High-dimensional data analysis

Lecture 5

Discriminant Analysis

Innopolis University

Fall 2018

Discriminant Analysis

- Goals of Multivariate data analysis:
  - understand the structure in the data;
  - summarize data in simpler ways;
  - find the relationship between parts of the data;
  - **make decisions based on the data.**

Discriminant Analysis

- Find projection of data which
  - Minimizes variability inside each class
  - Maximizes variability between classes

PCA vs. CCA vs. DA

- PCA – all data in one group
- CCA – <u>features</u> split into two groups
- DA – <u>samples</u> split into *n* groups.

Fischer’s Discriminant

- $\boldsymbol{X}^c, \boldsymbol{e} \in \mathbb{R}^d, \quad \boldsymbol{X}^c \sim (\boldsymbol{\mu}_c, \Sigma_c), \quad c = 1..k$
- $b(\boldsymbol{e}) = \sum \left| \boldsymbol{e}^T (\boldsymbol{\mu}_c - \overline{\boldsymbol{\mu}}) \right|$ – between-class variability
- $w(\boldsymbol{e}) = \sum \operatorname{var}\left(\boldsymbol{e}^T \boldsymbol{X}^c \right)$ – within-class variability

$$d = \max_{\|\boldsymbol{e}\|=1} \frac{b(\boldsymbol{e})}{w(\boldsymbol{e})}$$ – Fischer’s discriminant

$$\boldsymbol{\eta} = \operatorname{argmax}(d)$$ – discriminant direction

Fischer’s Discriminant

- $\boldsymbol{X}^c, \boldsymbol{e} \in \mathbb{R}^d, \quad \boldsymbol{X}^c \sim (\boldsymbol{\mu}_c, \Sigma_c), \quad c = 1..k$
- $B = \sum(\boldsymbol{\mu}_c - \overline{\boldsymbol{\mu}})(\boldsymbol{\mu}_c - \overline{\boldsymbol{\mu}})^T$
- $W = \sum \Sigma_c$
- Then:
  - $b(\boldsymbol{e}) = \boldsymbol{e}^T B \boldsymbol{e}$
  - $w(\boldsymbol{e}) = \boldsymbol{e}^T W \boldsymbol{e}$
  - $d = \max \mathrm{eigenvalue}(W^{-1}B)$
  - $\boldsymbol{\eta}$ is the corresponding eigenvector.

Fisher's Linear Discriminant Rule

- We have k classes with distributions $(\boldsymbol{\mu}_c, \Sigma_c)$
- $\boldsymbol{X}$ is random vector from one of the classes
- $\boldsymbol{\eta}$ is the discriminant direction
- Then:

$$\text{label}(X) = \underset{l=1..k}{\operatorname{argmin}} \left| \boldsymbol{\eta}^T (\boldsymbol{X} - \boldsymbol{\mu}_l) \right|$$

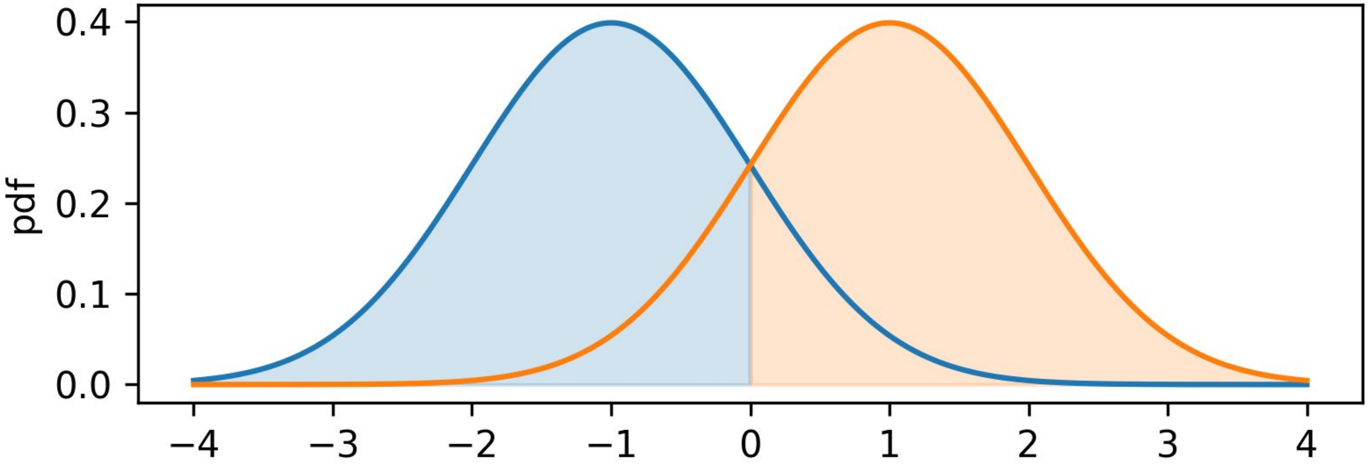
Sample case

- $\boldsymbol{X}_i^c, \boldsymbol{e} \in \mathbb{R}^d,\ i = 1..n_c,\ c = 1..k$
- $X_i^c \sim \mathrm{Sam}(\boldsymbol{\mu}_c, S_c), \quad \overline{\boldsymbol{\mu}} = \sum_c \boldsymbol{\mu}_c / k$
- $B = \sum(\boldsymbol{\mu}_c - \overline{\boldsymbol{\mu}})(\boldsymbol{\mu}_c - \overline{\boldsymbol{\mu}})^T$
- $W = \sum S_c$
- $b(\boldsymbol{e}) = \sum\left|\boldsymbol{e}^T(\boldsymbol{\mu}_c - \overline{\boldsymbol{\mu}})\right| = \boldsymbol{e}^T B \boldsymbol{e}$
- $w(\boldsymbol{e}) = \sum \frac{1}{n_c - 1} \sum\left[\boldsymbol{e}^T(\boldsymbol{X}_i^c - \boldsymbol{\mu}_c)\right]^2 = \boldsymbol{e}^T W \boldsymbol{e}$

Normal Discriminant Rule

- We have 2 classes with normal distributions $N\left(\mu_c, \sigma^2\right), \quad \mu_1 \neq \mu_2$
- X is random variable from one of the classes
- We choose class with max. probability of producing X:

$$\mathrm{label}(X) = \underset{c=1,2}{\operatorname{argmax}} \left| P\left(X \in \mathrm{class}(c)\right) \right| = \underset{c=1,2}{\operatorname{argmax}} |\mathrm{pdf}_c(X)|$$



Normal Discriminant Rule

- We have 2 classes with normal distributions $N(\boldsymbol{\mu}_c, \Sigma)$, $\quad \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$
- $\boldsymbol{X}$ is random vector from one of the classes
- $h(\boldsymbol{X}) = \left[\boldsymbol{X} - \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)\right]^T \Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ – decision function
- Then $\mathrm{pdf}_1(\boldsymbol{X}) > \mathrm{pdf}_2(\boldsymbol{X})$ if and only if $h(\boldsymbol{X}) > 0$

Bayesian Discriminant Rule

- We know not only likelihood functions, but also prior probabilities π_c
  - *E.g.*, we know that one class is much rarer than the other

- Normal Discriminant Rule: maximize $P\big(\boldsymbol{X} \in \text{class}(c)\big) = \text{pdf}_c(\boldsymbol{X})$
- $h(\boldsymbol{X}) = \left[\boldsymbol{X} - \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)\right]^T \Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$

- Bayesian Discriminant Rule: maximize $P\big(\boldsymbol{X} \in \text{class}(c)\big) = \ \text{pdf}_c(\boldsymbol{X})\,\pi_c$
- $h(\boldsymbol{X}) = \left[\boldsymbol{X} - \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)\right]^T \Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) - \log\frac{\pi_2}{\pi_1}$