EECS545 Lecture 5 Quiz Solutions

1. Suppose for a balanced dataset for binary classification (e.g., positive and negative examples are approximately 50% each for both train and test set) with n examples we run classification using K-nearest neighbors, with K $= n$. For simplicity, we assume that K is an odd number to make tie-breaking unambiguous. In this case, K-nearest neighbors is... :
   (a) The accuracy of this model would be higher than 90% over the test set.
   (b) n is a good choice for K.
   (c) K $= n$ leads to a high bias.
   (d) Leads to a classifier that ignores the input.

   Solution: (c) and (d). (a) The accuracy of this model would be approximately 50%, so (b) n is not a good choice for K.

2. Let's say we have four training examples $\{(x_1, x_2), y\} = [\{(-3, 3), 0\}, \{(4, 0), 1\}, \{(3, 4), 0\}, \{(0, -6), 1\}]$. When we provide (0, 0) as the query example and set K= 3 with L^2 norm as the distance function, what will be the predicted output class?

   Solution: 0. L^2 distances from $(0, 0)$ are $3\sqrt{2}(\approx 4.24), 4, 5, 6$. Since the model will not choose $\{(0, -6), 1\}$, the majority class among three points will be 0.

3. Continued from Q3. If we change the distance function from L^2 norm to L^1 norm, what will be the predicted output class?

   Solution: 1. L^1 distances from $(0, 0)$ are $6, 4, 7, 6$. Because $\{(3, 4), 0\}$ will be excluded, the majority class will be 1.

4. Select all that true:
   (a) Generative models in general learn fewer parameters than discriminative models.
   (b) Generative models model the joint probability distribution p(x, C), where x is data and C is class.
   (c) Generative models that models the joint distribution of data x and classes C can be converted to calculate p(C | x)
   (d) Regularization can only be used for discriminative models and not generative models

Solution: (b) and (c).

5. Suppose we are using GDA on a dataset with two classes. What happens when we use different covariance matrices for each class? Select all that are true:

   (a) Different covariance allows us to model a non-linear decision boundary

   (b) Calculating the MLE for different covariance takes less computation

   (c) Different covariance can increase the log likelihood

   (d) Different covariance is less likely to underfit the training data

   (e) Different covariance guarantees lower test error

Solution: (a), (c), and (d).

Learning difference covariance in GDA is analogous to adding more polynomial features to linear regression.

Suppose the training data is non-linear or skewed, and imagine we fit (i) GDA with fixed covariance and (ii) GDA with learned covariance. Then the likelihood for (ii) should always be higher because (ii) subsumes (i).

When the training data is skewed (e.g. ovular shaped) learning different covariance can learn this skew better than when the covariance is fixed, hence the log likelihood increases and it is less likely to underfit.