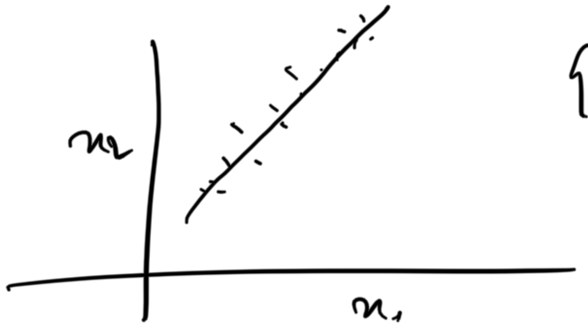
COL 774
Machine Learning
Nov 9, 2021

Last Class:-

PCA (Principal Component Analysis)

↳ Dimensionality Reduction



$$\underbrace{\{x^{(i)}\}_{i=1}^{m} \qquad x^{(i)} \in \mathbb{R}^n}_{\text{Given.}}$$

Goal:- Find

$(u_1 \ldots u_K)$, $\|u_k\|^2 = 1$

$u_k \in \mathbb{R}^n$ $\qquad u_{k_1}^T \cdot u_{k_2} = 0$

s.t. $\{x^{(i)}\}_{i=1}^{m}$ when projected onto $\quad k_1 \neq k_2$

$(u_1 \ldots u_K)$, variance is maximum.

$$\Rightarrow \underset{u_1 \ldots u_K}{\arg\max} \sum_{k=1}^{K} u_k^T \hat{\Sigma} u_k \quad \to \text{co-variance matrix}$$

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

$$X = \begin{bmatrix} \text{—} \; x^{(i)T} \; \text{—} \end{bmatrix}$$

$\Rightarrow$ The solution to above optimization problem is top $\overline{K \text{ eigenvectors}}$ of $\hat{\Sigma}$ [Note:- $\hat{\Sigma}$ is the semi-definite $\Rightarrow$ all eigenvalues of $\hat{\Sigma} \geq 0$)

How to compute these eigenvectors?

↳ Direct method → (a) compute $X^T X$ $(\hat{\Sigma})$ $\to \mathbb{R}^{n \times n}$

(b) compute Eigenvectors of

$$O(\overbrace{n^2 m}^{X^TX} + n^3)$$

Expensive if n is large

⇒ SVD:- Singular Value Decomposition
(more efficient)

$$\overset{\in \mathbb{R}^{m\times n}}{X} = UDV^T$$

U, V:- are matrices st their columns form an orthonormal basis ⇐

$$\begin{cases} U \in \mathbb{R}^{m\times m} \\ D \in \mathbb{R}^{m\times n} \quad \text{(Diagonal)} \\ V \in \mathbb{R}^{n\times n} \end{cases}$$

$n \geq m$

$$D = m\left[\begin{array}{ccc|c} \lambda_1 & & 0 & \\ & \ddots & & 0 \\ 0 & & \lambda_m & \end{array}\right]$$

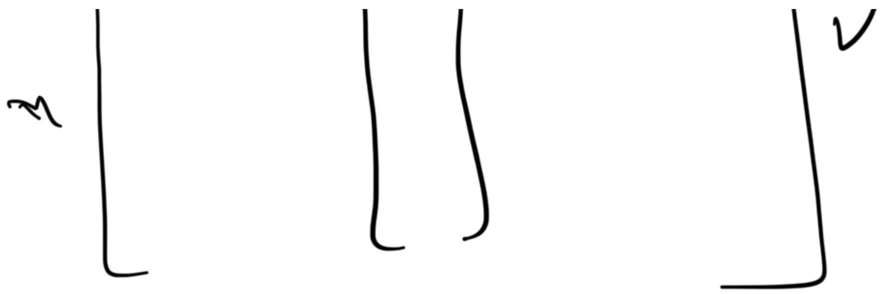
$$\underbrace{\quad\quad\quad}_{m}\ \underbrace{\quad}_{(n-m)}$$

⇒ Then (Also)
:- Columns of U are eigenvectors of XX^T
Column of V are eigenvectors of X^TX

⇒ λ_j:- λ_j's are square (roots) of eigenvalues of X^TX
$= XX^T$

$$U = \begin{bmatrix} \cdots & U_{:j} & \cdots \end{bmatrix} m \qquad \begin{bmatrix} \ddots & & \\ & D\lambda_j & \\ & & \ddots \end{bmatrix}$$

$V_{:j}$ n



$$X^TX = (UDV^T)^T\, UDV^T$$

$$= VD^T\underbrace{U^TU}_{I}DV^T$$

$$= VD^TDV^T = X^TX$$

~~Dst~~ Post multiplying by V

$$X^TXV = VD^TD\underbrace{V^TV}_{I}$$

$$\boxed{X^TXV = VD^TD}$$

↳ Show that columns of V are eigenvectors of X^TX & non-zero entries of D are square roots of eigenvalues of X^TX

$\Rightarrow O(\min(m^2n, n^2m))$

$\Rightarrow$ Fast Approximations to compute SVD

SVD:- X ↳ col. of V or SVD:- X^T ↳ col of U

Eigenfaces

$\{x^{(i)}\}_{i=1}^{m}$ $x^{(i)} \in \mathbb{R}^{h \times w}$

$\Rightarrow \quad u_1 \dots u_k$

"Eigenfaces" :- $x^{(i)} = \sum_{k=1}^{K} z_k^{(i)} u_k$

Learning Theory :- $\{x^{(i)}, y^{(i)}\}_{i=1}^{m} \equiv \underline{D}$ (training Data)

$y^{(i)} \in$ Discrete set (classification)

$\{0, 1\}$:- Binary classification

$x^{(i)} \in \mathbb{R}^n$

$\in$ Discrete space

Data :- $(x^{(i)}, y^{(i)}) \sim$ Dist. (unknown)

$\hookrightarrow \quad (x, y) \sim$ Dist Test Data

:- Goal :- $\mathcal{H}$:- hypotheses space.

$h \in \mathcal{H}$

$h(x^{(i)}) \simeq y^{(i)}$?

Training error

:- Training error :- $\hat{\varepsilon}(h) = 1\{y^{(i)} \neq h(x^{(i)})\}$ → ①

Test error :- $\varepsilon(h) = E_{(x,y) \sim Dist}[1\{h(x) \neq y\}]$ → ②

:- Question :- how good is $\hat{\varepsilon}(h)$ an approximation of $\varepsilon(h)$