CS 691 Final Exam        Name _____

## Short Problems (20 points)

1. Learning to drive is an example of imitation learning or reinforcement learning? Explain. (5 points)

2. Labels would never be helpful when using PCA. True or False? Choose one and explain. (5 points)

3. Why do we convert the SVM problem to its dual form? (5 points)

4. Consider the following situation: Landry is concerned about the security of their bank accounts. They call the bank and the manager asks them to detail all instances in the last six months wherein they might have shared their account details with another person for any kind of transaction, or may have accessed their online account from a public system, etc... Landry had ten such instances in reality, and they narrated twenty instances to finally spell out the ten correct instances. What would be Landry's recall and precision in this situation? (5 points)

691

# Models and Algorithms (20 points)

5. For each of the methods listed below. Detail the model and the algorithm ("iterative" is not enough, write out the algorithm in steps) associated with the method. Each algorithm should be no more than 3 steps. Keep it high level. Clearly label each.

   (a) 3-NN (5 points)

   (b) 5-Means (5 points)

   (c) Perceptron (5 points)

   (d) PCA (5 points)

691

# Unsupervised Pre-Processing (10 points)

6. I have the following data. I want to use an unsupervised learning method to structure the data prior to using a supervised method for classification. Choose an unsupervised learning method and a supervised learning method for your implementation. Detail the steps required to implement your combination (supervised + unsupervised) method. Justify your choice of both unsupervised and supervised methods. Assume the reader knows how the methods work, so specify hyperparameters, and what inputs and outputs would be and how you would use them.
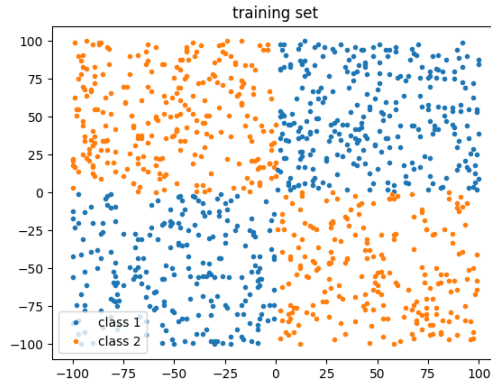
# SVM And Kernels (15 points)

7. Multiple kernels can be combined to produce new kernel functions. For example $K(x, z) = K_1(x, z) + K_2(x, z)$ is a valid kernel function. For the questions below, kernel $K_1$ has the associated feature transform $\phi_1$ and similarly $K_2$ has the feature transform $\phi_2$. Identify the feature transform associated with $K$ for the expressions given below. *Hint:* You may find the following operator helpful. The operator $[*, *]$ denotes concatenation of the two arguments. For example, $[x, z] = (x1, x2, z1, z2)$.

   (a) $K(x, z) = aK_1(x, z)$, for some scalar $a > 0$ (5 points)

   (b) $K(x, z) = aK_1(x, z) + bK_2(x, z)$, for some scalars $a, b > 0$ (5 points)

   (c) Suppose you are given the choice between using the normal perceptron algorithm, which directly works with $\phi(x)$, and the dual (kernelized) perceptron algorithm, which does not explictly compute $\phi(x)$ but instead works with the kernel function $K$. Keeping space and time complexities in mind, when would you prefer using the kernelized perceptron algorithm over the normal perceptron algorithm? Choose one and explain. *Note:* Here $N$ denotes the total number of training samples and $d$ is the dimensionality of $\phi(x)$. (5 points)

      i. $d >> N$ ($d$ is much greater than $N$)
      ii. $d << N$ ($d$ is much less than $N$)
      iii. Always
      iv. Never

# Decision Boundaries (15 points)

8. Consider a 2-class classification problem in a 2D feature space with labels (target values) that are either $+1$ or $-1$. The training data consists of 7 samples as shown below. (4 black diamonds for the positive class and 3 white diamonds for the negative class). The points are marked 1-7 for your convenience.
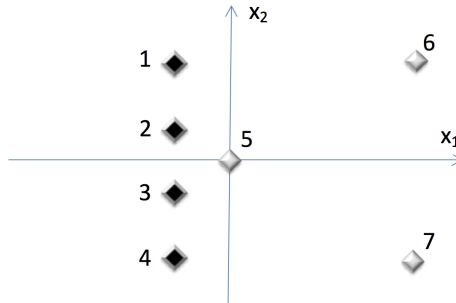


(a) K-Means: Draw on the plot the centers for the two classes. Mark them with a $\oplus$ for the positive class and $\ominus$ for the negative class. Draw the decision boundary for k-means using a dashed line. (3 points)

  i. What is the training error? (1 point)

  ii. Is there any sample such that upon its removal, the decision boundary changes in a manner that the removed sample goes to the other side? Answer ("yes" or "no"). (1 point)

(b) SVM: Draw the decision boundary for the hard-margin SVM using a thick solid line. Draw the margins on either side with thinner solid lines. Circle the support vectors. (3 points)

  i. What is the training error? (1 point)

  ii. The removal of which sample would change the decision boundary? Write "None" if none. (1 point)

(c) Which method, K-Means of SVM, is more generalizable in this setting? Explain. (2 points)

(d) Is there a setting in which K-Means and SVM could give the same decision boundary on this data? Yes or no. Explain. (3 points)

# Missing Data (10 points)

9. I have the following snippet (small portion) of a dataset containing medical data. We want to use this dataset to predict cervical cancer. There is at least one sample with a feature value missing (NaN) (there may be more, remember this is a small portion of the dataset). We learned about 3 unsupervised learning methods in this class. Choose one method and detail the steps you would use to "fill in" the missing feature values. Specify inputs and outputs to the unsupervised learning algorithm and how you would use them to replace the NaN values.

| | Age | Number of sexual partners | First sexual intercourse | Num of pregnancies | Smokes | Smokes (years) | Smokes (packs/year) | Hormonal Contraceptives | Hormonal Contraceptives (years) | IUD |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 18 | 4.0 | 15.0 | 1.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 1 | 15 | 1.0 | 14.0 | 1.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 2 | 34 | 1.0 | NaN | 1.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 3 | 52 | 5.0 | 16.0 | 4.0 | 1.0 | 37.0 | 37.0 | 1.0 | 3.0 | 0.0 |
| 4 | 46 | 3.0 | 21.0 | 4.0 | 0.0 | 0.0 | 0.0 | 1.0 | 15.0 | 0.0 |

691

# ML in the Real World (10 points)

10. You have developed a wonderful new theory of classification learning, called "the UNR algorithm". You input data, it outputs a formula. You decide to test your theory on the problem of predicting in January who will win a political election in November given attributes like office, region, party, incumbency, success in raising funds to date, current poll numbers, etc. You have a database with all this data for all American elections over the last twenty years. You ask a research assistant to test the UNR algorithm on this data. They come back with the good news that when they ran the UNR algorithm over the your data set, it output a formula that gives 80% accuracy on the data. What further tests should you run before release this method, claiming that you have promising new technique for machine learning? List at least three tests you would run, and how they would provide you with more information about the UNR algorithm.

691