Lecture 10: Linear Discriminant Analysis

LDA

Model $\underset{\sim}{X} | Y$ and Y

equiv. to i.e. gives the same classifier

$$\delta_c(\underset{\sim}{x}) \equiv \mathbb{P}(Y=c \mid \underset{\sim}{X}=\underset{\sim}{x}) \equiv \mathbb{P}(\underset{\sim}{X}=\underset{\sim}{x} \mid Y=c) \cdot \mathbb{P}(Y=c)$$

LDA: $\left\{ \begin{array}{l} \underset{\sim}{X} | Y \text{ is Normal / Gaussian} \\ Y \text{ is discrete} \end{array} \right.$

$\boxed{p=1}$ so X is a univar. RV

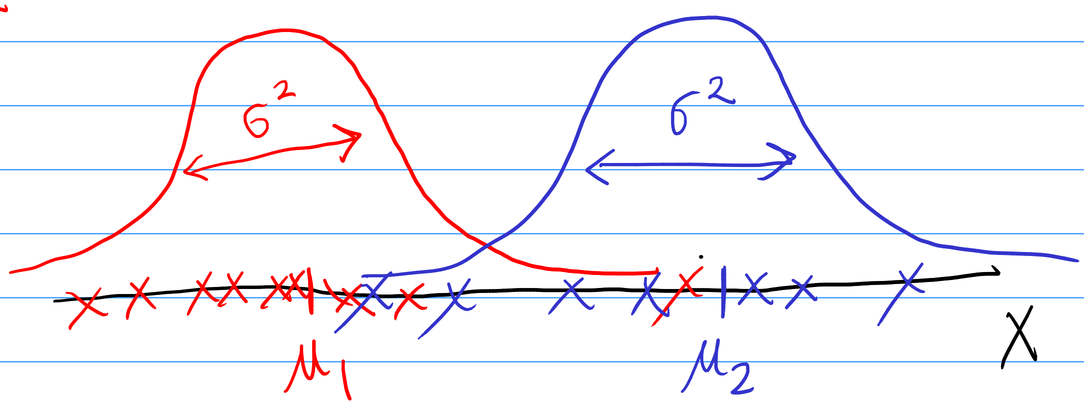
generically: $c \in \{1, 2, 3, \ldots, K\}$

LDA: $X | Y=c \sim N(\mu_c, \sigma^2)$ shared among classes

$\mathbb{P}(Y=c) = \pi_c$ where $\pi_c \geq 0$

and $\sum_{c=1}^{K} \pi_c = 1$

Ex. $K=2$



This model reduces learning δ_c's to learning μ_c, σ^2, π_c

$$\delta_c(x) \equiv P(Y=c \mid X=x)$$

$N(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$$

$$\equiv \underbrace{P(X=x \mid Y=c)}_{N(\mu_c,\, \sigma^2)} \underbrace{P(Y=c)}_{\pi_c}$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu_c)^2\right) \pi_c$$

$$\hat{f}(x) = \underset{c}{\arg\max}\ \delta_c(x)$$

take log

$$\equiv \underbrace{-\frac{1}{2}\log 2\pi - \frac{1}{2}\log(\sigma^2)}_{\text{doesn't depend on } c} - \frac{1}{2\sigma^2}(x-\mu_c)^2 + \log(\pi_c)$$

$$\equiv -\frac{1}{2\sigma^2}(x^2 + \mu_c^2 - 2\mu_c x) + \log \pi_c$$

$$\equiv -\cancel{\frac{x^2}{2\sigma^2}} - \frac{\mu_c^2}{2\sigma^2} + \frac{\mu_c x}{\sigma^2} + \log \pi_c$$

$$\equiv \underbrace{\left(\frac{\mu_c}{\sigma^2}\right)}_{\beta} x + \underbrace{\log \pi_c - \frac{\mu_c^2}{2\sigma^2}}_{\beta_0}$$

$$\equiv \beta_0 + \beta x$$

linear!

So LDA:

$$\delta_c(x) = \frac{\hat{\mu}_c}{\hat{\sigma}^2}x + \log\hat{\pi}_c - \frac{\hat{\mu}_c^2}{2\hat{\sigma}^2}$$

need to estimate: μ_c, σ^2, π_c as $\hat{\mu}_c, \hat{\sigma}^2, \hat{\pi}_c$ from my traing data

$\hat{\pi}_c$ = % of data in class c

$\hat{\mu}_c$ = mean of Xs in class c

$\hat{\sigma}^2$ = pooled variance of Xs

MLE

Pooled var: X s and Ys

$$S^2_{pooled} = \frac{\sum_n (X_n - \bar{X})^2 + \sum_m (Y_m - \bar{Y})^2}{N+M-2}$$

What about P > 1?

mean vector

covariance matrix

$$\underset{\sim}{X} = \underset{\sim}{x} \mid Y = c \sim N(\mu_c, \Sigma)$$

$$P(Y=c) = \pi_c, \quad \pi_c \geq 0, \quad \sum_c \pi_c = 1$$

$N(\mu, \Sigma)$, $\mu = (\mu_1, \dots, \mu_p)$

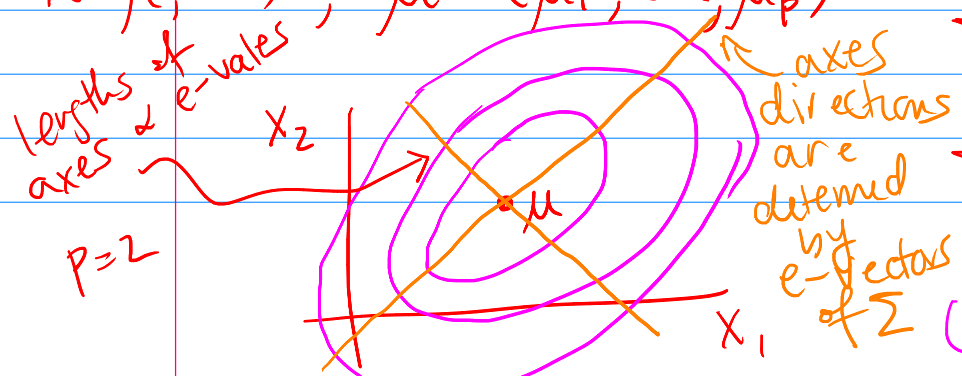
PXP

$\Sigma_{ij} = Cov(X_i, X_j)$

$\Sigma_{ii} = Cov(X_i, X_i)$

$= Var(X_i)$

↳ symmetric



Σ is positive definite ($\sigma^2 > 0$)

defn: $x^T \Sigma x > 0 \quad \forall x \neq 0$

for symmetric matrices: PD = e-vals > 0

uni: $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$

multivariate:

$$f(\underset{\sim}{x}) = (2\pi)^{-p/2} \det(\Sigma)^{-1/2} \exp\left(-\frac{1}{2}(\underset{\sim}{x}-\mu)^T \Sigma^{-1} (\underset{\sim}{x}-\mu)\right)$$

As before:

$$\delta_c(\underset{\sim}{x}) = \mathbb{P}(\underset{\sim}{X} = \underset{\sim}{x} \mid Y = c)\,\mathbb{P}(Y = c)$$

$$= (2\pi)^{-p/2} \det(\Sigma) \exp\left(-\frac{1}{2}(\underset{\sim}{x}-\mu_c)^T \Sigma^{-1} (\underset{\sim}{x}-\mu_c)\right)\pi_c$$