Computational Statistics

Homework 4

-Subash Kharel

Mar 11, 2020

1. **Using a little bit of algebra, prove that (4.2) is equivalent to (4.3). In other words, the logistic function representation and logit representation for the logistic regression model are equivalent.**

$$p(X) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}} \quad \text{——— (4.2)}$$

$$\text{or,} \quad p(X)\left(1 + e^{\beta_0 + \beta_1 X}\right) = e^{\beta_0 + \beta_1 X}$$

$$\text{or,} \quad p(X) + p(X)e^{\beta_0 + \beta_1 X} = e^{\beta_0 + \beta_1 X}$$

$$\text{or,} \quad p(X) = e^{\beta_0 + \beta_1 X}\left(1 - p(X)\right)$$

$$\text{or,} \quad \frac{p(X)}{1 - p(X)} = e^{\beta_0 + \beta_1 X} \quad \text{——— (4.3)}$$

2. **It was stated in the text that classifying an observation to the class for which (4.12) is largest is equivalent to classifying an observation to the class for which (4.13) is largest. Prove that this is the case. In other words, under the assumption that the observations in the kth class are drawn from a N(μk, σ2) distribution, the Bayes' classifier assigns an observation to the class for which the discriminant function is maximized.**

$$p_k(x) = \frac{\pi_k \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x-\mu_k)^2\right)}{\sum_{t=1}^{k} \pi_t \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x-\mu_k)^2\right)} \quad \text{— (4.12)}$$

Taking log on both sides,

$$\log p_k(x) = \log\left(\frac{\pi_k \exp\left(-\frac{1}{2\sigma^2}(x-\mu_k)^2\right)}{\sum_{t=1}^{k} \pi_t \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x-\mu_t)^2\right)}\right)$$

$$= \log \pi_k - \frac{1}{2\sigma^2}(x-\mu_k)^2 - \log\left(\sum_{t=1}^{k} \pi_t \frac{1}{\sqrt{2\pi}\sigma} \exp(\ldots)\right)$$

Omitting last term which is independent of k.

$$= \log \pi_k - \frac{1}{2\sigma^2}\left(x^2 - 2x\mu_k + \mu_k^2\right)$$

$$= \log \pi_k - \frac{x^2}{2\sigma^2} + \frac{x\mu_k}{\sigma^2} - \frac{\mu_k^2}{2\sigma^2}$$

omitting terms independent of k,

$$\delta_k(x) = \frac{\mu_k x}{\sigma^2} - \frac{\mu_k^2}{2\sigma^2} + \log \pi_k \quad \text{—— (4.13)}$$

3. **We now examine the differences between LDA and QDA.**
   a. **If the Bayes decision boundary is linear, do we expect LDA or QDA to perform better on the training set? On the test set?**
      Answer: With linear decision boundary, QDA will perform better on training set because it will try to align to the data as close as possible. But, LDA will perform better on test set since, QDA have high chances of overfitting the training set resulting errors in linear testing set data.
   b. **If the Bayes decision boundary is non-linear, do we expect LDA or QDA to perform better on the training set? On the test set?**
      Answer: For a non-linear decision boundary, QDA is expected to perform better on both training and test set since it will properly capture the non-linearity of the data than LDA.
   c. **In general, as the sample size n increases, do we expect the test prediction accuracy of QDA relative to LDA to improve, decline, or be unchanged?**
      Answer: The prediction accuracy of QDA tends to increase if the size n increases. With large training data, the chance of overfitting gets reduced and prediction becomes more reliable with the test data.
   d. **True or False: Even if the Bayes decision boundary for a given problem is linear, we will probably achieve a superior test error rate using QDA rather than LDA because QDA is flexible enough to model a linear decision boundary. Justify your answer.**
      Answer: For non-linear methods, number of training samples matters the most. The lesser the training data, the more the chance of overfitting increases. We cannot confirm this statement without the information on number of training samples. With enough number of training samples, QDA can be expected to achieve superior test rate than LDA if the Bayes decision boundary is linear.

4. **Suppose we collect data for a group of students in a statistics class with variables X1 = hours studied, X2 = undergrad GPA, and Y = receive an A. We fit a logistic regression and produce estimated coefficient, βˆ0 = −6, βˆ1 = 0.05, βˆ2 = 1.**
   a. **Estimate the probability that a student who studies for 40 h and has an undergrad GPA of 3.5 gets an A in the class.**

Probability of getting A (P_A) $= \dfrac{1}{1+e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2)}}$

$= \dfrac{1}{1+e^{-(-6+0.05\times40+3.5\times1)}}$

$= \dfrac{1}{1+e^{0.5}}$

$= 0.37754$

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

Probability of getting A (P_A) $= \dfrac{1}{1+e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2)}}$

or, $0.5 = \dfrac{1}{1+e^{-(-6+0.05X_1+3.5\times1)}}$

or, $e^{-(0.05X_1-2.5)} = 1$

or, $0.05X_1 - 2.5 = 0$

or, $X_1 = \dfrac{2.5}{0.05}$

$= 50$ hours.

∴ The student should study for 50 hours.

5. **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 ˆσ2 = 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) = ((√ 1 /2π) σ^2) *e^(−(x−μ) ^2/(2σ^2)). You will need to use Bayes' theorem.**

Given,

$\bar{X} = 10$

$\hat{X} = 0$

$\sigma^2 = 36$

Percentage issuing dividends = 80%

$P = 0.8$

Last year percentage return = 4

Probability the company will issue dividend this year = ?

Using the formula,

$$P' = \frac{P e^{-\frac{1}{2\sigma^2}(x-\mu)^2}}{P e^{-\frac{1}{2\sigma^2}(x-\mu)^2} + (1-P) e^{-\frac{1}{2\sigma^2}(x-\mu_0)^2}}$$

$$= \frac{0.8\, e^{-(1/72)(4-10)^2}}{0.8 \times e^{-(1/72)(4-10)^2} + (1-0.8) \times e^{-(1/72)(4-0)^2}}$$

$\simeq 0.75$ Answer.

6. **This problem has to do with odds.**
   a. **On average, what fraction of people with odds of 0.37 of defaulting on their credit card payment will in fact default?**

Odds of defaulting $= 0.37$

or, $\frac{P}{1-P} = 0.37$

or, $P = 0.37 - 0.37P$

or, $P(1+0.37) = 0.37$

or, $P = \frac{0.37}{1+0.37} = 0.27$ Ans.

b. Suppose that an individual has a 16 % chance of defaulting on her credit card payment. What are the odds that she will default?

Probability $= 0.16$

odds $= \dfrac{0.16}{1 - 0.16} = 0.19$ Ans.