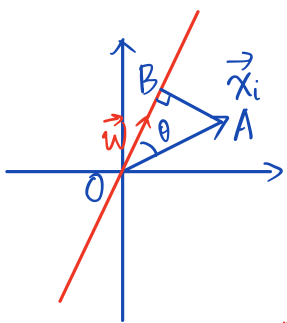
现有 X: $n\times p$, n是观测值的个数, p是features个数. X可以看成p维空间的一组点.

问题: 如何在尽可能保留信息的情况下, 把X的维度减少?

方案: 把X投影到q维空间上去, 使这n个点到投影点的距离最小.

从p维投影到过原点的一条直线上. (红线)



$\vec{w}$是红线的unit vector.

OB的长度是 $\|\vec{x}_i\|\cdot\cos\theta = \|\vec{x}_i\|\dfrac{\vec{x}_i\cdot\vec{w}}{\|\vec{x}_i\|\|\vec{w}\|} = \vec{x}_i\cdot\vec{w}$

OB的坐标是 $(\vec{x}_i\cdot\vec{w})\vec{w}$

目标: 希望投影过后仍然尽可能保持 $\vec{x}_i$的信息. 信息就是长度.

目标函数: $\min\limits_{\vec{w}} \frac{1}{n}\sum_{i=1}^{n}\|\vec{x}_i-(\vec{x}_i\cdot\vec{w})\vec{w}\|^2$

$$\|\vec{x}_i-(\vec{x}_i\cdot\vec{w})\vec{w}\|^2 = \vec{x}_i\cdot\vec{x}_i - 2(\vec{x}_i\cdot\vec{w})^2 + (\vec{x}_i\cdot\vec{w})^2\,\vec{w}\cdot\vec{w}$$
$$= \vec{x}_i\cdot\vec{x}_i - (\vec{x}_i\cdot\vec{w})^2$$

$$MSE(\vec{w}) = \frac{1}{n}\sum_{i=1}^{n}\|x_i\|^2 - \frac{1}{n}\sum_{i=1}^{n}(\vec{x}_i\cdot\vec{w})^2$$

最小化MSE, 也即最大化 $\frac{1}{n}\sum_{i=1}^{n}(\vec{x}_i\cdot\vec{w})^2$

$$\frac{1}{n}\sum_{i=1}^{n}(\vec{x}_i\cdot\vec{w})^2 = \Big(\underbrace{\frac{1}{n}\sum_{i=1}^{n}(\vec{x}_i\cdot\vec{w})}_{=0}\Big)^2 + Var(x_i\cdot\vec{w})$$

$$= \Big(\frac{1}{n}\sum_{i=1}^{n}\vec{x}_i\Big)\cdot\vec{w}$$

$\vec{x}_i$已作~~标准化~~, 所以第一项为0.

中心化, x均值为0, $\frac{1}{n}\sum_{i=1}^{n}x_i = 0$, $\forall x^j \in \{x^1, x^2, \dots, x^p\}$

若投影到q维空间

$\sum_{j=1}^{q}(\vec{x}_i\cdot\vec{w}_j)\vec{w}_j$ ， w_j 是q维子空间的标准正交基

$\|\underset{p\times1}{\vec{x}_i} - \sum_{j=1}^{q}(\vec{x}_i\cdot\vec{w}_j)\underset{p\times1}{\vec{w}_j}\|^2$

p维空间中q个方向上的和，与$\vec{x}$的距离不希望太大

$= \vec{x}_i\cdot\vec{x}_i - 2\sum_{j=1}^{q}(\vec{x}_i\cdot\vec{w}_j)\,\vec{x}_i\cdot\vec{w}_j + \underbrace{\left(\sum_{j=1}^{q}(\vec{x}_i\cdot\vec{w}_j)\vec{w}_j\right)\cdot\left(\sum_{j=1}^{q}(\vec{x}_i\cdot\vec{w}_j)\vec{w}_j\right)}_{☆}$

$☆ = \left[(\vec{x}_i\cdot\vec{w}_1)\vec{w}_1+\cdots+(\vec{x}_i\cdot\vec{w}_q)\vec{w}_q\right]\cdot\left[(\vec{x}_i\cdot\vec{w}_1)\vec{w}_1+\cdots+(\vec{x}_i\cdot\vec{w}_q)\vec{w}_q\right]$

$= (\vec{x}_i\cdot\vec{w}_1)^2\vec{w}_1\cdot\vec{w}_1 + (\vec{x}_i\cdot\vec{w}_2)^2\vec{w}_2\cdot\vec{w}_2 + \cdots + (\vec{x}_i\cdot\vec{w}_q)^2\vec{w}_q\cdot\vec{w}_q$

$= \sum_{j=1}^{q}(x_i\cdot w_j)^2$

$= \vec{x}_i\cdot\vec{x}_i - \sum_{j=1}^{q}(\vec{x}_i\cdot\vec{w}_j)^2$

最大化 $\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{q}(\vec{x}_i\cdot\vec{w}_j)^2 = \sum_{j=1}^{q}\frac{1}{n}\sum_{i=1}^{n}(\vec{x}_i\cdot\vec{w}_j)^2$

$= \frac{1}{n}\sum_{i=1}^{n}(\vec{x}_i\cdot\vec{w}_1)^2 + \frac{1}{n}\sum_{i=1}^{n}(\vec{x}_i\cdot\vec{w}_2)^2 + \cdots + \frac{1}{n}\sum_{i=1}^{n}(\vec{x}_i\cdot\vec{w}_q)^2$

$= \frac{1}{n}(\mathbf{X}\vec{w}_j)^T(\mathbf{X}\vec{w}_j)$

$= \frac{1}{n}\vec{w}_j^T\mathbf{X}^T\mathbf{X}\vec{w}_j$

$= \vec{w}_j^T\frac{\mathbf{X}^T\mathbf{X}}{n}\vec{w}_j$

$= \vec{w}_j^T\,\mathrm{Var}(\mathbf{X})\,\vec{w}_j$

约束条件： $\vec{w}_j\cdot\vec{w}_j = 1$.

Lagrange： $\vec{w}_j^T V\vec{w}_j - \lambda(\vec{w}_j\cdot\vec{w}_j - 1)$

$$\frac{\partial L}{\partial \lambda}: \quad w^T w - 1 = 0$$

$$\frac{\partial L}{\partial w}: \quad 2V\vec{w} - 2\lambda\vec{w} = 0$$

$$\Rightarrow \quad V\vec{w} = \lambda\vec{w}$$

因此 $\vec{w}$ 是 V 的特征向量. λ 是特征值.