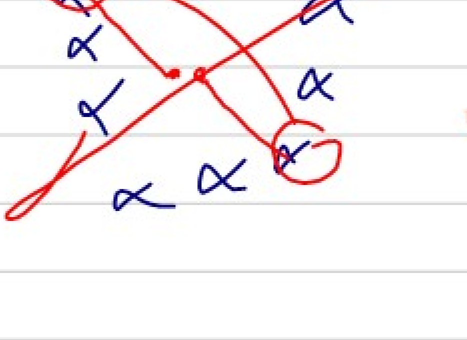
$$\mu_i = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$X - \mu_i$

② استخراج ماتریس COV

③ بردارهای ویژه و مقادیر ویژه

④ مرتب کردن بردارهای ویژه براساس مقادیر ویژه



PCA مشکلات خطی

داده‌های پرت

غیر خطی

t-SNE

غیرخطی

← استخراج اطلاعات از داده با ساختارهای پیچیده

← تصویرسازی بسیار بهتری از داده ها

× کند

داده متوسط و کوچک

UMAP

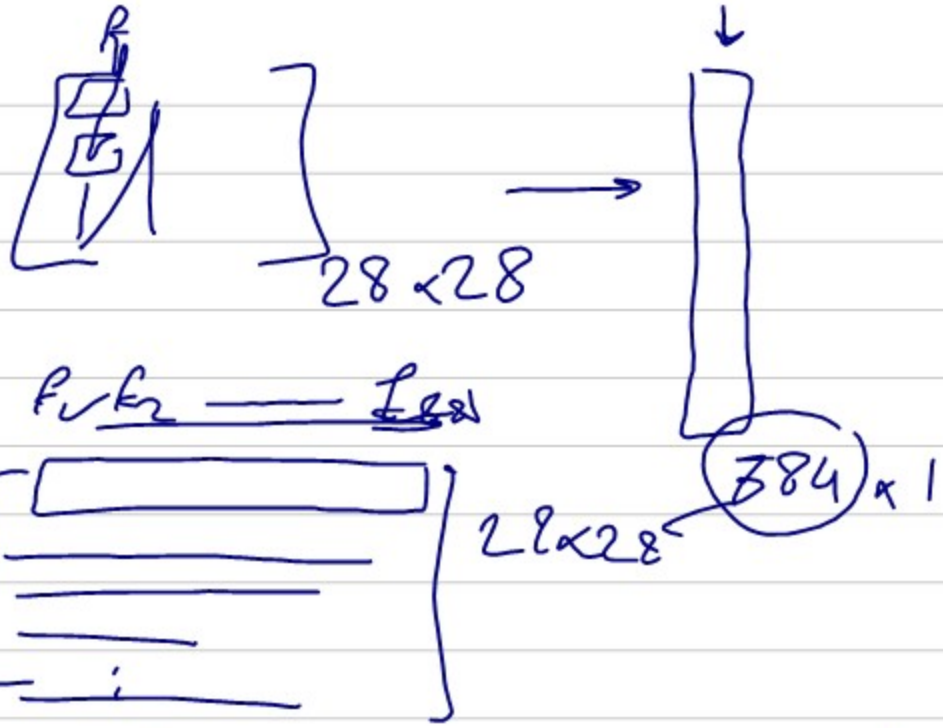
uniform

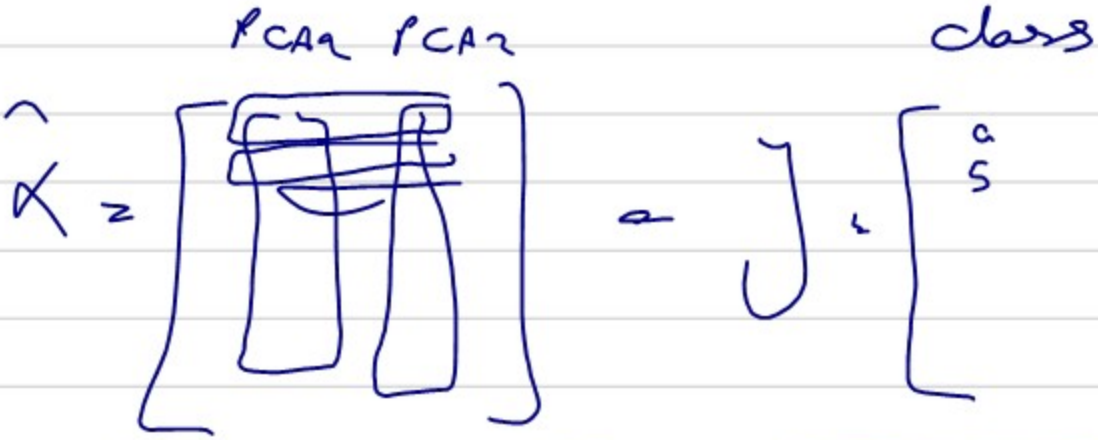
← عملکرد بالا و زمان محاسباتی کمتر

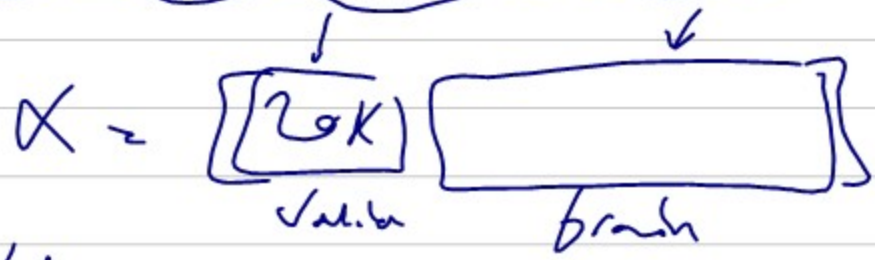
← غیرخطی

← سرعت بیشتر نسبت به t-SNE

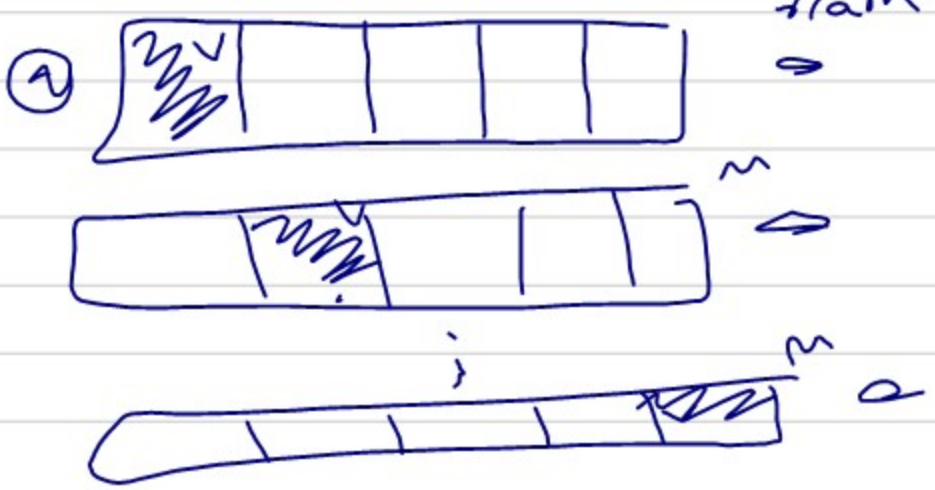
← برای داده های متوسط

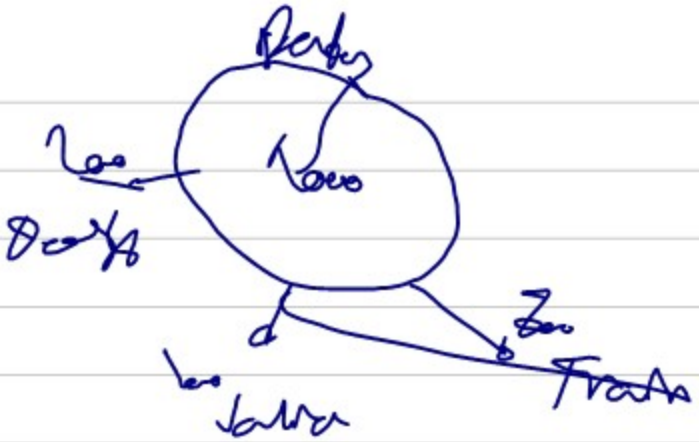

28×28
784 × 1
28×28


class


20K
valid
train

5fold


train
mean
acc
acc
acc
acc



overfitting ← Valid