Support Vector Machines and PCA

1) A principal components analysis was run and the following eigenvalue results were obtained: 2.731, 2.218, 1.442, 0.009, 0.00183, 0.00085. How many components would you retain assuming you want to retain 99% of variance? 

3

2) K-means is not deterministic. True/False? True

3) What would happen when you use a very small C (C~0) in the SVM?
   a) Misclassification would happen
   b) Data will be correctly classified
   c) Can’t say
   d) None of these

Context: Suppose you have trained an SVM with linear decision boundary after training SVM, you correctly infer that your SVM model is under fitting. (for questions 4-6)

4) Which of the following option would you more likely do when implementing SVM next time?
   a) You want to increase your number of data points
   b) You want to decrease your number of data points
   c) You will try to create more features
   d) You will try to reduce the number of features
   e) You will try a more complex kernel (e.g. Gaussian)

5) Suppose you gave the correct answer in the previous question. What is happening?
   Are you lowering or increasing the bias? The variance?
   Increasing variance, decreasing bias

6) In the question above, you want to change one of the SVM’s hyperparameter so that the model will not underfit?
   a) We will increase the parameter C
   b) We will decrease the parameter C
   c) Changing in C don’t effect
   d) None of these

7) We usually use feature normalization before using the Gaussian kernel in SVMs. Why?
   Because similarity metric
Context: Suppose you are dealing with a 4 class classification problem and you want to train a SVM model on the data in a One-vs-all method.

8) How many times do we need to train our SVM model in such case? 4

9) Suppose your problem has changed now. Your data only has 2 classes. How many times do we need to train SVM in such case? 1

Context: Suppose you are using SVM with a linear kernel of polynomial degree 2, you have applied this on data and found that training and validation accuracy is 100%.

10) Now, you increase the complexity (or degree of polynomial of this kernel). What would you think will happen?
   a) Increasing the complexity will overfit the data
   b) Increasing the complexity will underfit the data
   c) Nothing will happen since your model was already 100% accurate
   d) None of these

11) What is/are true about kernel in SVM?
   1. Kernel function maps low dimensional data to high dimensional space
   2. It’s a similarity function

12) Kernels, C, SVMs (adapted from CMU)
   Assume you have the following dataset. You fit a SVM with a quadratic kernel.

   a) What is the shape of a decision boundary? **Parabola**
   b) Draw the decision boundary for large values of C. Remember that C penalizes misclassifications (1/C acts as a regularizer).
   c) Draw the decision boundary for small values of C. Remember that C penalizes misclassifications (1/C acts as a regularizer).
d) Assume that you know that training data is representative of the true data. What value of C should you choose? **Very high**

e) What if the training data is known to be not reliable? **Very low**

f) Assume infinite value of C. What is the decision boundary? Draw a point that will change this decision boundary. Similarly, draw a point that will not.

Figure 2: Solutions for Problem 2