Final Exam Feedback

● Graded

Student

Yisak Tolla

Total Points

24.5 / 25 pts

Question 1

Decision Trees **5** / 5 pts

1.1 **Entropy of dataset** **1** / 1 pt

✔ **– 0 pts** No penalty

– 0.5 pts Math error / formula error / plug-in error

– 1 pt Incorrect

1.2 **Entropy of A1=1 branch** **1** / 1 pt

✔ **– 0 pts** No penalty

– 0.5 pts Math error / formula error / plug-in error

– 1 pt Incorrect

1.3 **Entropy of A1=0 branch** **1** / 1 pt

✔ **– 0 pts** No penalty

– 0.5 pts Math error / wrong formula / plug-in error

– 1 pt Incorrect

1.4 **Leaf node attributes** **1** / 1 pt

✔ **– 0 pts** No penalty

– 1 pt Neither attribute 2 or 3 would perfectly separate the training data for either branch.

1.5 **Does the tree fit the training data?** **1** / 1 pt

✔ **– 0 pts** No penalty

– 0.5 pts This tree misclassifies the training point $X^{(5)}$ (should be 1, tree says 0)

– 1 pt Blank

Question 2

Logistic Regression and k-NN — **5** / 6 pts

2.1 **Rough plot** — **1** / 1 pt

- ✔ **– 0 pts** No penalty
- **– 0 pts** No penalty, but this is **not** a useful graph of the data
- **– 1 pt** Blank

2.2 **k-NN output for X(6)** — **2** / 2 pts

- ✔ **– 0 pts** No penalty
- **– 1 pt** Reversed!
- **– 1 pt** Wrong for one of $k = 1$ or $k = 3$ but not the other (see later hint)
- **– 2 pts** This is not how k-NN works at all.

2.3 **Logistic regression output for X(6)** — **0** / 1 pt

- **– 0 pts** No penalty
- ✔ **– 1 pt** Incorrect
- **– 1 pt** Blank / missing / illegible

2.4 **Changing the labels for Logistic Regression on X(4)** — **2** / 2 pts

- ✔ **– 0 pts** No penalty
- **– 1 pt** There is no way to set the parameters such that the prediction for $X^{(4)}$ changes, but the prediction for $X^{(3)}$ does not.
- **– 2 pts** Blank / missing / illegible

Question 3

Neural Networks **2.5** / 4 pts

3.1 **Perceptron rule updates** **1.5** / 3 pts

– **0 pts** No penalty

✔ – **1.5 pts** Mistake in iteration 1 (no further penalties)

– **1 pt** Mistake in NN output for iteration 2 (no further penalties)

– **0.5 pts** Mistake in iteration 2 (no further penalties)

– **3 pts** Blank / Missing / Illegible

3.2 **SGD vs Perceptron Rule requirements** **1** / 1 pt

✔ – **0 pts** No penalty

– **1 pt** Stochastic Gradient Descent will not work with the step-function **activation function**.

– **1 pt** Blank / missing / illegible

– **0.5 pts** The weight updates *follow* from the architecture

Question 4

Multiple Choice / True False — **9** / 10 pts

4.1 **Overfitting** — **1** / 1 pt

✔ **– 0 pts** No penalty

– 1 pt Not fitting the training data is sometimes called **under**fitting

– 1 pt Being insensitive to noise is a good thing!

4.2 **Stochastic Gradient Descent for fitting** — **1** / 1 pt

✔ **– 0 pts** No penalty

– 1 pt All of these algorithms *can* use SGD

4.3 **"Deep" Neural Network** — **1** / 1 pt

✔ **– 0 pts** Three!

4.4 **Random Forest tree construction** — **1** / 1 pt

✔ **– 0 pts** Correct

– 1 pt **Random** Forests use **random** sampling

4.5 **Minimizing SSE = ?** — **0** / 1 pt

– 0 pts Correct

– 1 pt Actually, **none** of the proofs we showed actually relate to generalization performance or model complexity.

✔ **– 1 pt** There is a subtle connection to the L*2* norm, but we didn't discuss L1 except briefly in Linear Regression.

4.6 **T/F: Inadmissible heuristics are always sub-optimal** — **1** / 1 pt

✔ **– 0 pts** Correct!

– 1 pt Incorrect, see the midterm feedback.

4.7 **When does an agent search?** — **1** / 1 pt

✔ **– 0 pts** Correct!

– 1 pt Incorrect, see the midterm feedback.

4.8 **Variants on search** — **1** / 1 pt

✔ **– 0 pts** Correct!

– 1 pt Incorrect, see the midterm feedback.

4.9 **T/F: Generic Search w/ inconsistent heuristics** — **1** / 1 pt

✔ **– 0 pts** Correct!

– 1 pt Incorrect, see the midterm feedback.

4.10 **Comparing Heuristics** **1** / 1 pt

✔ **– 0 pts** Correct!

– 1 pt Incorrect, see the midterm feedback.

Question 5

Extra credit! **3** / 0 pts

✔ **+ 3 pts** Extra credit!

001033

NAME: Yisak Tella

G#: G01366265

CS 480
Final Exam
Section 001 (Hrolenok)

This test is closed-book. You are allowed two pages of notes which you must turn in with the exam. You may use a calculator (but no phones/laptops). Write your name and G# in the boxes above, and your name at the top of each additional page in case the pages become separated.

Question 1: Decision Trees (5 parts)

Given the following data

Data point	Attribute 1	Attribute 2	Attribute 3	Label (Y)
$x^{(1)}$	1	0	1	1
$x^{(2)}$	0	1	0	0
$x^{(3)}$	1	1	1	0
$x^{(4)}$	0	0	1	0
$x^{(5)}$	1	1	0	1

Part 1 (1pts): What is the entropy of the entire dataset (all five data points)? (Hint: make sure you use base 2 when calculating logarithms!)

Entropy

$H(S) = -\sum_{y_i} p(y_i) \log(p(y_i))$

$= -\frac{2}{5}\log_2(\frac{2}{5}) - \frac{3}{5}\log_2(\frac{3}{5})$

Part 2 (1pt): What is the entropy of the labels for the data where attribute 1 is 1?

$= -\frac{2}{3}\log_2(\frac{2}{3}) - \frac{1}{3}\log(\frac{1}{3})$

Part 3 (1pt): What is the entropy of the labels for the data where attribute 1 is 0?

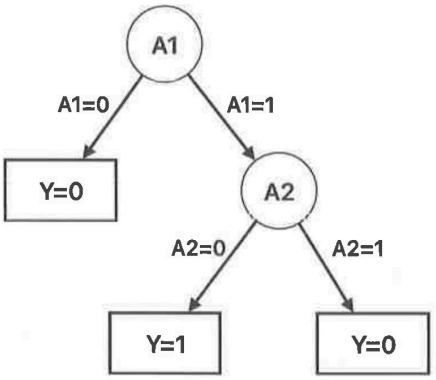
$= -\frac{2}{2}\log_2(\frac{2}{2}) - \frac{0}{2}\log_2(\frac{0}{2})$

Part 4 (1pt): If Attribute 1 is chosen first when building a tree from this data, the '0' branch immediately leads to a leaf node. Does any other attribute lead immediately to a leaf node if chosen first?

No other attribute leads to a leaf node if chosen first.

Write your name at the top of each separate page, in case pages become separated.
Name: Yisak Tolla.

Part 5 (1pt): Suppose we constructed the following decision tree:
Does this tree correctly classify all of the data points in the table? (Yes/no)



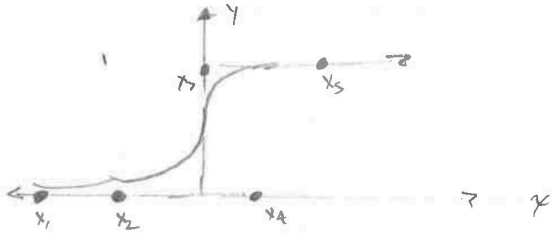
No, X^5 violates this tree

Question 2: Logistic Regression & k-NN (4 parts)

Consider the following data

Data point	X	Label
$x^{(1)}$	-2.5	0
$x^{(2)}$	-1.5	0
$x^{(3)}$	0.0	1
$x^{(4)}$	0.5	0
$x^{(5)}$	1	1

Part 1 (1pt): Draw a rough plot of the data. Note: This data is 1-D, so you should be marking points on a number line (do not plot the label as a "y-value").



Part 2 (2pt): If you were using k-NN to classify this data, how would you label the point $X^{(6)} = 0.55$ when k=1? What about when k=3?

When $k=1$, the label will be 0 as x^4 is the closest neighbor.

When $k=3$, the label will be 1 as 2 of its closest neighbors have a label of 1.

Write your name at the top of each separate page, in case pages become separated.
Name: Yisak Tolla

Part 3 (1pt): Recall the 1D form for logistic regression is $h(x) = \sigma(\theta_1 x + \theta_0) = p(y = 1 \mid x)$ where $\sigma(z) = \frac{1}{1 + e^{-z}}$. Assuming that after training the values for the parameters are $\theta_1 = 10$ and $\theta_0 = 0$, which label (0 or 1) would $X^{(6)}$ (from the previous question) be given under this model?

$h(x) = \sigma(10x + 0)$

$h(x) = \sigma(10x)$

$\sigma(10x) = \frac{1}{1 + e^{\frac{1}{10x}}}$

$\sigma(10 \cdot 0.55) = \sigma(5.5)$

$\sigma(5.5) = \frac{1}{1 + e^{1/5.5}}$

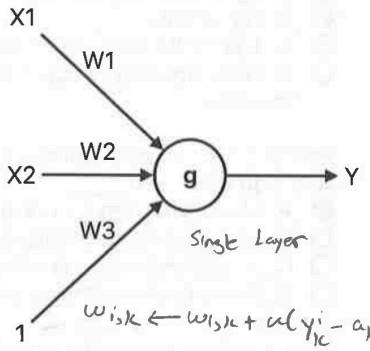
label for x^6 would be 0.

Part 4 (2pts): Notice that by changing from k=3 to k=1, the k-NN prediction for $X^{(4)}$ changes from one label to the other (this is a hint), without any of the other labels changing. Is there a similar way we can change the logistic regression parameters (θ_0, θ_1) from part 3 so that the label for $X^{(4)}$ changes **without any other labels changing**? If yes, give the new (θ_0, θ_1), if no give the two word reason why.

NO, _____ _____.

Question 3: Neural Networks (2 parts)

Consider a single layer neural network with two inputs, a bias input, and a single output, using the step function for the activation function (g(z)=1 if z>0 otherwise 0)



Part 1 (3pts): Perform two iterations of weight updates using the Perceptron Rule. Each weight is initialized to 0 and the learning rate (α) is 1. The first step has been partially completed for you.

$w_{i,k} \leftarrow w_{i,k} + \alpha(y_k^i - a_k) \cdot g'(in_k) x_i^j$

	X1	X2	Y	NN output	W1	W2	W3
Iteration 0	-	-	-	-	0	0	0
Iteration 1	0	1	1	0	0+(1*0)=0	0	0
Iteration 2	1	1	1	0	0	0	0

Part 2 (1pt): Say that we wanted to use Stochastic Gradient Descent instead of the Perceptron Rule, what other part of the network's "architecture" would we have to change? (Answer is two words).

Activation Fonction

Write your name at the top of each separate page, in case pages become separated.
Name: Yisak Tella

Question 4: Multiple choice / True False (10 parts)

Fill in the circle next to the most appropriate answer for the following questions. Each question is worth 1 point.

Part 1: In Machine Learning, overfitting occurs when ____
- ● A. The trained hypothesis fits the training data but does not generalize well to test data.
- ○ B. The trained hypothesis does not fit the training data.
- ○ C. The hypothesis class is not complex enough to represent the target concept.
- ○ D. The trained hypothesis is not sensitive to noise.

Part 2: Which machine learning model classes can use Stochastic Gradient Descent to find the best parameters?
- ○ A. Neural Networks.
- ○ B. Linear Regression.
- ○ C. Logistic Regression.
- ● D. All of the above.

Part 3: How many layers must a Neural Network have in order to represent any computable function (given enough neurons in each layer)?
- ○ A. Three.
- ○ B. 3.
- ○ C. III.
- ● D. All of the above.

Part 4: How does the "forest" based on the Bootstrap Aggregation ("Bagging") technique we discussed for Decision Trees choose which attributes/thresholds are chosen?
- ○ A. By computing the Information Gain of each possible attribute/threshold.
- ● B. Randomly.
- ○ C. Using the Entropy of each data split.
- ○ D. By weighting misclassified data more highly than correctly classified data at the previous iteration.

Part 5: Using a probabilistic approach, we showed that minimizing the sum of squared errors was equivalent to ____
- ● A. Minimizing the L1 norm.
- ○ B. Maximizing generalization performance.
- ○ C. Assuming errors are Gaussian and Independent and Identically Distributed.
- ○ D. Picking a hypothesis with minimum complexity.

Part 6: Using a heuristic which is not admissible is guaranteed to give you a sub-optimal path to the goal. True or false?
- ○ True.
- ● False.

Part 7: Generally speaking, **when** does a search-based agent perform search?
- ● A. Once, at the very beginning before the agent starts taking actions in the environment.
- ○ B. Once, just before the agent reaches the goal to ensure the path taken was optimal.
- ○ C. At every step as the agent moves through the environment towards the goal.
- ○ D. Whenever the agent encounters an unexpected obstacle.

Write your name at the top of each separate page, in case pages become separated.
Name: Yisak Tolla

Part 8: Minimax and alpha-beta pruning use a modified form of which kind of search?
○ A. UCS.
○ B. BFS.
● C. DFS.
○ D. A*.

Part 9: It is possible to modify Generic Search to be able to use **inconsistent** (but admissible) heuristics and still find the correct/optimal solution. True or false?
● True.
○ False.

Part 10: Given two heuristics for a given problem, how would you determine which is "better"?
● A. The heuristic that's always **higher** is better because it expands **fewer** nodes.
○ B. The heuristic that's always **higher** is better because it expands **more** nodes.
○ C. The heuristic that's always **lower** is better because it expands **fewer** nodes.
○ D. The heuristic that's always **lower** is better because it expands **more** nodes.

Extra credit

Which courses do you think the CS department should add to the undergraduate curriculum? Below are three examples, but feel free to add your own. Please provide a ranking from 1 (most preferred) to 5 (least preferred). There are no wrong answers!

Class	Description	Ranking (1-5)
CS 4XX: Machine Learning / Deep Learning / Applied ML	Focused on (circle your choice): A) ML topics not covered in CS 484 (data mining) B) Neural Networks / Deep Learning C) The latest AI models, project focused	1
CS 380: Data Structures w/ AI Focus	A version of Data Structures (CS 310) that uses AI to motivate more advanced CS concepts	1
CS 280: Advanced Python	A class which covers more advanced programming techniques (and possibly some math prerequisites) useful for this class (CS 480)	1
CS 4XX: The Mathematics and Theory of ML	Covering the mathematical foundations and theoretical results of ML topics (formal proofs!)	3
Your Class Here!	Provide a description below if you like! (You'll still get credit if no description provided)	1