CS 307 Lecture 24

PCA: Reduce dimensionality of the input vectors to a lower dimensional space such that variance of the projected data in the lower dimensional space is maximized.

Before PCA, normalize data so that mean is 0 & variance is 1.

To reduce data to 1-dimension we need to find vector u s.t.

$$\max_{\|u\|^2=1} \frac{1}{n}\sum_{i=1}^{n}\left(x^{(i)^T} u\right)^2 = \max_{\|u\|=1} u^T A u$$

where $A = \frac{1}{n}\sum_{i=1}^{n} x^{(i)}(x^{(i)})^T$ ← symmetric

Proposition: For any symmetric matrix $A \in \mathbb{R}^{d\times d}$ with max eigenvalue λ_1 and min eigenvalue λ_d, $\forall x \in \mathbb{R}^d$ such that $\|x\| = 1$,

$$\lambda_d \le x^T A x \le \lambda_1$$

Furthermore that max and min are achieved when

$x = u_1$ (eigenvector corresponding to λ_1)

and $x = u_d$ (eigenvector corres. to λ_d)

Data d-dimensions $\longrightarrow$ k dimensions $(k < d)$

Then you should project the training set to the k eigenvectors that correspond to the k-largest eigenvalues of matrix $A = \frac{1}{n} \sum_{i=1}^{n} x^{(i)} x^{(i)T}$.

Recall any symmetric matrix A can be written as

$$A = U \Lambda U^T$$

where Λ is a diagonal matrix of eigenvalues.

$$\Lambda = \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_d \end{bmatrix} \qquad \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d.$$

$$U = \begin{bmatrix} | & | & & | \\ u_1 & u_2 & \dots & u_d \\ | & | & & | \end{bmatrix}$$ where $\{u_1, u_2 \dots u_d\}$ is orthonormal.

and $A u_i = \lambda_i u_i$.

PCA reduction to k-dimensions

$$\underset{\in \mathbb{R}^d}{x^{(i)}} \longmapsto \begin{bmatrix} x^{(i)T} u_1 \\ x^{(i)T} u_2 \\ \vdots \\ \cancel{x}\, x^{(i)T} u_k \end{bmatrix} \in \mathbb{R}^k$$

where $u_1 \dots u_k$ are k (principal) eigenvector ~~ver~~ for A.

$$X = \begin{bmatrix} \text{—} & x^{(1)T} & \text{—} \\ \text{—} & x^{(2)T} & \text{—} \\ \text{—} & x^{(n)T} & \text{—} \end{bmatrix} \qquad U_k = \begin{bmatrix} | & | & & | \\ u_1 & u_2 & \dots & u_k \\ | & | & & | \end{bmatrix}$$

Reduced training set is $X U_k$.

Applications of PCA

- Visualization.
- Compression.

For compression:

$$x \in \mathbb{R}^d \xrightarrow{\text{compression}} y \in \mathbb{R}^k$$

$$\hat{x} \in \mathbb{R}^d \xleftarrow{\text{uncompress}}$$

Compressing: $\underbrace{W}_{k \times d} \underbrace{x}_{d \times 1} \in \mathbb{R}^k$

Uncompressing: $\underbrace{U}_{d \times k} \underbrace{y}_{k \times 1} \in \mathbb{R}^d$

Find U and W such that $\forall x^{(i)}$

$$\min \sum \| x^{(i)} - UWx^{(i)} \|^2$$

Take $W = U_k$ (k-principal orthonormal eigenvectors of A)

$U = U_k^T$.

- Learning
  * Reducing dimensions helps reduce computational costs during ~~...~~ learning
  * Reducing dimension may help ~~...~~ avoid overfitting.

PCA algorithm : Compute k-principal orthonormal eigenvectors for matrix A where

$$\underbrace{A}_{\in d\times d} = \frac{1}{n}\sum_{i=1}^{n} x^{(i)} x^{(i)T} = \frac{1}{n} X X^T$$

$\hookrightarrow \in \mathbb{R}^d$.

$$X = \begin{bmatrix} \text{---} & x^{(1)T} & \text{---} \\ \text{---} & x^{(2)T} & \text{---} \\ & \vdots & \\ \text{---} & x^{(n)T} & \text{---} \end{bmatrix}$$

$x^{(i)}$ is a 100×100 images $\longrightarrow d = 10{,}000$

A is $(10{,}000)^2$

Singular Value Decomposition (SVD)

Definition : For a matrix $A \in \mathbb{R}^{n\times m}$, $\sigma \in \mathbb{R}$ is a singular value of A and u and v are the left and right singular vectors of A w.r.t σ if the following to conditions hold.

(a) $A v = \sigma u$ and — $v \in \mathbb{R}^m$, $u \in \mathbb{R}^n$

(b) $A^T u = \sigma v$.

Theorem : Let $A \in \mathbb{R}^{n\times m}$ and let rank(A) = r. There are orthonormal vectors $u_1, u_2 \ldots u_r$ and $v_1, v_2 \ldots v_r$. and $\sigma_1 \geq \sigma_2 \ldots \geq \sigma_r > 0$ where

linearly independent columns/rows of A.

u_i and v_i are the left and right singular vector for σ_i

such that

$\rightarrow A = U \Sigma V^T$ where $\Sigma = \begin{bmatrix} \sigma_1 & & 0 \\ & \sigma_2 & \\ & & \ddots \\ 0 & & & \sigma_r \end{bmatrix}$

$$U = \begin{bmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_r \\ | & | & & | \end{bmatrix} \quad \text{and} \quad V = \begin{bmatrix} | & & | \\ v_1 & \cdots & v_r \\ | & & | \end{bmatrix}$$

Proposition: σ, u, v are ~~the~~ a singular value, left singular vector, right singular vector of A iff σ^2, v (or σ^2, u) are an (non-zero) eigen value λ of A^TA (of AA^T)

Proof: If σ, u, v are singular ... of A Then

$$Av = \sigma u \quad \text{and} \quad A^Tu = \sigma v.$$

$$A^TAv = A^T(\sigma u) = \sigma A^Tu = \sigma^2 v.$$

$$AA^Tu = \sigma^2 u.$$

Suppose $\lambda \neq 0$ is an eigen value of A^TA.

Since A^TA is symmetric, $\lambda > 0$.

Let v be the eigenvector corresponding to λ.

$$u = \frac{1}{\sqrt{\lambda}} Av$$

$$A^Tu = A^T\left(\frac{1}{\sqrt{\lambda}} Av\right) = \frac{1}{\sqrt{\lambda}} A^TAv = \frac{1}{\sqrt{\lambda}} \lambda . v = \sqrt{\lambda}\, v.$$

$$\sqrt{\lambda}\, u = \sqrt{\lambda}\left[\frac{1}{\sqrt{\lambda}} Av\right] = Av.$$

Coming back to PCA:

Need to compute eigenvectors of $A = \frac{1}{n} XX^T$.

Instead we compute the SVD of X. (or X^T)

SVD of X will give eigenvectors / eigenvalues of XX^T.

$$A = U \Sigma V^T \quad \text{(SVD)}$$

$$A^+ = V \Sigma' U^T \quad \text{where} \quad \Sigma' = \begin{bmatrix} \frac{1}{\sigma_1} & & 0 \\ & \frac{1}{\sigma_2} & \\ 0 & & \ddots \; \frac{1}{\sigma_r} \end{bmatrix}$$