**INF552 Assignment3**

**3.7.4.**

*(a) Suppose that the true relationship between X and Y is linear. Consider the training residual sum of squares (RSS) for the linear regression, and also the training RSS for the cubic regression. Would we expect one to be lower than the other, would we expect them to be the same, or is there not enough information to tell? Justify your answer.*

Not enough information about the training data, so it’s hard to tell which one is better. However, since the true relationship between X and Y is linear, we may expect a lower RSS of linear regression than that of cubic regression.

*(b) Answer (a) using test rather than training RSS.*

In this case, we have not enough information to conclude. Maybe cubic regression will have a higher test due to the overfitting.

*(c) Suppose that the true relationship between X and Y is not linear, but we don’t know how far it is from linear. Consider the training RSS for the linear regression, and also the training RSS for the cubic regression. Would we expect one to be lower than the other, would we expect them to be the same, or is there not enough information to tell? Justify your answer.*

Cubic regression has lower train RSS than the linear fit because it is more flexible. The linear regression will not provide a good fit for the non-linear relationship.

*(d) Answer (c) using test rather than training RSS.*

Hard to tell. The statement says we don’t know “how far it is from linear”. So we don’t know the level of flexibility fit the data better. If it is closer to linear than cubic, the linear regression test RSS could be lower. Otherwise, the cubic regression RSS would be lower.

**4.7.3**

*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 kth class then X comes from a one-dimensional normal distribution, X ∼ N(μ k , σ k ). Recall that the density function for the one-dimensional normal distribution is given in (4.11). Prove that in this case, the Bayes’ classiﬁer is not linear. Argue that it is in fact quadratic.*

*Hint: For this problem, you should follow the arguments laid out in 2 Section 4.4.2, but without making the assumption that σ1^2 = . . . = σK^2.*

In this equation, the first tern is a distinct for each class, and the second one is the quadratic function of x. That is, the classifier is quadratic.

**4.7.7**

*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 σˆ 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.*

prior = Probability(Dividend = yes) = 0.8  
likelihood = density(X = 4 | Mean = 10, Variance = 36) = dnorm(4, mean = 10, sd = 6) = 0.0403285  
evidence = sum of prior \* likelihood for all classes = 0.8 \* dnorm(4, mean = 10, sd = 6) + (1 - 0.8) \* dnorm(4, mean = 0, sd = 6) = 0.042911  
posterior = prior \* likelihood / evidence = 0.7518525