- HW2 due Friday

Outline

LDA / QDA

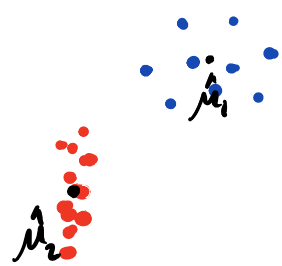
Naive Bayes

Recap LDA

$$\log \hat{\pi}_1 - \frac{1}{2}(x-\hat{\mu}_1)^T \Sigma^{-1} (x-\hat{\mu}_1) \underset{2}{\overset{1}{\gtrless}} \log \hat{\pi}_2 - \frac{1}{2}(x-\hat{\mu}_2)^T \Sigma^{-1}(x-\hat{\mu}_2)$$

$$\Sigma = \frac{1}{n}\sum_{i=1}^{n} (x_i-\hat{\mu}_{y_i})\cdot(x_i-\hat{\mu}_{y_i})$$

$$\underline{\hat{\mu}_k} = \frac{1}{n_i}\sum_{\substack{i=1 \\ i:\, y_i=k}}^{n} \underline{x_i}$$



- $\Sigma = I$, $\hat{\pi}_1 = \hat{\pi}_2 = \frac{1}{2}$ $\qquad -\frac{1}{2}\|x-\hat{\mu}_1\|_2^2 \gtrless -\frac{1}{2}\|x-\hat{\mu}_2\|_2^2$

- $\Sigma \neq I$

  Mahalanobis distance: $d_M(x) = \sqrt{(\underline{x}-\underline{\mu})^T \underline{\underline{\Sigma}}^{-1} (\underline{x}-\underline{\mu})}$

$\underline{\underline{\Sigma}}$ PSD $\Rightarrow \Sigma = \underline{\underline{U}}\,\underline{\underline{\Lambda}}\,\underline{\underline{U}}^T$ $\qquad \Lambda = \begin{bmatrix} \lambda_1 & \\ & \lambda_2 \end{bmatrix}$ $\qquad U^T U = I$

$$\Sigma^{-1} = U\Lambda^{-1}U^T \quad \text{since} \quad \Sigma^{-1}\Sigma = U\Lambda U^T U\Lambda^{-1}U^T$$

$$= U\Lambda\Lambda^{-1}U^T = UU^T = I$$

matrix squareroot: $\Sigma^{-\frac{1}{2}} = U\Lambda^{-1/2}U^T$

$$\Sigma^{-1/2}\cdot\Sigma^{-1/2} = U\Lambda^{-1/2}\underbrace{U^T U}\Lambda^{-1/2}U^T = \Sigma^{-1}$$

$$d_M = \sqrt{(x-\mu)^T\Sigma^{-1/2}\Sigma^{-1/2}(x-\mu)}$$

$$= \sqrt{\|\Sigma^{-1/2}(x-\mu)\|_2^2}$$

$$= \|\Sigma^{-1/2}(x-\mu)\|_2$$

Whitening

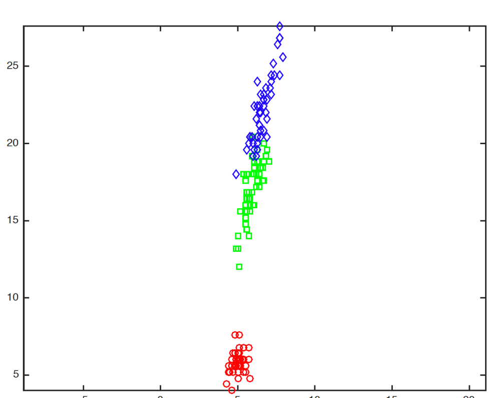
$$\Sigma = \frac{1}{n}\sum_{i=1}^n (x_i-\mu)(x_i-\mu)^T \qquad \tilde{x}_i = \Sigma^{-1/2}\cdot x_i$$

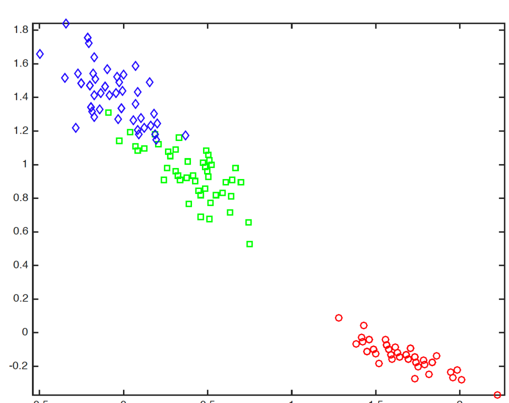
$$\tilde{\Sigma} = \frac{1}{n}\sum_{i=1}^n (\tilde{x}_i-\tilde{\mu})\cdot(\tilde{x}_i-\tilde{\mu})^T$$

$$= \frac{1}{n}\sum_{i=1}^n \Sigma^{-1/2}(x_i-\mu)\cdot(x_i-\mu)^T\cdot\Sigma^{-1/2}$$

$$= \frac{1}{n}\Sigma^{-1/2}\cdot\sum_{i=1}^n (x_i-\mu)(x_i-\mu)^T\,\Sigma^{-1/2}$$

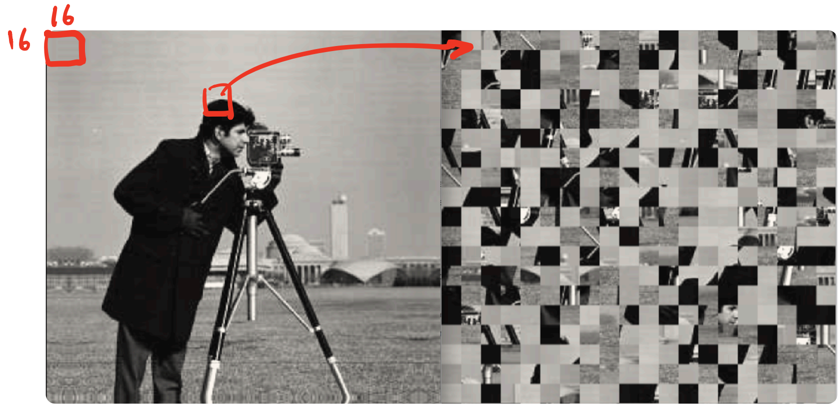
$$= \Sigma^{-1/2}\,\Sigma\,\Sigma^{-1/2} = \Sigma^{-1/2}\Sigma^{1/2}\cdot\Sigma^{1/2}\cdot\Sigma^{-1/2} = I$$

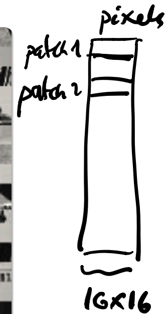




Σ $\qquad\qquad$ $\tilde{x}_i = \Sigma^{-1/2} x_i$

Example:





whitened $\underline{\underline{\Sigma}}^{-1/2} \underline{x}_i$ patch pixels

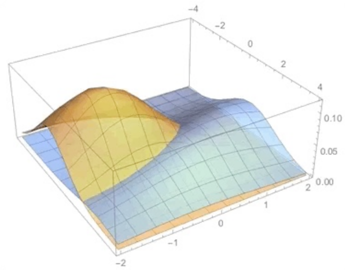
Quadratic Discriminant Analysis

$$\Sigma_1 \neq \Sigma_2 \neq \Sigma_3$$

Estimate $\quad \Sigma_k = \frac{1}{n_k} \sum_{i:\, y_i = k} (x_i - \hat{\mu}_k)(x_i - \hat{\mu}_k)^T$

$$\log \hat{\pi}_1 - \frac{1}{2}(x - \hat{\mu}_1)^T \Sigma_1^{-1} (x - \hat{\mu}_1) \underset{2}{\overset{1}{\gtrless}} \log \hat{\pi}_2 - \frac{1}{2}(x - \hat{\mu}_2)^T \Sigma_2^{-1} (x - \hat{\mu}_2)$$

expand : $\quad x^T Q x + b^T x + c \underset{2}{\overset{1}{\gtrless}} 0 \quad$ for some Q, b, c



K classes, $x_i \in \mathbb{R}^d$ $i=1,\dots,n$

LDA $\underline{\mu}_1, \underline{\mu}_2, \dots \underline{\mu}_K$, $\underline{\underline{\Sigma}}$

parameters $O(K \cdot d + d^2)$

QDA $\underline{\mu}_1, \underline{\mu}_2, \dots \underline{\mu}_K$, $\underline{\underline{\Sigma}}_1, \underline{\underline{\Sigma}}_2, \dots$

$O(K \cdot d + K d^2)$

kNN

parameters $O(dn)$

Naive Bayes (NB) Let $(X^{(1)}, \dots, X^{(d)}) \in \mathbb{R}^d$ denote random feature vectors and Y corresponding label.

NB assumes that $X^{(1)}, \dots, X^{(d)}$ are <u>independent</u> given Y

Example: Document classification

Classify into: sports, politics, business etc.

A document is $x = [x^{(1)}, \dots x^{(d)}]$

$$x^{(j)} = \begin{cases} 1 & \text{if } j\text{'th word occurs in document} \\ 0 & \text{otherwise} \end{cases}$$

vocabulary of d-words

Let $g_k(x)$ be the pmf of $X \mid Y=k$

NB assumption is $g_k(x) = \prod_{j=1}^{d} g_k^{(j)}(x)$ ← marginal pmf of $X^{(j)} \mid Y=k$

$(x_1, y_1), \ldots, (x_n, y_n)$ training data

$$\hat{\pi}_k = \frac{|\{i : y_i = k\}|}{n}$$

$\hat{g}_k^{(j)}$ estimate of $g_k^{(j)}$

NB classifier $\hat{f}(x) = \arg\max_k \hat{\pi}_k \cdot \prod_{j=1}^{d} g_k^{(j)}(x)$

Estimating $g_k^{(j)}$ (discrete Bernoulli model)

$$n_k = |\{i : y_i = k\}|$$

$$n_{kl}^{(j)} = |\{i : y_i = k \ \& \ x_i^{(j)} = z_l\}|$$

The natural (maximum likelihood) estimate of

$g_k^{(j)}(z_l) = P(X^{(j)} = z_l \mid Y=k)$ is

$$\hat{g}_k^{(j)}(z_l) = \frac{n_{kl}^{(j)}}{n_k}$$

Example $= \dfrac{\text{\# "win" occurs in sports}}{\text{\# sports documents}}$

Laplace Smoothing: adding 1 to counts