Classification

Machine Learning 2023-2024 - UMONS
Souhaib Ben Taieb

1 Exercise 1

Suppose we collect data for a group of students in a statistics class with variables:

- X_1 = hours studied.
- X_2 = undergrad GPA.
- Y = receive an A.

We fit a logistic regression and produce estimated coefficients:

- $\hat{\beta}_0 = -6$
- $\hat{\beta}_1 = 0.05$
- $\hat{\beta}_2 = 1$

a) How would the model write and how do you interpret its coefficients ?

b) Estimate the probability that a student who studies for 40h and has an undergrad GPA of 3.5 obtains an A in the class.

c) How many hours would the above student need to study to have a 50% chance of getting an A in the class ?

Exercise 2

Consider the following dataset with $n = 8$ observations, three binary input features and a binary response.

X_1	X_2	X_3	Y
1	0	1	1
1	1	1	1
0	1	1	0
1	1	0	0
1	0	1	0
0	0	0	1
0	0	0	1
0	0	1	0

Assume we are using a naive Bayes classifier to predict the value of Y from the values of the other variables.

- a) What is $P\Big(Y = 1|X_1 = 1, X_2 = 1, X_3 = 0\Big)$?

- b) What is $P\Big(Y = 0|X_1 = 1, X_2 = 1\Big)$?

Now, suppose that we are using a joint Bayes classifier to predict the value of Y from the values of the other variables.

- c) What is $P\Big(Y = 1|X_1 = 1, X_2 = 1, X_3 = 0\Big)$?

- d) What is $P\Big(Y = 0|X_1 = 1, X_2 = 1\Big)$?

Exercise 3

This problem relates to the QDA model, in which the observations within each class are drawn from a normal distribution with a class specific mean vector and a class specific covariance matrix. We consider the simple case where $p = 1$; i.e. there is only one feature.

Suppose that we have K classes, and that if an observation belongs to the k^{th} class, then X comes from a one-dimensional normal distribution, $X \sim \mathcal{N}(\mu_k, \sigma_k^2)$. Prove that, in that case, the Bayes' classifier is not linear. Argue that it is in fact quadratic.

Exercise 4

Bob is playing a bar game, for which the principle is the following: While being blindfolded, Bob has to throw a dart at random on a target that only contains number between 0 and 1. Once he has thrown, he can take off the blindfold, and look at the target value x he got. Based on this value, the bartender secretly pours a beer with probability $0.2+0.4x$, a mojito with probability $0.6-0.4x$, and a glass of wine with probability 0.2. If Bob correctly guesses the beverage that has been served, he gets it for free, otherwise he is obliged to pay for it.

1. Depending on the target value obtained, what could be the optimal prediction that Bob could make and what would be the name of such a classifier ? You will need to derive the boundary decisions of the classifier.

2. What would be the misclassification error rate of this classifier ? Your answer should be a scalar.

Exercise 5

Suppose that you are given a set of n i.i.d. observations $\mathscr{D} = \{(x_i, y_i)\}_{i=1}^n$ where y_i is a categorical variable belonging to K categories, $\mathscr{Y} = \{C_1, ..., C_K\}$. You wish to fit a multiclass logistic regression model to $\mathscr{D}$, i.e.

$$\mathbb{P}(y_i = C_k | x_i) = p_k(x_i; \boldsymbol{\beta}) = \frac{e^{\beta^{(k)} x_i}}{\sum_{l=1}^K e^{\beta^{(l)} x_i}}$$

Write the expression of the conditional log-likelihood as a function of the data and the unknown coefficients $\boldsymbol{\beta}$.