Section B

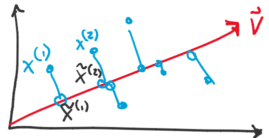
Monday, November 28, 2022 9:45 AM

Objectives for PCA

Written in terms of the first principal component, called $\vec{v}_1$ here

$\vec{v}_1$

(1) Minimize the Reconstruction Error



$$\vec{v}_1 = \underset{\vec{v}}{\operatorname{argmin}} \ \frac{1}{N} \sum_{i=1}^{N} \text{distance}\left(\vec{x}^{(i)}, \tilde{x}^{(i)}\right)^2$$

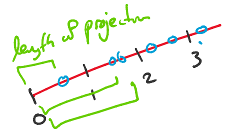
$$= \underset{\vec{v}}{\operatorname{argmin}} \ \frac{1}{N} \sum_{i=1}^{N} \Big\| x^{(i)} - \overbrace{\underbrace{\left(\text{vector projection of } \vec{x}^{(i)} \text{ onto } \vec{v}\right)}_{(\vec{v}^T \vec{x}^{(i)})\vec{v} / \|\vec{v}\|_2^2}}^{\tilde{x}^{(i)}} \Big\|_2^2$$

$$= \underset{\substack{\vec{v} \text{ s.t.} \\ \|\vec{v}\|_2 = 1}}{\operatorname{argmin}} \ \frac{1}{N} \sum_{i=1}^{N} \left\| x^{(i)} - (\vec{v}^T \vec{x}^{(i)})\vec{v} \right\|_2^2$$

(2) Maximize the Variance

Variance of a sample $x^{(1)}, \dots, x^{(N)} \in \mathbb{R}$ is $\sigma^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x^{(i)} - \mu)^2$

$\mu = \frac{1}{N} \sum_{i=1}^{N} x^{(i)}$



$$\vec{v}_1 = \underset{\vec{v}}{\operatorname{argmax}} \ \frac{1}{N} \sum_{i=1}^{N} \left(\text{length of vector proj. of } x^{(i)} \text{ onto } \vec{v} \right)^2$$

$$= \underset{\substack{\vec{v} \text{ s.t.} \\ \|\vec{v}\|_2 = 1}}{\operatorname{argmax}} \ \frac{1}{N} \sum_{i=1}^{N} \left(\vec{v}^T \vec{x}^{(i)} \right)^2$$

$$= \underset{\substack{\vec{v} \text{ s.t.} \\ \|\vec{v}\|_2 = 1}}{\operatorname{argmax}} \ \underbrace{\frac{1}{N} (\vec{v}^T X^T)(X \vec{v})}_{\vec{v}^T \Sigma \vec{v}}$$

$$= \underset{\substack{\vec{v} \text{ s.t.} \\ \|\vec{v}\|_2 = 1}}{\operatorname{argmax}} \ \vec{v}^T \Sigma \vec{v}$$

← why not use Gradient Descent?