# CS 491 Final Exam   Name _____

Short Problems (20 points)

1. 491 True/False: When training a neural network, you should check to see when the error is at a minimum on the test data, and stop training at that point. Explain. **Solution:** False. We use validation data.

2. 691 Let's say we have two polynomial feature maps: $\phi_1(x) = \{x, x^2\}$, and $\phi_2(x) = \{2x, 2x^2\}$. In general, is the margin we would attain using $\phi_2(x)$ greater, equal, or smaller, in comparison to the margin resulting from $\phi_1(x)$? **Solution:** Greater

3. Give one similarity and one difference between feature selection and PCA. **Solution:** Similarity=reduce the dimension of data, difference=feature selection finds a subset of features, while PCA produces a smaller, new set of features.

4. 691 True/False: We would expect the support vectors to remain the same in general as we move from a linear kernel to higher order polynomial kernels. Explain. **Solution:** False

5. 491 True/False: The maximum margin decision boundaries that support vector machines construct have the lowest generalization error among all linear classifiers. Explain. **Solution:** False

6. Which of the following statements are true? No need to explain.

   (a) Training a k-nearest neighbors classifier takes more computational time than applying it. **Solution:** False

   (b) The more training examples, the more accurate the prediction of a k-nearest neighbors classifier. **Solution:** True

   (c) k-nearest neighbors cannot be used for regression (to predict a real-value). **Solution:** False

   (d) A k-nearest neighbors is sensitive to outliers. **Solution:** True

7. Explain how you might use the result of PCA on the below data to perform classification.
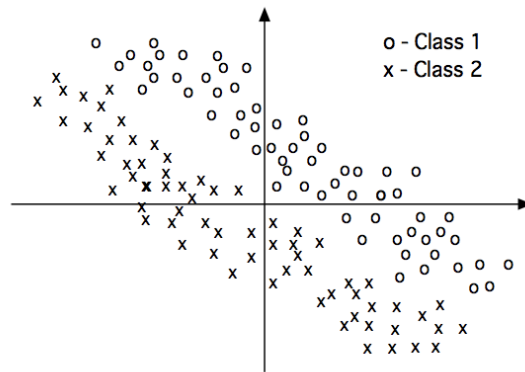
8. Consider two types of Neural Network activations:

- linear: $h = w \cdot x + b$
- hard threshold: $h = 1$ if $w \cdot x + b \geq 0$, and $h = 0$ otherwise.

Which of the following functions can be exactly represented by a neural network with one hidden layer which uses linear and/or hard threshold activation functions? For each case, justify your answer.

(a) Polynomials of degree one
(b) Hinge loss: $h(x) = max(1 - x, 0)$
(c) Polynomials of degree two
(d) Piece-wise constant functions

9. For each of the following situations, indicate whether the classifier will produce a linear decision boundary always, sometimes or never. Give a short explanation of your answer. In every case, assume that the input has two features x_1 and x_2, which have continuous values ranging from -1 to 1. For example, the value of x_1 is any real number between -1 and 1. And suppose the label, y has two possible values, 1 or -1. In all cases, assume there is plenty of training data (at least 100 examples for each class).

   (a) We build a decision tree with two levels.

   (b) We build a decision tree with one level.

   (c) We train a Perceptron.

   (d) We use K-nearest neighbor, with K $= 1$.

   (e) We train an SVM using a quadratic kernel.

   (f) We use a neural network. The network has two input units, two hidden units and one output unit. We use ReLU in the hidden units (remember ReLU(x) $=$ max(0,x)) and a sigmoid after the output unit. The input is classified as belonging to class 1 if the output is greater than $\frac{1}{2}$, and as class -1 if the output is less than $\frac{1}{2}$.

10. Run K-Means manually for the following dataset. Circles are data points and the filled-in circles in the first panel are the initial cluster centers. Draw the cluster centers and the decision boundaries that define each cluster. Use a different panel for each iteration. Use as many panels as you need until convergence.
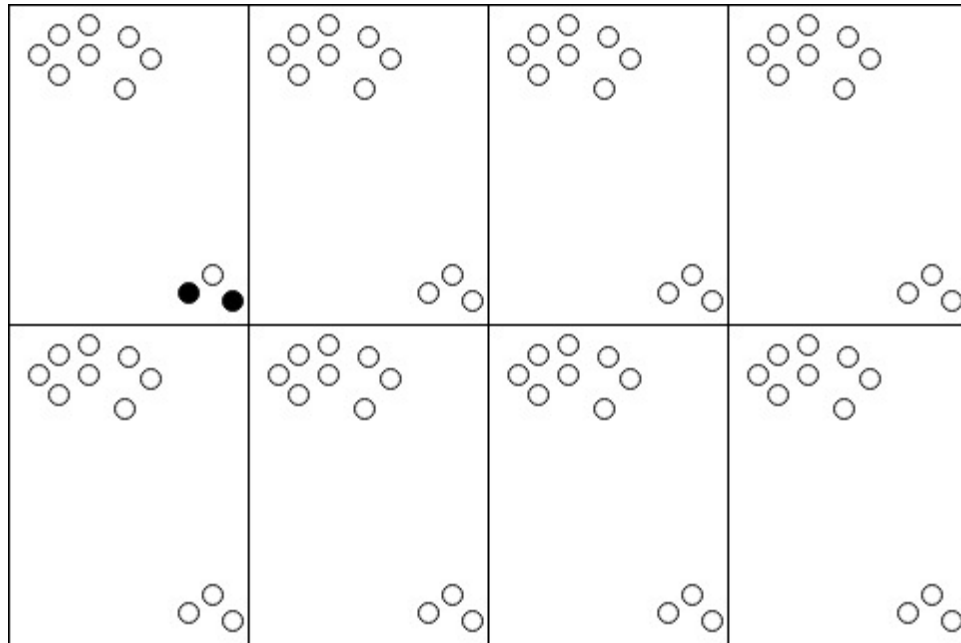
11. The following figure shows a dataset with two real-valued features (x_1 and x_2) and a binary output class. The positive points are shown as solid dots and the negative points are small circles.

6 –

5 – ● ●

4 – ● ●

3 – ●

2 – ○ ○

x2 1 – ○ ○ ○ ○

0 –

　0　1　2　3　4　5　6

X1 →
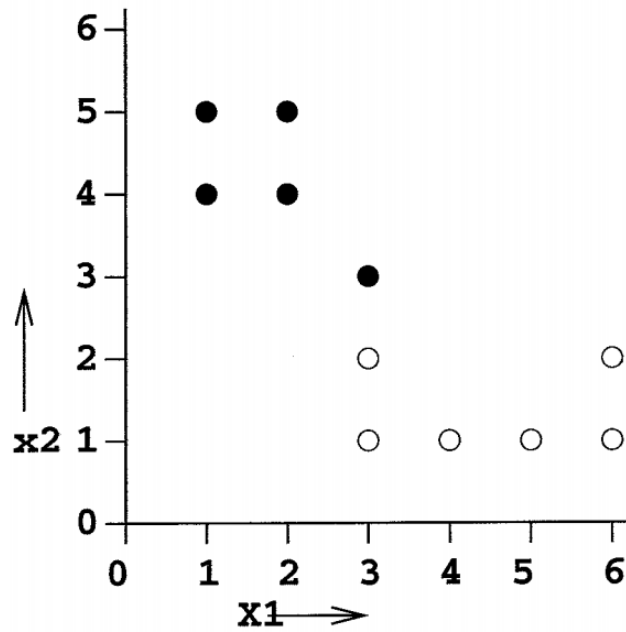
(a) Suppose you are using a linear hard-margin SVM. Draw three lines on the above diagram, showing the decision boundary and the two sides of the margin. Circle the support vector(s).

(b) Calculate the values of w and b learned for part (a).

12. Recall that Adaboost learns a classifier H using a weighted sum of weak learners h_t as follows:

$$H(x) = sign(\sum_{t=1}^{T} \alpha_t h_t(x))$$

In this question, we will use decision trees as our weak learners, which classify a point as $\{1, -1\}$ based on a sequence of threshold splits on its features (x_1 and x_2). Red squares are negative points and blue circles are positive points. **In the questions below, be sure to mark which regions are marked positive/negative, and assume that ties are broken arbitrarily.**

(a) Now assume that our weak learners are decision trees of maximum depth 2, which minimize the weighted training error. Using the dataset below, draw the decision boundary learned by h_1.
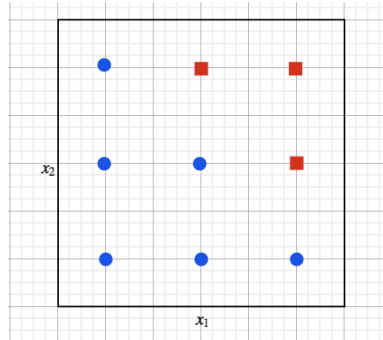
(b) On the dataset below, circle the point(s) with the highest weights on the second iteration, and draw the decision boundary learned by h_2.
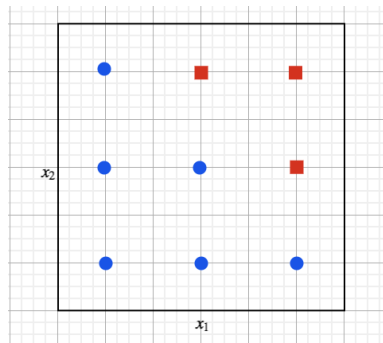
(c) On the dataset below, draw the decision boundary $H = sign(\alpha_1 h_1 + \alpha_2 h_2)$. (Hint: you do not need to explicitly compute the α's).
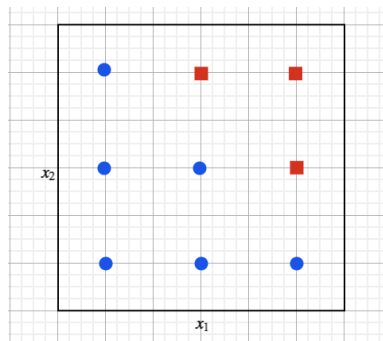
13. **BONUS QUESTION:** Write and answer your own exam question based on the material in the course. This question can be anything we covered in class or was in the course notes. This is your chance to show what you learned that may not have been otherwise tested in the exam. Up to four bonus points will be given based on difficulty, correctness and originality of the question. No points will be given for reproducing a question from this exam.