DSE 220 Final Exam

Name: ______________________________

Problem 1. (2 points) Consider the point $x = (-2, -2, \ldots, -2)$ in $\mathbb{R}^d$. What is $\|x\|_2$? Your answer may be a function of d.

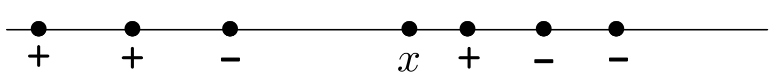
Problem 2. (4 points) A particular data set has 4 possible labels, with the following frequencies:

Label	Frequency
A	10%
B	50%
C	30%
D	10%

(a) What is the error rate of a classifier that always predicts C?

(b) What is the error rate of a classifier that picks a label (A,B,C,D) at random, each with probability 1/4?

Problem 3. (3 points) A training set of six labeled one-dimensional points is shown on the line below, along with a new test point x.



(a) How will x be classified by 1-NN? Just give the label, $+$ or $-$.

(b) How will x be classified by 3-NN? Just give the label, $+$ or $-$.

(c) Suppose leave-one-out cross-validation (LOOCV) is used to assess the error rate of the 3-NN classifier based on the training set of six points. What is the LOOCV error rate?

Problem 4. (2 points) We use 4-fold cross-validation to figure out the right value of k to use when running k-nearest neighbor on a data set of size 10,000. When checking a particular value of k, we look at four different training sets. What is the size of each of these training sets?

Problem 5. (3 points) For each of the following prediction problems, say whether it is best thought of a *classification* problem or a *regression* problem.

(a) We want to predict a person's salary based on their education level, geographical location, and field of work.

(b) We want to predict whether a person will default on a loan based on their income and credit history.

(c) We want to predict how many times a week a person visits the grocery store, based on their age and sex.

Problem 6. (3 points) Consider the vector $x = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$.

(a) What is $\|x\|_1$?

(b) What is $x^T x$?

(c) What is $x x^T$?

Problem 7. (6 points) Which of these distance functions is a *metric*? If it is not a metric, state which of the four metric properties it violates.

(a) Let $\mathcal{X} = \mathbb{R}^k$ and $d(x, y) = \|x - y\|_\infty$.

(b) Let $\mathcal{X} = \mathbb{R}^k$ and let $d(x, y) = |\{i : x_i \neq y_i\}|$ be the *number of coordinates* on which x, y differ.

(c) Let $\mathcal{X} = \{\text{dog}, \text{cat}, \text{rabbit}\}$ and let d be given by the table below:

	dog	cat	rabbit
dog	0	1	4
cat	1	0	2
rabbit	4	2	0

Problem 8. (4 points) Consider the following data set of four points (x, y):

$$(1, 2), (2, 2), (2, 4), (3, 4).$$

(a) Find the line $y = ax + b$ that minimizes the mean squared error (MSE) on these points.

(b) Sketch (or plot) the four points and the line.

Problem 9. (4 points) Consider a logistic regression model for data in $\mathbb{R}^2$ that is based on the linear function $w \cdot x + b$, where $w = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $b = -4$.

(a) Draw the decision boundary in $\mathbb{R}^2$. Make sure to clearly indicate where it intersects the x_1- and x_2-axes.

(b) Consider the point $x = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$. Would the logistic regression model classify this point as $+1$ or -1?

(c) For the point x from part (b), what would the logistic regression model return as the conditional probability that $y = 1$? You may leave an exponent in your answer.

Problem 10. (4 points) We have a training set of points $x^{(1)}, \ldots, x^{(n)} \in \mathbb{R}^d$, with response values $y^{(1)}, \ldots, y^{(n)}$. We will use regression to fit these points with a linear function $y = w \cdot x + b$.

(a) In class, we studied three different loss functions for linear regression. Which of them would be most suitable if we want a **sparse** vector w? (Just name the method.)

(b) Which of the three methods would be most **unsuitable** if we have a shortage of training data?

Problem 11. (5 points) Consider the following loss function on vectors $w \in \mathbb{R}^3$:

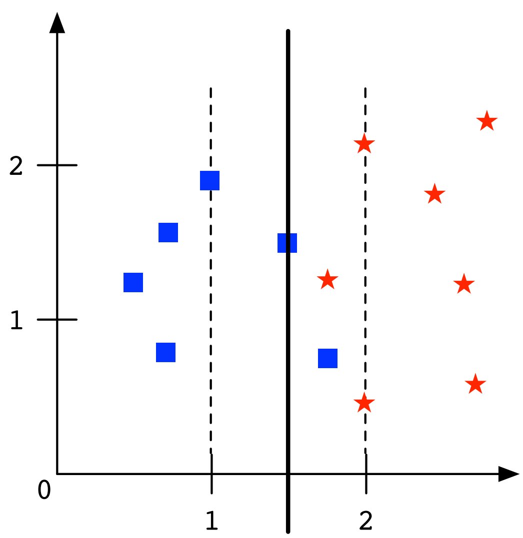
$$L(w) = w_1^2 + 5w_3^2 + 2w_1w_3 - w_2w_3.$$

(a) What is $\nabla L(w)$?

(b) Suppose we use gradient descent to minimize this function, and that the current estimate is $w = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$.

If the step size is $\eta = 0.1$, what is the next estimate?

Problem 12. (5 points) The picture below shows the decision boundary—and margins—obtained from running soft-margin SVM on a small data set of squares (the positive points) and stars (the negative points) in $\mathbb{R}^2$.



(a) Circle all the support vectors. For each, give the approximate value of the corresponding slack variable.

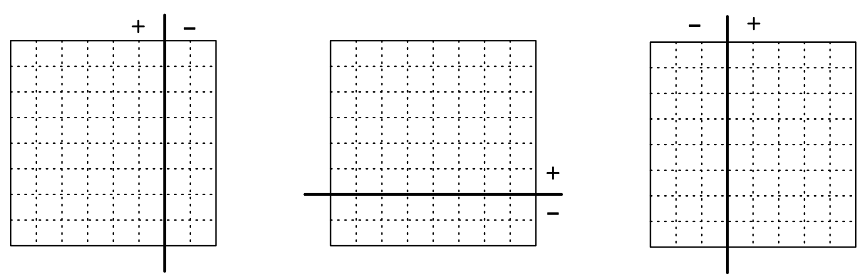
(b) What is $\|w\|$, approximately? Just state the value.

Problem 13. (4 points) Multiclass logistic regression is run on a 2-d data set with three labels $(1, 2, 3)$. It returns the following three linear functions:

$$w_1 = \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \quad w_2 = \begin{pmatrix} 2 \\ -1 \end{pmatrix}, \quad w_3 = \begin{pmatrix} -1 \\ 2 \end{pmatrix},$$

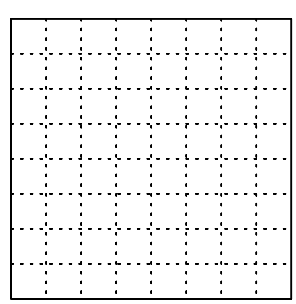
with $b_1 = b_2 = b_3 = 0$. Show the decision boundaries for the three classes. Mark each region carefully.

Problem 14. (4 points) Suppose that we run boosting with decision stumps on a certain data set in $\mathbb{R}^2$, with labels in $\{-1, 1\}$. After three iterations, we get the following weak rules h_t and weights α_t.



h_1 $\alpha_1 = 0.25$ | h_2 $\alpha_2 = 0.50$ | h_3 $\alpha_3 = 0.50$

In the figure below, draw the exact decision boundary for the final rule that combines the three weak rules. Clearly indicate the $+$ and $-$ regions.



Problem 15. (7 points) Consider the following two unit vectors in $\mathbb{R}^3$:

$$u_1 = \begin{pmatrix} 3/5 \\ 4/5 \\ 0 \end{pmatrix}, \quad u_2 = \begin{pmatrix} -4/5 \\ 3/5 \\ 0 \end{pmatrix}.$$

(a) Notice u_1 and u_2 are orthogonal. Find a third unit vector u_3 that is orthogonal to both u_1 and u_2.

(b) A particular data set in $\mathbb{R}^3$ has covariance matrix M with eigenvectors u_1, u_2, u_3 (from part (a)) and eigenvalues $4, 2, 1$, respectively. What is M? (You can leave it as a product of matrices.)

(c) Suppose principal component analysis is used to project the data set (from part (b)) into $\mathbb{R}^2$. What is the 2-dimensional projection of $x = (10, 5, 5)$?

(d) What is the three-dimensional reconstruction of x from the projection you obtained in part (c)?