~~Reduce~~ reduce n-dim to k-dim

Dataset: $\{x^{(i)};\ i=1,\dots,m\}$

$$\vec{x}^{(i)} = \begin{pmatrix} x_1^{(i)} \\ \vdots \\ x_n^{(i)} \end{pmatrix}$$

~~每个分量~~ 标准化不同分量的均值、方差.

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

对每个值 $x_j^{(i)} \leftarrow x_j^{(i)} - \mu$ (使得均值 $\mu = 0$)

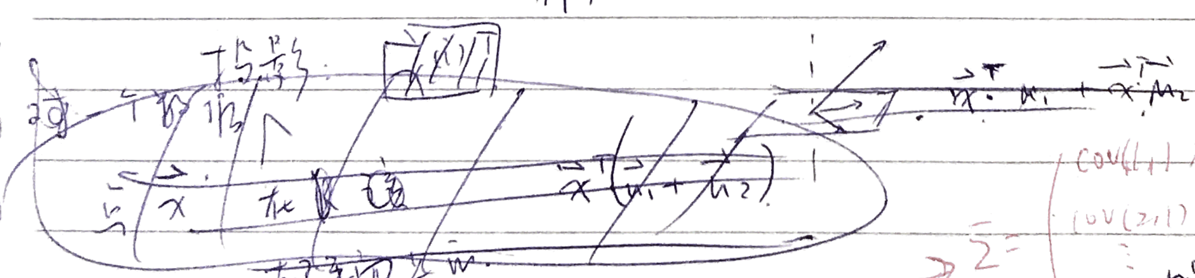
$$\sigma_j^2 = \frac{1}{m}\sum_{i=1}^{m}\left(x_j^{(i)}\right)^2$$

对每个分量 $x_j^{(i)} \leftarrow \dfrac{x_j^{(i)}}{\sigma_j}$ (使得方差 $\sigma_j = 1$)

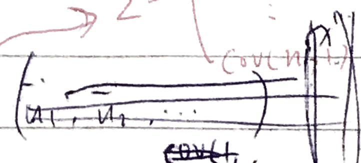
这种处理称为"白化"

目标: 找到一个 k 维超平面(子空间) (一组基 $u_1 \cdots u_k$)

s.t. 全体数据点(样本)在这个超平面上的投影 方差最大.



$$\Sigma = \begin{pmatrix} cov(1,1) & \cdots & cov(1,n) \\ cov(2,1) & & \vdots \\ \vdots & & \\ cov(n,1) & & cov(n,n) \end{pmatrix}$$



协方差矩阵.

$$\Sigma_{n\times n} = \frac{1}{m}\sum_{i=1}^{m} \vec{x}^{(i)}\,\vec{x}^{(i)T}$$

$\vec{u}_k$ 求出 Σ 的特征向量 u

Σ_{ij} 中元素的值越大,则说明对应下标的特征之间相关性越高

~~求~~ 找到 k-dim 子空间的一组基 $\vec{u}_1, \dots, \vec{u}_k$ 后

将 原来的数据点（样本）$x^{(i)} = \begin{pmatrix} x_1^{(i)} \\ \vdots \\ x_n^{(i)} \end{pmatrix}$ 投影到这个 k-dim 子空间

上 即可. 这样 新数据的 新的坐标点是

$$y^{(i)} = \begin{pmatrix} \vec{u}_1^T \vec{x}^{(i)} \\ \vec{u}_2^T \vec{x}^{(i)} \\ \vdots \\ \vec{u}_k^T \vec{x}^{(i)} \end{pmatrix} \in \mathbb{R}^k.$$

线性~~变换~~映射 $\Sigma: \mathbb{R}^n \to \mathbb{R}^k$.

原先的基：

$\vec{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ $\vec{e}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}$ ... $\vec{e}_n$

A

$x^{(i)}$ 是在这个基下的坐标.

~~变换~~的

线性映射后的基：$\vec{u}_1$, $\vec{u}_2$, ... $\vec{u}_k$

$y^{(i)}$ 是在这个基下的坐标