STAT 430: Basics of Statistical Learning

Quiz I - Review Questions

Exercise 1

Consider the following joint probability mass function for X and Y.

	$X = 0$	$X = 1$	$X = 2$
$Y = 0$	0.4	0.1	0.0
$Y = 1$	0.1	0.3	0.1

(a) Are X and Y independent? Justify your answer.

(b) What is the distribution of $X \mid Y = 1$?

(c) Calculate $P[Y = 1 \mid X = 1]$.

Exercise 2

Consider a normal random variable X,

$$X \sim N(\mu = 3, \sigma^2 = 25).$$

(a) Consider $x = 2$ and $x = 5$. At which of these values does the density take a larger value?

(b) Calculate the value of the density at $x = 3$.

Exercise 3

Consider two models for the weight, Y, of an individual in kilograms.

$$Y = \beta_0 + \beta_1 x_1 + \epsilon \quad \text{(Model 1)}$$

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon \quad \text{(Model 2)}$$

where x_1 is the height of the individual in meters and x_2 is a dummy variable such that

$$x_2 = \begin{cases} 1 & \text{female} \\ 0 & \text{male} \end{cases}$$

and for both models

$$\epsilon \sim N(0, \sigma^2).$$

(a) What is the mean weight of a female individual using Model 1?

(b) What is the mean weight of a female individual using Model 2?

(c) Suppose the true parameter values for Model 2 are given by

$$\begin{aligned}\beta_0 &= 0\\ \beta_1 &= 50\\ \beta_2 &= -5\\ \sigma^2 &= 4.\end{aligned}$$

What is the probability that a male individual that is 1.7 meters tall weighs more than 89 kilograms?

Exercise 4

Consider five regression models, each fit to the same training data.

- Model 1: `y ~ 1`
- Model 2: `y ~ x1`
- Model 3: `y ~ x1 + x2 + x3`
- Model 4: `y ~ x1 + x2 + x3 + I(x1 ^ 2) + I(x2 ^ 2)`
- Model 5: `y ~ x1 * x2 * x3 + I(x1 ^ 2) + I(x2 ^ 2)`

Assume that the true model is given by

$$Y = 5 + 3x_1 + 2x_2 - 4x_3 + \epsilon$$

where $\epsilon \sim N(0, \sigma^2 = 4)$.

(a) Which model is the least flexible?

(b) Which model is the most flexible?

(c) How many predictors are used in model 5?

(d) Which model will have the lowest training RMSE?

(e) Which model will have the lowest test RMSE?

Exercise 5

Refer to the previous exercise. Which model will ...

(a) Be the least biased?

(b) Be the most biased?

(c) Be the least variable?

(d) Be the most variable?

Exercise 6

Consider the following dataset:

x	y
1	3
2	3
3	6
4	9
5	10
6	12
7	16
8	16
9	20
10	20

(a) Use KNN with $k = 3$ to predict y when $x = 9.5$.

(b) Use KNN with $k = 5$ to predict y when $x = 9.5$.

(c) Consider the above data as training data. Also consider a test dataset with a single observation $(1.5, 5)$. Use KNN with $k = 2$ to train a model. Calculate test RMSE for this model.

Exercise 7

Consider the following joint probability mass function for X and Y.

	$X = 0$	$X = 1$	$X = 2$
$Y = 0$	0.4	0.1	0.0
$Y = 1$	0.1	0.3	0.1

(a) Use the Bayes Classifier to obtain a classification when $x = 0$.

(b) Use the Bayes Classifier to obtain a classification when $x = 2$.

Exercise 8

Consider a categorical response Y which takes possible values 0 and 1 as well as a single numerical predictor X. Recall that

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

Consider the model

$$\log\left(\frac{p(x)}{1 - p(x)}\right) = \beta_0 + \beta_1 x$$

and estimated coefficients

- $\hat{\beta}_0 = 2.4$
- $\hat{\beta}_1 = -1.2$

(a) Provide a classification when $x = 2.2$ that attempts to minimize the classification error.

(b) Calculate an estimate of $P[Y = 1 \mid X = 1]$

(c) Calculate an estimate of $P[Y = 0 \mid X = 2.5]$

(d) Find a value c that splits x into regions classified as 1 and 0. Define these regions.

Exercise 9

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

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

and estimated coefficients

- $\hat{\beta}_0 = 5$
- $\hat{\beta}_1 = -4$
- $\hat{\beta}_2 = -2$

(a) Derive and sketch the decision boundary for the classifier that results from this model. Shade the region that will be classified as **one**.

(b) Suppose the data used to estimate the coefficients has the response, y, flipped. That is, *one* becomes *zero*, and *zero* becomes *one*. What effect would this have on the:

- Decision Boundary?
- Classifications?
- Estimated coefficients?

Exercise 10

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
$\hat{\mu}_A = 2$	$\hat{\mu}_B = 4$
$\hat{\sigma}^2_A = 3$	$\hat{\sigma}^2_B = 5$
$n_A = 10$	$n_B = 30$

Use LDA to answer the following questions. Assume all estimates given are unbiased. All estimates used should be unbiased.

(a) Classify y when $x = 2.7$.

(b) Give an estimate of the probability $P[Y = B \mid X = 3.8]$.

(c) Find the decision boundary. (Hint: It should be a single value of x. The result may be somewhat surprising.)

Exercise 11

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 three-class classification problem:

Class A	Class B	Class C
$\hat{\mu}_A = 10$	$\hat{\mu}_B = 14$	$\hat{\mu}_C = 19$
$\hat{\sigma}_A^2 = 1$	$\hat{\sigma}_B^2 = 4$	$\hat{\sigma}_C^2 = 9$
$n_A = 10$	$n_B = 10$	$n_C = 20$

Assume all estimates given are unbiased.

(a) Use QDA to classify y when $x = 16$.

Exercise 12

(**ISL** 4.8) Suppose that we take a data set, divide it into equally-sized training and test sets, and then try out two different classification procedures.

First we use logistic regression and get an error rate of 20% on the training data and 30% on the test data. Next we use 1-nearest neighbors (i.e. $k = 1$) and get an average error rate (averaged over both test and training data sets) of 18%.

Based on these results, which method should we prefer to use for classification of new observations? Why?