Last Lecture 14

recap course on Sunday?

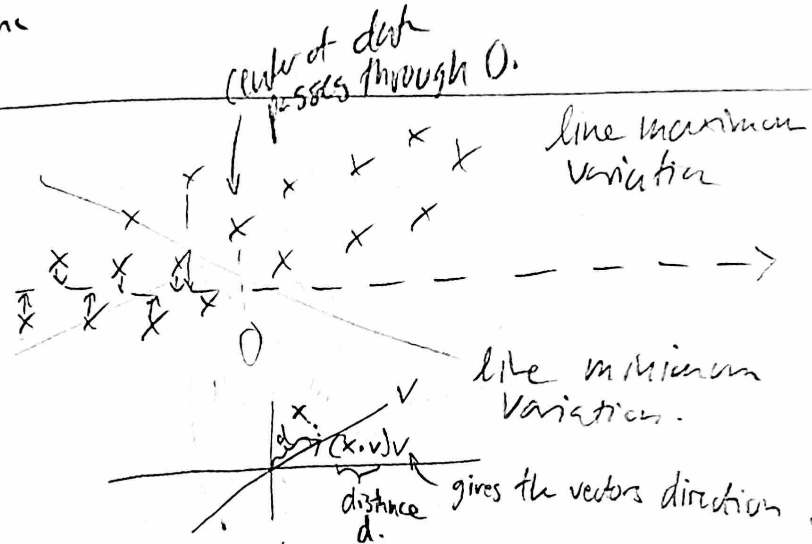
Recap

Tips! Kika på test questions i jupyter-filen för denna

Today

Single Value Decomposition
PCA

- Projections



1. We project all x's to the line.
2. How much variance do I now have? = How much variation have I kept?
3. Find line with max variation.

The vector v describes that line.
We project x on v

Y_i = distance from x's projection on V to 0 = Distance d.

We want to find the direction v when the variance of $(Y_i - \bar{Y}_n)^2$ is maximized.

PCA first steps

Gram matrix

Gram matrix

$$A = \begin{pmatrix} \overset{\text{vector } x_1}{x_1} \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^{n \times \overset{\text{dimension}}{m}}$$

so in our case $A \in \mathbb{R}^{\overset{\#\text{ points}}{14} \times \overset{2 \text{ dimensions}}{2}}$

$$\underset{n\times m}{A}\ \underset{m\times 1}{v} = \begin{pmatrix} x_1 \circ v \\ x_2 \circ v \\ \vdots \\ x_n \circ v \end{pmatrix}$$

$$|Av|^2 = \sum_{i=1}^{n} (x_i \cdot v)^2$$

With that we can replace our problem with:

$$v_1 = \arg\max_{\|v\|=1} |Av|$$

The vector v_1 is called the $\underbrace{\text{first}}$ $\underbrace{\text{singular vector}}_{\text{the squareroot of the empirical variance}}$ of A. $\sigma_1(A) := |Av_1|$

Tips! #samples, #features skilearn usually has the correct form for Gram Matrix

1 - np.sum (w ** 2)) in constrains as we have equality constraint &

Now we want to find the direction with the second max variance.

We search for the $\underset{\text{first}}{\text{the}}$ vector that is orthogonal to the vector (first max variance vector).

$$v_2 := \arg \max_{\|v\| = 1;\, v \perp v_1} |Av|$$

we do so by plotting out all datapoints that are orthogonal to v_1.

Code: A - (A @ normal) ...
with that we remove all points that are orthogonal to v_1.

What is the second singular vector?

let's restrict the first singular value $\sigma_1(A)$

$= \sqrt{\lambda_1} = \sigma_1(A)$

eigenvalue

$$|Av|^2 = Av \circ Av = v \cdot A^T A v \qquad A^T A \in \mathbb{R}^{m \times m}$$

now let $v_1, \dots, v_m$ be eigenvectors of $A^T A$

$\lambda_1, \dots, \lambda_m$ eigenvalues.

$$v_i \circ v_j = 0 \quad \text{if } i \neq j$$

$$\boxed{v = \sum_{j=1}^{m} (v \circ v_j) v_j}$$

basis v.

$$v \cdot A^T A v = \left(\sum_{j=1}^{m} (v \cdot v_j) v_j\right) \cdot \left(A^T A \sum_{i=1}^{m} (v \cdot v_i) v_i\right) =$$

$$= \left(\sum_{j=1}^{m} (v \cdot v_j) v_j\right) \cdot \left(\sum_{i=1}^{m} \lambda_i (v \cdot v_i) v_i\right)$$

$$= \sum_{j=1}^{m} \lambda_i (v \cdot v_i)$$

v should be v_1 here to get out λ_1.

v_1 maximizes the expression.

Skapad med Tiny Scanner

$$A = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \underbrace{\begin{pmatrix} \sum_{j=1}^{m} (x_i \cdot v_j) v_j \\ \vdots \\ \sum_{j=1}^{m} (x_n \cdot v_j) v_j \end{pmatrix}}_{A} =$$

$$= \underbrace{\sum_{j=1}^{m} (A v_j)}_{\begin{pmatrix} \sum (x_1 \cdot v_i) v_j \\ \vdots \\ \sum (x_n \cdot v_i) v_j \end{pmatrix}} = \sum_{j=1}^{m} u_j v_j^T \sigma_j = UDV^T$$

left singular vectors

Dvs $A = UDV^T$

$A \in \mathbb{R}^{n \times m}$

if we want to multiply by v: $Av \Rightarrow v \in \mathbb{R}^m$

$$U = \begin{pmatrix} u_1, u_2, \dots u_m \end{pmatrix}$$

$$D = \begin{pmatrix} \sigma_1 & & & 0 \\ & \sigma_2 & & \vdots \\ 0 & & \ddots & \\ & & & \sigma_m \end{pmatrix}$$

$$V = \underset{m \times m}{\begin{pmatrix} v_1, v_2 \dots v_m \end{pmatrix}}$$

How to do SVD on a Compute

U, D, VT = np.linalg.svd(X, full_matrices = False)

D = np.diag(D)

if correct:

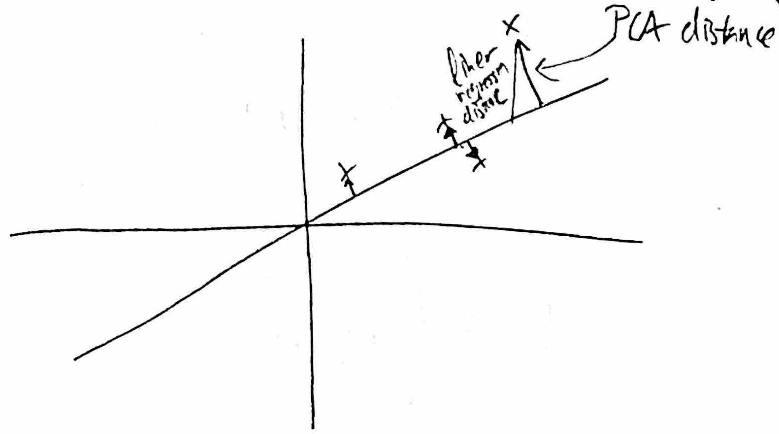
X − U@D@VT should be 0.

& norm(X − U@D@VT) is small.

What is PCA?

$$PCA(X_1, \dots, X_n) = AV = UDV^TV = UD$$

Is PCA the same as linear regression?



No, as PCA minimizes the orthogonal distance & linear regression minimizes the absolute distance