Midterm Exam

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. [18 points] Suppose we are given two random variables $X$ and $Z$ with the following joint distribution:

<table>
<thead>
<tr>
<th></th>
<th>X=0</th>
<th>X=1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z=0</td>
<td>$\frac{1}{5}$</td>
<td>$\frac{1}{5}$</td>
</tr>
<tr>
<td>Z=1</td>
<td>$\frac{3}{5}$</td>
<td>$\frac{1}{3}$</td>
</tr>
</tbody>
</table>

Answer the following questions.

(a) [4 points] Write down the marginal distribution of $Z$.

(b) [6 points] Write down the conditional distributions $X|Z=0$ and $X|Z=1$.

(c) [4 points] Calculate the conditional entropy $H(X|Z)$.

(d) [4 points] Calculate the information gain $I(X,Z) = H(X) - H(X|Z)$. 
2. [12 points] Let $x$ be a $d \times 1$ vector such that $\|x\| \leq 1$. Answer the following questions.

(a) [4 points] Let $M$ be a $d \times d$ matrix such that $M_{i1} = 0$ for all $i$ (that is, the first column of $M$ is all zeros) and for all $i$ and $j$, $|M_{ij}| \leq 1$. Write down the maximum possible value of $\|Mx\|$. Write down a $d \times d$ matrix $M$ and a $d \times 1$ $x$ which achieve this maximum value. (Hint: To write down a $d \times d$ matrix, you can specify what its entries $M_{ij}$ will be.)

(b) [2 points] In part (a), what is the minimum possible value of $\|Mx\|$? Write down a $d \times d$ matrix $M$ and a $d \times 1$ vector $x$ that achieve this minimum value.
(c) [4 points] Let $D$ be a $d \times d$ diagonal matrix such that $D_{i1} = 0$ for all $i$ (that is, its first column is all zeros), and for all $i$, $|D_{ii}| \leq 1$. Write down the maximum possible value of $\|Dx\|$. Write down a $d \times d$ matrix $D$ and a $d \times 1$ vector $x$ which achieve this maximum value.

(d) [2 points] In part (c), what is the minimum possible value of $\|Dx\|$? Write down a $d \times d$ matrix $D$ and a $d \times 1$ vector $x$ that achieve this minimum value.
3. [20 points] For each of the following questions, several possible answers are given. Only one of these answers is completely correct; mark the one which is completely correct. No justification is needed. Note: Some of the answers may be partially correct. You will not get any credit for marking an answer which is only partially correct. You will also not get any credit if you mark multiple answers (even if one of the answers marked is fully correct.)

(a) [4 points] Let $x$ be any arbitrary vector, and let $u_1$ be an unit vector. Which of the following statements is true, for all $x$?

1. $\|x\| = |\langle u_1, x \rangle|$.
2. $\|x\| \leq |\langle u_1, x \rangle|$.
3. $\|x\| \geq |\langle u_1, x \rangle|$, and $\|x\| = |\langle u_1, x \rangle|$ only if $x = u_1$.
4. None of the above.

(b) [4 points] Two $k$-NN classifiers are said to be equal if they have exactly the same decision boundary. Alice, Bob and Carol construct two 3-NN classifiers using the same training dataset $S$. Alice uses the Euclidean distance, Bob uses the distance:

$$d_{Bob}(x, z) = 30e^{-0.01\|x-z\|} - 30$$

and Carol uses the distance:

$$d_{Carol}(x, z) = e^{\|x-z\|+1} - e$$

Which of the following statements is true? For all training sets $S$,

1. Alice, Bob and Carol get the same classifiers.
2. Alice and Bob get the same classifiers, Carol might get a different one.
3. Carol and Alice get the same classifiers, Bob might get a different one.
4. Bob and Carol get the same classifiers, Alice might get a different one.
5. Alice, Bob and Carol may all get different classifiers.

(c) [4 points] Let $X$ be a random variable, and let $Y = X^2 + 2$. Which of the following statements are true for all $X$?

1. $H(Y) = H(X)$.
2. $H(Y|X) = 0$.
3. $H(Y) = 2H(X) + \log 2$.
4. All of the above.
(d) [4 points] Let \( u \) and \( v \) be two arbitrary non-zero vectors such that \( \langle u, v \rangle = 0 \), and let \( A \) be the matrix:

\[
A = 3uu^\top + 5vv^\top
\]

Which of the following statements is true for any such \( u \) and \( v \)?

1. \( A \) has rank 2.
2. \( \|Au\| = 3 \).
3. \( v^\top Av = 5\|v\|^2 \).
4. All of the above.

(e) [4 points] Which of the following statements about training and test data is true?

1. Training and test data should always be kept separate.
2. The ID3 algorithm (without the pruning step) will always overfit.
3. Since training and test data are drawn from the same distribution, if you have low training error, then you will also have low test error.
4. All of the above.