

STAT 432: Basics of Statistical Learning

Quiz I - Extra Review Questions

Exercise 1

Consider a categorical response Y which takes possible values 0 and 1 as well as two numerical predictors X_1 and X_2 . Recall that

$$p(x) = P[Y = 1 \mid X = x]$$

Consider the model

$$\log\left(\frac{p(x)}{1-p(x)}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

and estimated coefficients

- $\hat{\beta}_0 = 1$
- $\hat{\beta}_1 = 2$
- $\hat{\beta}_2 = 3$
- $\hat{\beta}_3 = 4$

- (a) Provide a classification when $x_1 = 1$, $x_2 = -1$, and $x_3 = 1$.
(b) Provide an estimate for $P[Y = 1 \mid X_1 = 0, X_2 = 0, X_3 = 0]$.
(c) Provide an estimate for $P[Y = 1 \mid X_1 = 1, X_2 = 1, X_3 = -1]$.
(d) Provide an estimate for $P[Y = 0 \mid X_1 = -2, X_2 = 0, X_3 = 0]$.

Exercise 2

Recall that the pdf of a Normal random variable, X , with mean μ and variance σ^2 is given by

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

Consider the following estimates and information from data for a two-class classification problem:

	Class A	Class B
$\hat{\mu}_A = 22$	$\hat{\mu}_B = 25$	
$\hat{\sigma}_A^2 = 5$	$\hat{\sigma}_B^2 = 3$	
$n_A = 60$	$n_B = 40$	

Assume all estimates given are unbiased. All estimates used should be unbiased.

- (a) Use LDA to estimate the probability $P[Y = B \mid X = 23]$.
(b) Use QDA to estimate the probability $P[Y = B \mid X = 23]$.
(c) Calculate the LDA decision boundary.

Solutions

Exercise 1

- (a) 1.
- (b) 0.7311
- (c) 0.8808.
- (d) 0.9526.

Exercise 2

- (a) 0.318154.
- (b) 0.3281159.
- (c) 24.068.