4/4/2023

How to choose w to maximize $\frac{1}{n}\sum_{i=1}^{n}(w^T x^{(i)})^2$?

Recall: we restrict w to have $\|w\|=1$

$$\frac{1}{n}\sum_{i=1}^{n}\underbrace{(w^T x^{(i)})}_{1\times 1}\underbrace{(x^{(i)T} w)}_{1\times 1}$$

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

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

$$xx^T=\begin{bmatrix} x_1^2 & x_1x_2 & & \\ x_1x_2 & \ddots & & \\ & & & x_d^2 \end{bmatrix}$$

is symmetric

Covariance matrix of data $\{x^{(1)},\ldots,x^{(n)}\} = \underbrace{\frac{1}{n}\sum_{i=1}^{n}x^{(i)}x^{(i)T}}_{\text{also symmetric}}$

because mean of $x^{(i)}$'s is 0

$$= w^T\Sigma w$$

covariance matrix, is symmetric

Every symmetric matrix can be written as

$$\Sigma = UDU^T \quad \text{where: } D=\begin{bmatrix}\lambda_1 & \ddots & 0 \\ 0 & & \lambda_d\end{bmatrix} \text{ diagonal}$$

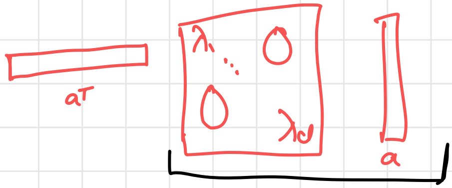
and U is orthonormal $\begin{bmatrix} u_1 & \cdots & u_d \end{bmatrix}$

Each u_i is unit vector and $u_i^T u_j = 0$ for $i \neq j$ (ie, they're orthogonal)

$$\underset{w}{\text{maximize}} \quad w^T \Sigma w = \underbrace{w^T U}_{a^T} D \underbrace{U^T w}_{a}$$

Define $a = U^T w$ — still a unit vector since w is unit vector as we just changed the basis

→ same as $\underset{a}{\text{maximizing}} \quad a^T D a$



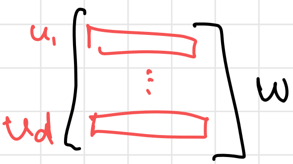
$$= \sum_{j=1}^{d} \lambda_j a_j^2 \qquad \begin{bmatrix} \lambda_1 a_1 \\ \lambda_2 a_2 \\ \vdots \\ \lambda_d a_d \end{bmatrix}$$

Constraint: $\sum_{j=1}^{d} a_j^2 = 1$

we can assume $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_d$

optimal solution: $a_1 = 1$, all other entries of $a = 0$

$$a = U^T w = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$



so: $w = u_1$

Recap: Given data $\{x^{(1)}, \ldots, x^{(n)}\}$

1. Mean-center data
2. Compute $\Sigma = \frac{1}{n} \sum_{i=1}^{n} x^{(i)} x^{(i)T}$
3. Decompose Σ as UDU^T
4. Choose w to be eigenvector corresponding to largest eigenvalue

What if you want >1 dimension?

e.g. $x^{(i)}$ is 1000-dimensional

want to visualize the data, so reduce it to 2 dimensions

Solution: Choose eigenvectors for 2 largest eigenvalues

- HW3 due next Tuesday
- Section next 2 weeks
  - Language models (this week)
  - Vision models including image generation (next week)