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.

You can see your graded final in the student affairs office starting Mon Mar 27.

<table>
<thead>
<tr>
<th>Question</th>
<th>Points</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>Total:</td>
<td>70</td>
<td></td>
</tr>
</tbody>
</table>
1. [10 points] Suppose you are given the following feature vectors:

\[ x_1 = (0, 4), x_2 = (0, 2), x_3 = (2, 0), x_4 = (4, 0), x_5 = (5, 0) \]

Their corresponding labels are

\[ y_1 = 1, y_2 = -1, y_3 = -1, y_4 = 1, y_5 = 1 \]

Consider four weak learners \( h_1, h_2, h_3, h_4 \), which are: \( h_i(x) = \text{sign}(\langle w_i, x \rangle + b_i) \), for \( i = 1, 2, 3, 4 \). The parameters are \( w_1 = w_2 = (1, 0), b_1 = -1, b_2 = -3, w_3 = w_4 = (0, 1), b_3 = -1, b_4 = -3 \).

Using these weak learners, we will now create an ensemble classifier \( H \) using boosting. In each round, we will pick the best weak learner – that is, the learner that has the lowest classification error. Write down the weak learners \( h_t \) picked in the first three rounds of boosting as well as the corresponding \( \alpha_t \) values.
2. [10 points] Suppose you are given the following feature vectors:

\[ x_1 = (1, 0), x_2 = (4, 2), x_3 = (0, -1), x_4 = (-1, -1), x_5 = (-2, 1) \]

Their corresponding labels are

\[ y_1 = 1, y_2 = 1, y_3 = -1, y_4 = -1, y_5 = -1 \]

Suppose we run perceptron on this dataset starting with \( w_1 = (0, 0) \). Write down the values of \( w_2, w_3, w_4, w_5 \) and \( w_6 \).
3. [25 points] State whether each of the following statements is true or false. If it is true, provide a brief justification or proof; if it is false, provide a counterexample or a justification.

(a) [5 points] Suppose we are given a data set \((x_1, y_1), \ldots, (x_n, y_n)\) that is linearly separable. If the feature vectors have zero mean (that is, if \(\frac{1}{n} \sum_{i=1}^{n} x_i = 0\)), then the data set \((x_1, y_1), \ldots, (x_n, y_n)\) is also linearly separable through the origin.

(b) [5 points] Let \(X\) and \(Z\) be two random variables such that \(Z = X - 2\). Then \(H(Z) = H(X)\).

(c) [5 points] If the training data is linearly separable, then running perceptron on one pass of the data is guaranteed to learn a classifier with zero training error.
(d) [5 points] The decision boundary of averaged perceptron is always a hyperplane, no matter what the training data is.

(e) [5 points] Two decision trees that have the same training error on a training dataset \( S \) also have the same test error on the same test dataset \( T \), no matter what \( S \) and \( T \) are.
4. [25 points] State whether each of the following statements is true or false. If it is true, provide a brief justification or proof; if it is false, provide a counterexample or a justification.

(a) [5 points] $x, z$ are real $d$-dimensional vectors. Then the function $K(x, z) = \frac{(x \cdot z)}{\|x\|^2 \|z\|^2}$ is a kernel.

(b) [5 points] $x, z$ are real $d$-dimensional vectors. Then the function $K(x, z) = \|x - z\|^2$ is a kernel.

(c) [5 points] If $K(x, z)$ and $L(x, z)$ are kernels, then $M(x, z) = 2K(x, z) - 3L(x, z)$ is a kernel, no matter what $K$ and $L$ are.
(d) [5 points] If $K(x,z)$ and $L(x,z)$ are kernels, then $M(x,z) = K(x,z)L(x,z)^2$ is also a kernel, no matter what $K$ and $L$ are.

(e) [5 points] The matrix $B$ is a valid kernel matrix:

$$B = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$$  \hspace{1cm} (1)