Instructions: read these first!

Do not open the exam, turn it over, or look inside until you are told to begin.

Switch off cell phones and other potentially noisy devices.

Write your full name on the line at the top of this page. Do not separate pages.

You may refer to a calculator and your cheat-sheet. You may not refer to any other printed material, and any other computational device (such as laptops, phones, iPads, friends, enemies, pets, lovers).

Read questions carefully. Show all work you can in the space provided.

Where limits are given, write no more than the amount specified. The rest will be ignored.

Avoid seeing anyone else’s work or allowing yours to be seen.

Do not communicate with anyone but an exam proctor.

If you have a question, raise your hand.

When time is up, stop writing.
1. [12 points] In this question, we will again look at how \( k \)-nearest neighbor classifiers can be robust to noise. Suppose we have two labels 0 and 1. We are given a test point \( x \), and its \( k \) nearest neighbors \( z_1, \ldots, z_k \), where \( z_i \) is the \( i \)-th closest neighbor of \( x \) (so \( z_1 \) is the closest neighbor, \( z_2 \) is the second closest neighbor and so on).

Suppose that the probability that the label of \( z_i \) is not equal to the label of \( x \) is \( p_i \). Also assume that all distances are unique, so the \( i \)-th closest neighbor of \( x \) is unique for all \( i \).

Answer the following questions:

(a) [3 points] Now, suppose, for this question that \( p_1 = 0.1 \), and \( p_i = 0.2 \) for \( i > 1 \). What is the probability that the 1-nearest neighbor classifier makes a mistake on \( x \)?

(b) [6 points] In the setting of part (a), what is the probability that the 3-nearest neighbor classifier makes a mistake on \( x \)?

(c) [3 points] Based on your calculation, what can you conclude about the relative robustness of 1 and 3-nearest neighbor classifiers in this case?
2. [12 points] Recall that the projection of a vector $x$ onto another vector $y$ is defined as the vector $\frac{(x \cdot y)}{\|y\|^2} y$. Suppose $d > 1$ and suppose that $u_1 \neq u_2$ are two $d \times 1$ vectors such that $\|u_1\| = \|u_2\| = 1$, and $u_1$ and $u_2$ are not orthogonal to each other. Let $x$ be a third non-zero vector.

Alice and Bob have been asked to take projections of $x$ onto $u_1$ and $u_2$. Alice first projects $x$ onto $u_1$ and then projects the result onto $u_2$ to get the vector $a$. Bob first projects $x$ onto $u_2$ and then projects the result onto $u_1$ to get the vector $b$.

(a) [3 points] Is $a = b$ always?

(b) [5 points] Is $\|a\| = \|b\|$ for all $x$? Justify your answer if this is the case; if it is not the case, provide a counter-example.

(c) [4 points] Does your answer change when $u_1$ and $u_2$ are orthogonal?
3. [10 points] Answer the following questions. In these questions, \( x \) is a \( d \times 1 \) vector with \( \|x\| \leq 1 \).

(a) [5 points] Suppose \( M \) is a \( d \times d \) matrix such that for each \( i \) and \( j \), \(|M_{ij}| \leq 1\). What are the maximum and minimum values of \( \|Mx\| \)? Justify your answer.

(b) [5 points] Remember that a diagonal matrix \( D \) is one where \( D_{ij} = 0 \) when \( i \neq j \). Suppose \( D \) is a \( d \times d \) diagonal matrix such that for each \( i \), \(|D_{ii}| \leq 1\). What are the maximum and minimum values of \( \|Dx\| \)? Justify your answer.
4. [16 points] For each of the following statements, say whether they are correct or incorrect. Justify your answer.

(a) [4 points] For every training data set \((x_1, y_1),\ldots,(x_n, y_n)\), the training error of the 3-nearest neighbor classifier is always zero. You may assume that each \(x_i\) is unique.

(b) [4 points] As we increase \(k\), the training error of the \(k\)-nearest neighbor classifier always increases.

(c) [4 points] If \(u\) and \(v\) are any two orthogonal unit vectors, then \(\|u + v\| = 1\).

(d) [4 points] Let \(X\) be a matrix of rank \(r\). If we add the same vector \(y\) to each row of \(X\), then the resulting matrix always has rank \(r + 1\).