PS5841

Data Science in Finance & Insurance

Discriminant Analysis

Yubo Wang

Spring 2022

Columbia University

Bayes Classifier

- Bayes Classifier

$$p_k(\boldsymbol{x}) = \frac{\pi_k f_k(\boldsymbol{x})}{\sum_{l=1}^{K} \pi_l f_l(\boldsymbol{x})}$$

- Features of a response from the k-th class
  - Linear Discriminant Analysis

$$\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma})$$

  - Quadratic Discriminant Analysis

$$\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

COLUMBIA UNIVERSITY

Example: 1D Feature (1)

$$p_k(x) = \frac{\pi_k \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left[-\frac{1}{2\sigma_k^2}(x-\mu_k)^2\right]}{\sum_{l=1}^{K} \pi_l \frac{1}{\sqrt{2\pi}\sigma_l} \exp\left[-\frac{1}{2\sigma_l^2}(x-\mu_l)^2\right]}$$

$$\ln p_k(x) = \ln \pi_k - \ln\sqrt{2\pi} - \ln\sigma_k - \frac{1}{2\sigma_k^2}\left(x^2 - 2\mu_k x + \mu_k^2\right) - \ln D$$

$$\hat{y} = \operatorname*{argmax}_k p_k(x)$$

$$\hat{y} = \operatorname*{argmax}_k \left[\ln \pi_k - \ln\sigma_k - \frac{1}{2\sigma_k^2}\left(x^2 - 2\mu_k x + \mu_k^2\right)\right]$$

COLUMBIA UNIVERSITY

Example: 1D Feature (2)

$$\hat{y} = \underset{k}{\operatorname{argmax}} \left[\ln \pi_k - \ln \sigma_k - \frac{1}{2\sigma_k^2} \left(x^2 - 2\mu_k x + \mu_k^2 \right) \right]$$

- If features of a response from the k-th class

$$X \sim \mathcal{N}(\mu_k, \sigma^2)$$

$$\hat{y} = \underset{k}{\operatorname{argmax}} \left[\ln \pi_k + \frac{\mu_k x}{\sigma^2} - \frac{\mu_k^2}{2\sigma^2} \right]$$

- If $K = 2$, and $\pi_1 = \pi_2$
  - classify x as class 1 when $\mu_1 x - 0.5\mu_1^2 > \mu_2 x - 0.5\mu_2^2$
  - decision boundary $x = \frac{1}{2}(\mu_1 + \mu_2)$

COLUMBIA UNIVERSITY

Example: 1D Feature (3)

$$\hat{y} = \underset{k}{\operatorname{argmax}}\left[\ln \pi_k - \ln \sigma_k - \frac{1}{2\sigma_k^2}\left(x^2 - 2\mu_k x + \mu_k^2\right)\right]$$

- If features of a response from the k-th class

$$\mathrm{X} \sim \mathcal{N}(\mu_k, \sigma_k^2)$$

$$\hat{\mathrm{y}} = \underset{\mathrm{k}}{\operatorname{argmax}}\left[\ln \pi_k - \ln \sigma_k - \frac{1}{2\sigma_k^2}(x - \mu_k)^2\right]$$

- If $K = 2$, and $\pi_1 = \pi_2$
  - classify x as class 1 when

$$-\ln \sigma_1 - \frac{(x - \mu_1)^2}{2\sigma_1^2} > -\ln \sigma_2 - \frac{(x - \mu_2)^2}{2\sigma_2^2}$$

Columbia University

LDA (1)

- Features of a response from the k-th class

$$\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma})$$

- Likelihood

$$f_k(\boldsymbol{x}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_k)\right]$$

- Posterior

$$p_k(\boldsymbol{x}) = \frac{\pi_k f_k(\boldsymbol{x})}{\sum_{l=1}^{K} \pi_l f_l(\boldsymbol{x})}$$

$$= \frac{1}{D} \frac{\pi_k}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_k)\right]$$

LDA (2)

- Classification

$$\hat{y} = \underset{k}{\operatorname{argmax}} \left(\ln \pi_k + \boldsymbol{x}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_k - \frac{1}{2} \boldsymbol{\mu}_k^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_k \right)$$

- Discriminant function

$$\delta_k(\boldsymbol{x}) = \ln \pi_k + \boldsymbol{x}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_k - \frac{1}{2} \boldsymbol{\mu}_k^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_k$$

- Decision boundary (linear in $\boldsymbol{x}$)

$$\{\boldsymbol{x} | \delta_k(\boldsymbol{x}) = \delta_l(\boldsymbol{x})\}$$

COLUMBIA UNIVERSITY

Estimation

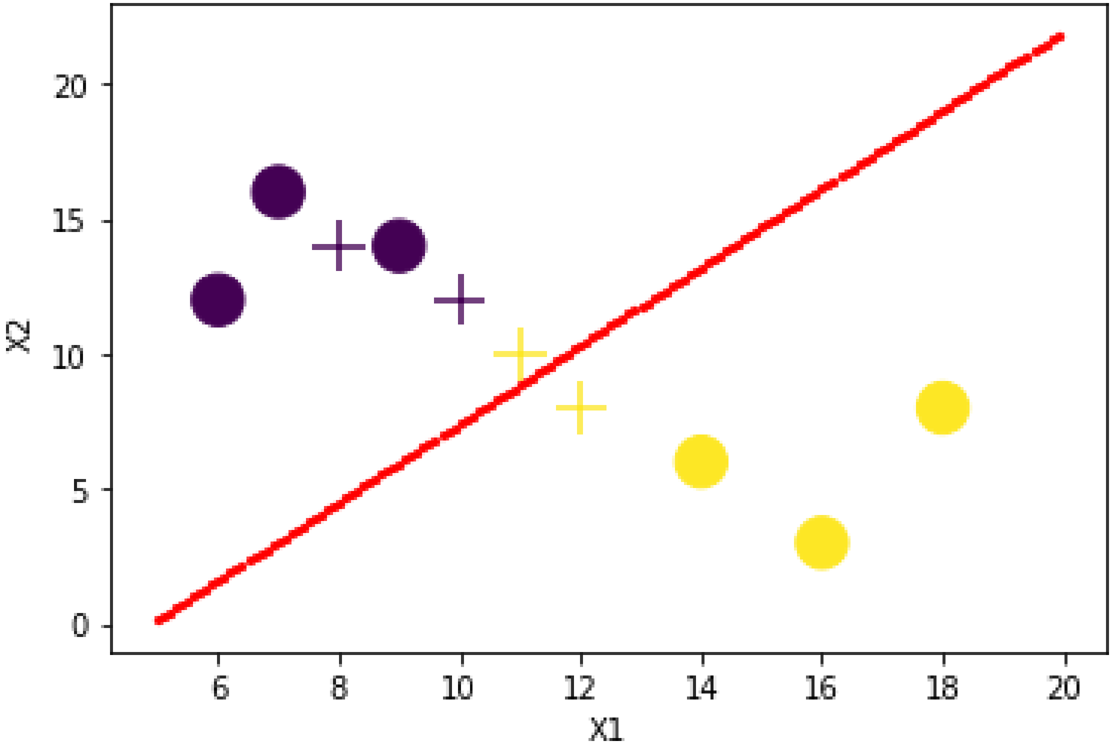
1 x 1
$$\hat{\pi}_k = \frac{n_k}{N}$$

p x 1
$$\hat{\boldsymbol{\mu}}_k = \frac{1}{n_k} \sum_{i:y_i=k} \boldsymbol{x}_i = \left(\frac{1}{n_k} \sum_{i:y_i=k} x_{i1}, \dots, \frac{1}{n_k} \sum_{i:y_i=k} x_{ip} \right)^T$$

p x p
$$\hat{\boldsymbol{\Sigma}} = \frac{1}{N-K} \sum_{k=1}^{K} \sum_{i:y_i=k} (\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}_k)(\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}_k)^T$$
$$= \frac{(n_1 - 1)\hat{\boldsymbol{\Sigma}}_1 + \cdots + (n_K - 1)\hat{\boldsymbol{\Sigma}}_k}{(n_1 - 1) + \cdots + (n_K - 1)}$$

COLUMBIA UNIVERSITY

LDA Decision Boundary



For additional emphasis, can adjust the discrimination threshold

QDA (1)

- Features of a response from the k-th class

$$\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

- Likelihood

$$f_k(\boldsymbol{x}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}_k|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_k)\right]$$

- Posterior

$$p_k(\boldsymbol{x}) = \frac{\pi_k f_k(\boldsymbol{x})}{\sum_{l=1}^{K} \pi_l f_l(\boldsymbol{x})}$$

$$= \frac{1}{D} \frac{\pi_k}{(2\pi)^{p/2}|\boldsymbol{\Sigma}_k|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_k)\right]$$

COLUMBIA UNIVERSITY

QDA (2)

- Classification

$$\hat{y} = \operatorname*{argmax}_k \left(\ln \pi_k - \frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_k) - \frac{1}{2} \ln |\boldsymbol{\Sigma}_k| \right)$$

- Discriminant function

$$\delta_k(\boldsymbol{x}) = \ln \pi_k - \frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_k) - \frac{1}{2} \ln |\boldsymbol{\Sigma}_k|$$

- Decision boundary (quadratic in $\boldsymbol{x}$)

$$\{\boldsymbol{x} | \delta_k(\boldsymbol{x}) = \delta_l(\boldsymbol{x})\}$$

COLUMBIA UNIVERSITY

Estimation

1 x 1

$$\hat{\pi}_k = \frac{n_k}{N}$$

p x 1

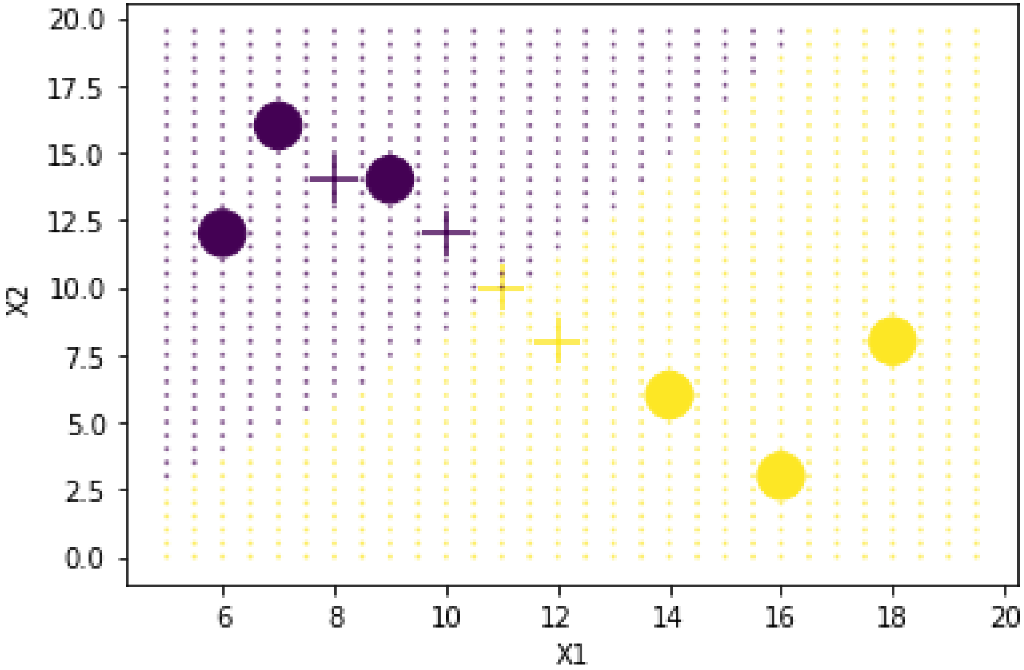
$$\hat{\boldsymbol{\mu}}_k = \frac{1}{n_k} \sum_{i:y_i=k} \boldsymbol{x}_i = \left(\frac{1}{n_k} \sum_{i:y_i=k} x_{i1}\,, \ldots, \frac{1}{n_k} \sum_{i:y_i=k} x_{ip} \right)^T$$

p x p

$$\hat{\boldsymbol{\Sigma}}_k = \frac{1}{n_k - 1} \sum_{i:y_i=k} (\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}_k)(\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}_k)^T$$

COLUMBIA UNIVERSITY

QDA Decision Boundary



For additional emphasis, can adjust the discrimination threshold

LDA vs QDA

- Features of a response from the k-th class
  - Linear Discriminant Analysis $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma})$
    - Common cov matrix
    - Fewer parameters, less flexible
    - Lower variance, more suitable for small training sets
  - Quadratic Discriminant Analysis $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$
    - Class-specific cov matrix
    - More parameters, more flexible
    - Higher variance, more suitable for large training sets

COLUMBIA UNIVERSITY

That was

COLUMBIA UNIVERSITY