B4M36SAN

LDA and Logistic Regression

Anh Vu Le & Jan Bláha

Outline

- From Linear to Logistic regression
- How is LDA approach different?
  - What assumptions are relaxed in QDA?
- LDA as dimensionality reduction technique

From Linear to Logistic regression

What we have

- Linear model
  - Simple to fit
  - Predicts on the **scale from** 0 to ∞

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots$$

What we want

- To predict **probabilities**
  - **Scale 0 to 1**

Examples of odds:

Prob. of heads in a coin toss *p=0.5*

$$\text{odds} = \frac{0.5}{1-0.5} = \mathbf{1}$$

Prob. of 6 in a dice roll *p=1/6*

$$\text{odds} = \frac{1/6}{1-1/6} = \mathbf{0.2}$$

Prob. of passing SAN *p=0.9*

$$\text{odds} = \frac{0.9}{1-0.9} = \mathbf{9}$$

Adapter function:

Sigmoid

$$p = \frac{1}{1+e^{-Y}}$$

$$p = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots)}}$$

Derivation:

$$Y = \frac{p}{1-p} = \text{odds}$$

$$Y = \log\left(\frac{p}{1-p}\right) = \text{logit}$$

Now express p to get the sigmoid

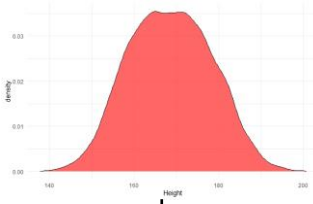
LDA approach

- To which class *c* the instance with features *x* most likely belongs to?

$$P(Y = y_c | X = x)$$

$$P(sex = Male | Height = 184)$$

- Bayes formula
- $P(sex = Male | Height = 184) =$



$P(male) = 0.45$
$P(female) = 0.55$

$$\frac{P(Height{=}184 | sex{=}male)\, P(sex{=}male)}{\sum_{sex_i} P(Height{=}184 | sex{=}sex_i)\, P(sex{=}sex_i)}$$

Assuming same „sizes“

same for all classes

$P(Height = 184 | sex = male) \sim N(\mu_{male}, \sigma^2)$ → **LDA/QDA**

LDA summary

- Assumes data have a normal distribution (within each class)
  - Decision boundary = region of equal posterior probability
  - Boundary linear if covariances are the same

- But what are the “linear discriminants”?
  - Let’s check ***Elhabian_LDA09.pdf***

```
Call:
lda(Gender ~ ., data = life_data)

Prior probabilities of groups:
Female   Male
   0.5    0.5

Group means:
         Height   Weight
Female 161.8203 61.68048
Male   175.3269 84.90736

Coefficients of linear discriminants:
              LD1
Height -0.06766107
Weight  0.15646959
```

- macbookpro_1477920514956.jpeg (800×533) (gadgets360cdn.com)
- 19" CRT monitor SAMSUNG SyncMaster 997MB LH19ISBBS/EDC slono | kak.cz
- 87b9db78647912d4bf3bca0309e86648.png (480×359) (pinimg.com)