Name:

USC ID:

Notes:

- Write your name and ID number in the spaces above.
- No books, cell phones or other notes are permitted. Only two letter size cheat sheet (back and front) and a calculator are allowed.
- Problems are not sorted in terms of difficulty. Please avoid guess work and long and irrelevant answers.
- Show all your work and your final answer. Simplify your answer as much as you can.
- Open your exam only when you are instructed to do so.
- Make sure you submit ALL pages of your answers. Answers submitted after the exam is adjourned WILL NOT BE ACCEPTED.

Problem	Score	Earned
1	25	
2	20	
3	20	
4	25	
5	25	
Total	115	

1. Consider a MLP with one input, two layers, one neuron in each layer, and one output. The activation function of the first layer is $f^{(1)}(n) = \frac{1}{1+e^{-n}}$ and the activation function of the second layer is $f^{(2)}(n) = \frac{e^n - e^{-n}}{e^n + e^{-n}}$. The initial weights of the first and the second layers are respectively $w^{(1)} = -2$ and $w^{(2)} = 1$. There are no bias terms, so $b^{(1)} = b^{(2)} = 0$ and they are kept zero during training. Assume that we present the data point with $x = 3$ and $y = -2$ to the network. Perform one step of the Stochastic Gradient Descent algorithm by using the backpropagation algorithm, assuming the learning rate $\alpha = 0.2$. Use the objective function $J = (y - a^{(2)})^2$. This means that you should calculate the updated weights.

$$x \longrightarrow \boxed{w^{(1)}} \xrightarrow{n^{(1)}} \boxed{f^{(1)}} \xrightarrow{a^{(1)}} \boxed{w^{(2)}} \xrightarrow{n^{(2)}} \boxed{f^{(2)}} \xrightarrow{a^{(2)}}$$

Forward path

$$n^{(1)} = w^{(1)} x = -6$$

$$a^{(1)} = \frac{1}{1+e^{+6}} \approx 0.002$$

$$n^{(2)} = w^{(2)} a^{(1)} = \frac{1}{1+e^{+6}} = .002$$

$$a^{(2)} = \frac{e^{.002} - e^{-.002}}{e^{.002} + e^{-.002}} = .002$$

7

Backward path

$$F'^{(1)}(n^{(1)}) = \left.\frac{df^{(1)}}{dn^{(1)}}\right|_{-6} = \frac{0-(-e^{-n})}{(1+e^{-n})^2} = \frac{e^6}{(1+e^6)^2} \approx 0.002$$

7

$$F'^{(2)}(n^{(2)}) = \left.\frac{df^{(2)}}{dn^{(2)}}\right|_{.002} = \frac{(e^n + e^{-n})^2 - (e^n - e^{-n})^2}{(e^n + e^{-n})^2} \approx 1$$

$$s^{(2)} = -2\,\underbrace{F'^{(2)}(n^{(2)})}_{1}\,\underbrace{(y-a)}_{=-2-.002} = 4.004$$

$$s^{(1)} = F'^{(1)}(n^{(1)})\, w^{(2)}\, s^{(2)} = .002(1)(4.004) = .008$$

6

$$w^{(2)} = -2 - .2\,a^{(1)}\,s^{(2)} = -2.002$$

$$w^{(1)} = 1 - .2\,\underbrace{a^{(0)}}_{x}\,s^{(1)} = 1-(.2)(3)(.008)$$

5

$$= .995$$

2. We are trying to estimate the temperature of consecutive years based on *observations* on tree ring sizes. Possible ring sizes are Very Small = VS, Small = S, Medium = M, Large = L, and Very Large = VL. Years can be Cold = C or Hot = H. Assume that we observed VS, VL tree ring sizes in two consecutive years. Also, Assume that $\pi = [0.2\ 0.8]$ shows the initial distribution of C and H, respectively. Which of the following HMMs is more likely to have given rise to the observation O= {VS,VL} and why? First rows of A_1, B_1, A_2, B_2 represent C and second rows represent H.

(a)

$$A_1 = \begin{array}{c} \begin{array}{cc} C & H \end{array} \\ \begin{bmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{bmatrix} \end{array} \quad B_1 = \begin{array}{c} \begin{array}{ccccc} VS & S & M & L & VL \end{array} \\ \begin{bmatrix} 0.1 & 0.4 & 0.2 & 0.2 & 0.1 \\ 0.3 & 0.2 & 0.1 & 0.1 & 0.3 \end{bmatrix} \end{array}$$

(b)

$$A_2 = \begin{array}{c} \begin{array}{cc} C & H \end{array} \\ \begin{bmatrix} 0.6 & 0.4 \\ 0.5 & 0.5 \end{bmatrix} \end{array} \quad B_2 = \begin{array}{c} \begin{array}{ccccc} VS & S & M & L & VL \end{array} \\ \begin{bmatrix} 0.1 & 0.2 & 0.2 & 0.1 & 0.4 \\ 0.6 & 0.1 & 0.1 & 0.1 & 0.1 \end{bmatrix} \end{array}$$

(a)

State	P(O, State)	
CC	(.2)(.1)(.6)(.1)	$= 12 \times 10^{-4}$
CH	(.2)(.1)(.4)(.3)	$= 24 \times 10^{-4}$
HC	(.8)(.3)(.4)(.1)	$= 96 \times 10^{-4}$
HH	(.8)(.3)(.6)(.3)	$= 432 \times 10^{-4}$

9 pts

$P(O) = 564 \times 10^{-4}$

Solution:

(b)

State	P(O, State)
CC	$(.2)(.1)(.6)(.4) = 48 \times 10^{-4}$
CH	$(.2)(.1)(.4)(.1) = 8 \times 10^{-4}$
HC	$(.8)(.6)(.5)(.4) = 960 \times 10^{-4}$
HH	$(.8)(.6)(.5)(.1) = 240 \times 10^{-4}$

9pts

$P(O) = 1256 \times 10^{-4}$

(b) is a better fit

2pts

3. Assume the following co-training (multiview learning) self-training scenario: the positive class contains $\mathbf{x}_1 = [1\ 0]^T$ and the negative class contains $\mathbf{x}_2 = [0\ 1]^T$. Assume that we first train a maximum margin classifier only based on the first feature of training vectors, then label the unlabeled vector $\mathbf{x}_3 = [2/3\ 1/3]^T$, and then train a maximum margin classifier based on the second feature of $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$. Explain why the class asasociated with the point $\mathbf{x}_4 = [2\ 2]^T$ is *indeterminate* if it is classified using a majority poll between the maximum margin classifier that uses the first feature and the maximum margin classifier that used the second feature for classification.

$\underline{x}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ $\underline{x}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

Maximum margin classifier based on X_1 $X_1 \underset{+}{\overset{-}{\lessgtr}} 1/2$

$x_3 = \begin{bmatrix} 2/3 \\ 1/3 \end{bmatrix}$ classified as $+$,

10 pts

because $2/3 > 1/2$ The new data set $\left\{ \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, + \right), \left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}, - \right), \left(\begin{bmatrix} 2/3 \\ 1/3 \end{bmatrix}, + \right) \right\}$

maximum margin classifier 10 pts

based on X_2: $X_2 \underset{\geq}{\overset{+}{<}} 2/3$

$\begin{bmatrix} 2 \\ 3 \end{bmatrix} \xrightarrow[X_1]{\text{based on}} 2 > 1/1 \rightarrow$ positive

$\xrightarrow[X_2]{\text{based on}} 2 > 2/3 \rightarrow$ neg

Majority polling

indeterminate

4. Suppose that for a particular data set, we perform hierarchical clustering using single linkage (minimal intercluster dissimilarity) and using complete linkage (maximal intercluster dissimilarity). We obtain two dendrograms.

   (a) At a certain point on the single linkage dendrogram, the clusters $\{1, 3, 5\}$ and $\{8, 9\}$ fuse. On the complete linkage dendrogram, the clusters $\{1, 3, 5\}$ and $\{8, 9\}$ also fuse at a certain point. Which fusion will occur higher on the tree, or will they fuse at the same height, or is there not enough information to tell?

   (b) At a certain point on the single linkage dendrogram, the clusters $\{8\}$ and $\{9\}$ fuse. On the complete linkage dendrogram, the clusters $\{8\}$ and $\{9\}$ also fuse at a certain point. Which fusion will occur higher on the tree, or will they fuse at the same height, or is there not enough information to tell?

(a) Not enough information to tell.

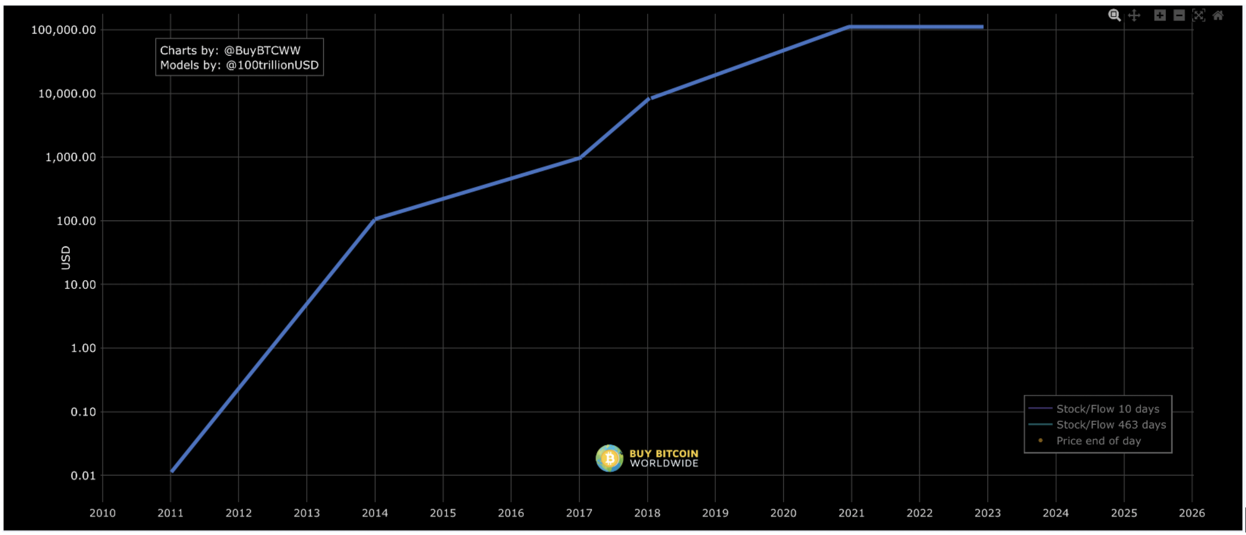
Reason: The maximal intercluster dissimilarity could be equal or not equal to the minimal intercluster dissimilarity. If the dissimilarities were equal, they would fuse at the same height. If they were not equal, the single linkage dendrogram would fuse at a lower height.

10 points for correct answer
3 points for correct reason

b) They would fuse at the same height. Reason: because linkage does not affect leaf-to-leaf fusion

6 pts for correct answer
6 pts for correct reason

5. The bitcoin Stock to Flow (S2F) model was created by the famous twitter user PlanB. S2F models the price of bitcoin based on its rarity. A slightly modified version of S2F is shown below. Note that the dependent variable y is $\log_{10} price$ and the independent variable is time in years since 2010 t; therefore $t \in [1, 13]$. Show this model using a decision tree and clearly determine the internal nodes and terminal nodes. Remember that this is a *model* tree, so the terminal nodes may contain *regression models*.



$1 \leq t \leq 4$ $\qquad \log_{10}\text{price} - (-2) = \dfrac{2-(-2)}{4-1}(t-1)$

$\qquad \Rightarrow \log_{10}\text{price} = \dfrac{4}{3}t + \dfrac{8}{3}$

$4 < t \leq 7$ $\qquad \log_{10}\text{price} - 2 = \dfrac{3-2}{7-4}(t-4)$

$\qquad \log_{10}\text{price} = \dfrac{1}{3}t + \dfrac{2}{3}$

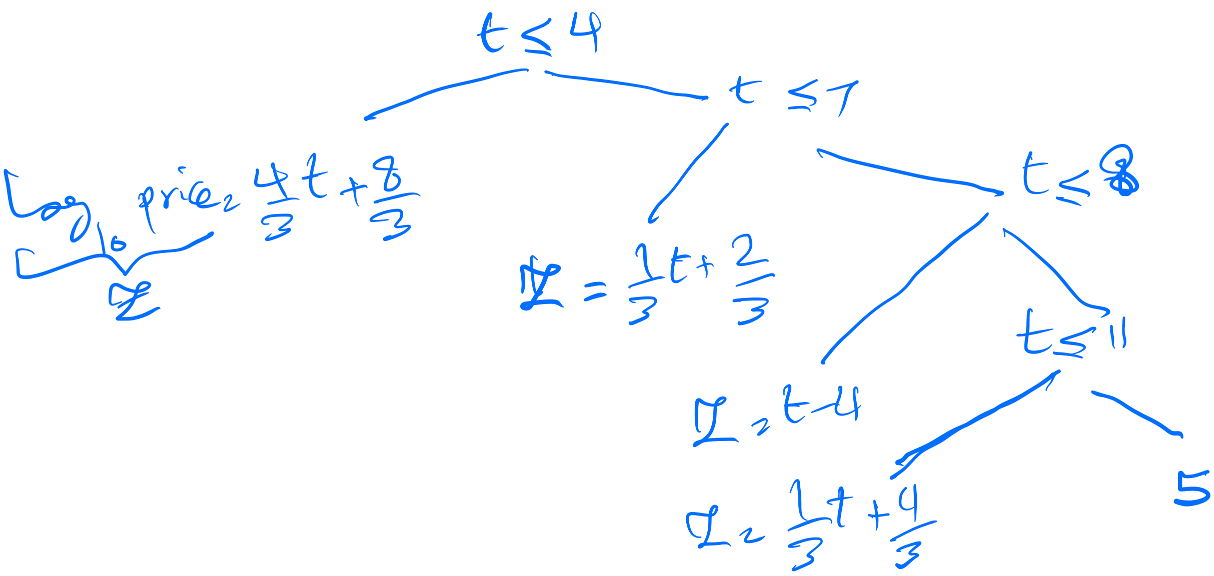
$7 < t \leq 8$ $\qquad \log_{10}\text{price} - 3 = \dfrac{4-3}{8-7}(t-7)$

$\qquad \Rightarrow \log_{10}\text{price} = t - 4$

$8 < t \leq 11$ $\qquad \log_{10}\text{price} - 4 = \dfrac{5-4}{11-8}(t-8)$

$\qquad \log_{10}\text{price} = \dfrac{1}{3}t + \dfrac{4}{3}$

$11 < t < 13$ $\qquad$ $\log_{10} \text{Price} = 5$ $\qquad$ } 13 pts



12 pts

terminal nodes = models

internal nodes = ranges of t

Solution:

Scratch paper

Name:

USC ID:

Scratch paper

Name:

USC ID: