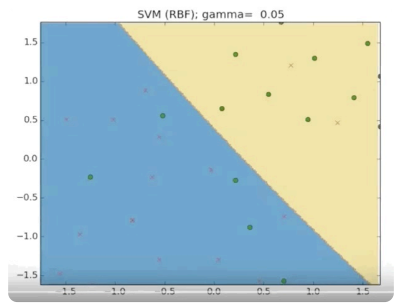
Outline: PCA
Unsupervised Methods

Recap: Kernel SVM

$$\max_{\alpha_1,\ldots,\alpha_n} \; -\frac{1}{2}\,\alpha^T \mathrm{Diag}(y)\, K\, \mathrm{Diag}(y)\,\alpha + \sum_{i=1}^{n}\alpha_i$$

$$\sum_{i=1}^{n}\alpha_i y_i = 0$$

$$0 \le \alpha_i \le \frac{C}{n} \quad \forall i$$



$$K_{ij} = k(x_i, x_j)$$

Classify $\hat{y} \rightarrow \mathrm{sign}\left(\sum \alpha_i^* k(x_i, x) + b^*\right)$

(RBF) Gaussian Kernel $K_{ij} = e^{-\frac{\|x_i - x_j\|_2^2}{2\sigma^2}}$

$$k(x_i, x) = e^{-\frac{\|x - x_i\|_2^2}{2\sigma^2}}$$

$$K_{ij} = (\langle x_i, x_j\rangle + 1)^p$$

Approximate K : Nystrom Method.

Principal Component Analysis

Dimension Reduction (DR)

Suppose we have $x_1, \ldots, x_n \in \mathbb{R}^d$

Goal: Transform $x_i \longrightarrow \theta_i \in \mathbb{R}^k$ where $k < d$, such that information loss is minimized.

Uses: (1) Visualization $k=1,2,3$

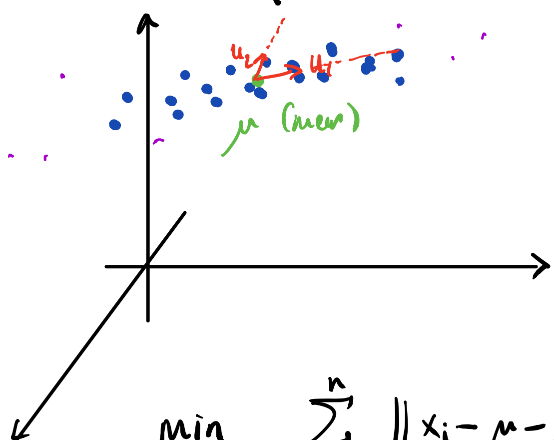
(2) Removing noise

(3) Computational Savings

DR can be supervised or unsupervised.

PCA is an unsupervised DR method

minimizes squared errors



$\underline{x}_i \approx \underline{\mu} + \underline{\underline{A}}\,\underline{\theta}_i$

$\mu \in \mathbb{R}^d$

$A \in \mathbb{R}^{d \times k}$ basis matrix

orthonormal columns

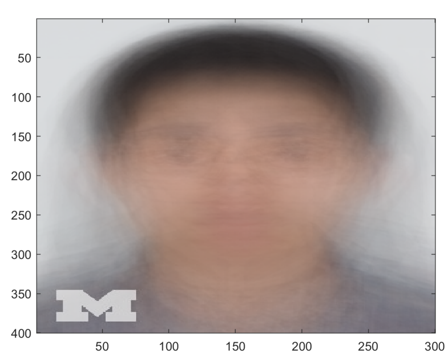
$\theta_i \in \mathbb{R}^k$ coefficients.

$$\min_{\substack{\mu \\ A \in \mathbb{R}^{d \times k} \\ \theta_1, \dots, \theta_n}} \sum_{i=1}^{n} \| x_i - \mu - A\theta_i \|_2^2$$

A orthonormal

where $A = [u_1, \dots, u_k]$

$A\theta = \sum_{i=1}^{k} \underline{u_i} \underline{\theta_i}$



$J(\mu, \theta_i, u_1)$

If $k=1$

$$\min_{\substack{\mu \\ u_1^T u_1 = 1 \\ \theta_i}} \overbrace{\sum_{i=1}^{n} \|x_i - \mu - u_1 \cdot \theta_i\|_2^2}^{J(\mu,\theta_i,u_1)} = \sum_{i=1}^{n} \|x_i - \mu\|_2^2 + \overbrace{\|u_1 \theta_i\|_2^2}^{\theta_i^2} - 2\theta_i u_1^T (x_i - \mu)$$

$$\frac{\partial}{\partial \theta_j} J = 2\theta_j - 2 \cdot u_1^T (x_j - \mu) = 0 \Rightarrow \boxed{\theta_j = u_1^T (x_j - \mu)}$$

$$J(\mu, \theta_i^*, u_1) = \sum_i \|x_i - \mu\|_2^2 - (u_1^T (x_i - \mu))^2$$

$$= \sum_i (x_i - \mu)^T (I - u_1 u_1^T) (x_i - \mu)$$

projection matrix

$(I - u_1 u_1^T)^2 = I - u_1 u_1^T$

$I - u_1 u_1^T - u_1 u_1^T + u_1 \underbrace{u_1^T u_1}_{1} u_1^T$

$$\frac{\partial}{\partial \mu} J(u_1, \mu_1, \theta_i^*) = \sum_i 2 \cdot (I - u_1 u_1^T) \cdot (x_i - \mu)$$

$$= 2 \cdot (I - u_1 u_1^T) \cdot \sum_{i=1}^{n} (x_i - \mu) = 2 (I - u_1 u_1^T) \cdot \left(\sum_{i=1}^{n} x_i - \mu n\right)$$

$$\boxed{\mu^* = \frac{1}{n} \sum_{i=1}^{n} x_i} + \beta u_1$$

for any β

$$\arg\min_{u_1^T u_1 = 1} \sum_{i=1}^{n} \|x_i - \bar{x}\|_2^2 - (u_1^T (x_i - \bar{x}))^2 = \arg\max_{u_1^T u_1 = 1} \sum_{i=1}^{n} (u_1^T \cdot (x_i - \bar{x}))^2$$

General Case: The solution is given by the eigendecomposition of the sample covariance matrix

$$S = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x}) \cdot (x_i - \bar{x})^T$$

$$S = V \Lambda V^T \quad \text{Eigenvalue Decomposition}$$

$U_1 = [u_1 \; \dots \; u_1]$

$U = [u_1, \dots, u_d]$

$$U^T U = U U^T = I \qquad \Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_d \end{bmatrix}$$

$\mu = \bar{x}$

$A = [u_1, \dots, u_k]$

$\theta_i = A^T (x_i - \bar{x})$

Principal component transform $\quad x \longrightarrow A^T (x - \bar{x}) \in \mathbb{R}^k$

j^{th} principal component $\quad \theta^{(j)} = u_j^T (x - \bar{x}) \in \mathbb{R}$

j^{th} principal eigenvector $\quad u_j \in \mathbb{R}^d$