COL774 Machine Learning

Unsupervised Technique

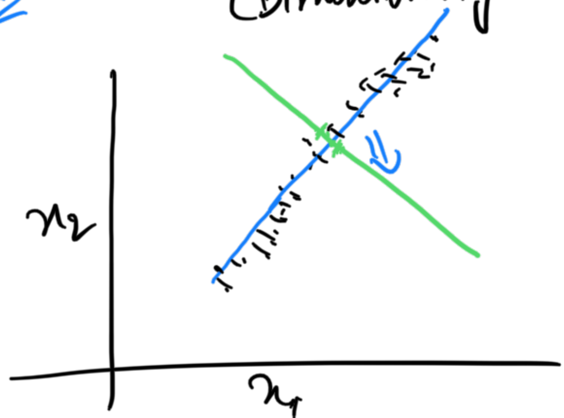
Oct 29, 2021

Assignment 4

Learning Theory

PCA: Principal Component Analysis (Linear)

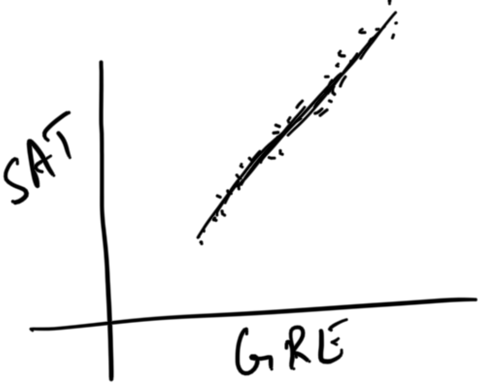
(Dimensionality Reduction)



$\{x^{(i)}\}_{i=1}^{m}$ $\quad x^{(i)} \in \mathbb{R}^n$

n: very very large

↳ (Pre-processing step)



$k << n$

such that most of the variation (variance) in the data can be captured by projecting $\{x^{(i)}\}_{i=1}^{m}$ to a lower dimensional sub-space

$\underline{(u_1 \dots u_k)}$ $\quad u_k \in \mathbb{R}^n$

↳ Eigenfaces:

$\{x^{(i)}\}_{i=1}^{m}$

$x^{(i)} \in \mathbb{R}^{(h \times w) \times 3}$

$\to \mathbb{R}^{(h \times w) \times 3}$ $\quad h, w = 1000$

$u_{k_1}^T u_{k_2} = 0$

↳ orthonormal basis

$u_k^T u_k = 1$

$\Rightarrow [u_1 \dots u_{100}]$

↳ will capture "most" of the features in the data

$$x^{(i)} = \sum_{k=1}^{100} z_k^{(i)} u_k$$

↳ is helpful with computational efficiency

↳ avoids overfitting by eliminating noise

$K = n$: you will get original data back in projected space

$K << n$:- We can retrieve most of the important features in the data.

$Z_K^{(i)} = x^{(i)T} u_K$ $\qquad \{x^{(i)}\}_{i=1}^{m}$

coefficient of projection along u_K $\qquad u_1 \ldots u_K$

Goal of PCA :- To find (given K), the "optimal" set of orthonormal vectors $\in [u_1 \ldots u_K$ (such that the variance of projections ~~to~~ is maximized).

Principal components

Goal!!

PCA :- $\{x^{(i)}\}_{i=1}^{m}$ $\quad x^{(i)} \in \mathbb{R}^n$ $\quad u_1 \ldots u_K$

$u_K \in \mathbb{R}^n$

Pre-processing :-

$j \in \{1 \ldots n\}$ $\quad \{x^{(i)}\}_{i=1}^{m}$ are converted to be zero mean & unit variance.

$\forall j$:- $x_j^{(i)} \leftarrow \dfrac{x_j^{(i)} - \mu_j}{\sigma_j}$ $\qquad \mu_j = \dfrac{\sum_{i=1}^{m} x_j^{(i)}}{m}$

$\sigma_j^2 = \frac{1}{m}\sum_{i=1}^{m}(x_j^{(i)} - \mu_j)^2$

~~After~~ ~~th~~

Derive the expression for $u_1 \ldots u_K$ when $K = 1$

u :- $Z^{(i)} = x^{(i)T} u$

Variance of the projected data :- $\{Z^{(i)}\}_{i=1}^{m}$

$$\mu_Z = \frac{1}{m}\sum_{i=1}^{m} Z^{(i)} = \frac{1}{m}\sum_{i=1}^{m}(x^{(i)T} u)$$

$$= \frac{1}{m}\left[\sum_{i=1}^{m} x^{(i)}\right]^T u$$

$$= \overline{0}$$

$$\text{Var}(z) = \frac{1}{m}\sum_{i=1}^{m} z^{(i)2} \qquad z^{(i)} = x^{(i)T}u$$

$$= \frac{1}{m}\sum_{i=1}^{m} \left(x^{(i)T}u\right)^2$$

$$= \frac{1}{m}\sum_{i=1}^{m} \left(u^T x^{(i)}\right)\left(x^{(i)T}u\right)$$

$$= \frac{1}{m}\sum_{i=1}^{m} u^T \underbrace{\left(x^{(i)}x^{(i)T}\right)}\, u \qquad \frac{1}{m}\sum_{i=1}^{m}(x^{(i)}-\underbrace{\mu}_{=0})(x^{(i)}-\underbrace{\mu}_{=0})^T$$

$$= u^T \underbrace{\left[\frac{1}{m}\sum_{i=1}^{m} \left(x^{(i)}x^{(i)T}\right)\right]}\, u$$

$\hookrightarrow$ Empirical Co-variance (Recall: ~~mean~~ data mean $=0$)

$$= u^T \Sigma u$$

Therefore:- The objective is to find u s.t. $\text{Var}(z)$ is maximized: $\to$ empirical covariance matrix

$$\max_u \underbrace{u^T \Sigma u} \quad \therefore \text{objective}$$

$$\Sigma = X^T X \quad (\text{to prove})$$

$$= \frac{1}{m}\sum_{i=1}^{m} x^{(i)}x^{(i)T}$$

Aside / Note:-

$X =$
$\hookrightarrow$ Design matrix

$$\begin{bmatrix} — & x^{(i)T} & — \end{bmatrix}$$

$\Rightarrow$ Solution to

$\max_u \; u^T \hat{\Sigma} u$ is the principal eigenvector of $\hat{\Sigma} \in \mathbb{R}^{n \times n}$

$u \in \mathbb{R}^n$

$\because \hat{\Sigma}$:- symmetric $\equiv X^T X$

$\hookrightarrow$ +ve semi-definite

$\Leftrightarrow$ All eigenvalues of $\hat{\Sigma}$ are real & non negative

$\hat{\Sigma} u = \lambda u$

max eigenvalue

$$\begin{cases} P^T (X^T X) P \geq 0 \\ (XP)^T (XP) \geq 0 \end{cases}$$

$$\left.\begin{array}{ll} \max_u & u^T \hat{\Sigma} u \\ \text{subject} & u^T u = 1 \end{array}\right]$$

$\Leftrightarrow$ Theory of Langrangian.

$$L(u, \lambda) = u^T \hat{\Sigma} u + \lambda(1 - u^T u)$$

Primal:- $\max_u \min_\lambda \left[u^T \hat{\Sigma} u + \lambda(1 - u^T u) \right]$

Dual:- $\min_\lambda \left[\max_u \left[u^T \hat{\Sigma} u + \lambda (1 - u^T u) \right] \right]$

$$\nabla_u \left[u^T \hat{\Sigma} u + \lambda (1 - u^T u) \right] = 0$$

$$\Rightarrow 2 \hat{\Sigma} u + 2(-\lambda) u = 0$$

$$\Rightarrow \boxed{\hat{\Sigma} u = \lambda u}$$

maximum eigenvalue

In order to project on u, s.t. variance in projection is maximized
u:- (principal) eigenvector of $\hat{\Sigma}_0$

$(u_1 \dots u_K)$:- $\{z^{(i)}\}_{i=1}^{m}$

$z^{(i)} \in \underline{\mathbb{R}^K}$

$\underline{(z^{(i)}_1 \dots z^{(i)}_K)}$

$\mathrm{Var}(z)$

$$= \sum_{k=1}^{K} u_k^T \hat{\Sigma} u$$

In one dim.
$\mathrm{Var}(z) = u^T \hat{\Sigma}_0 u$
constrained optimization problem.

:- Objective:-

$$\begin{cases} \max\limits_{u_1 \dots u_K} \sum\limits_{k=1}^{K} u_k^T \hat{\Sigma}_0 u \\ u_k^T u_k = 1 \qquad u_{k_1}^T u_{k_2} = 0 \quad k_1 \neq k_2 \end{cases}$$

$\Rightarrow$ You can show that $(u_1 \dots u_K)$ correspond to eigenvectors of $\hat{\Sigma}_0$ corresponding to K eigenvalue.

$$\hat{\Sigma} u_k = \lambda_k u_k$$

$(\lambda_1 - \lambda_K)$:- Highest eigenvalue of $\hat{\Sigma}_0$

$(u_1 - u_K)$:- principal components.

$\Rightarrow$ Objective:- finding top K eigenvectors of

$$\hat{\Sigma}_0 = \underline{X^T X}$$

$\Rightarrow$ Different ways of solving the problem.-

<u>SVD:-</u> Singular value Decomposition

$\underline{X} \in \mathbb{R}^{m \times n}$:- Help us get eigenvalues of $(\underline{X^T X})$

eigenvectors