Classification

Machine Learning 2021-2022 - UMONS
Souhaib Ben Taieb

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

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

2

Suppose that we wish to predict whether a given stock will issue a dividend this year ("Yes" or "No") based on X, last year's percent profit. We examine a large number of companies and discover that the mean value of X for companies that issued a dividend was $X = 10$, while the mean for those that didn't was $X = 0$. In addition, the variance of X for these two sets of companies was $\sigma^2 = 36$. Finally, 80% of companies issued dividends. Assuming that X follows a normal distribution, predict the probability that a company will issue a dividend this year given that its percentage profit was $X = 4$ last year.

Hint: Recall that the density function for a normal random variable is :

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(- \frac{1}{2\sigma^2}(x-\mu)^2 \Big)$$

You will need to use Bayes' theorem.

3

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.

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

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

4

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?

5

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.