Name:

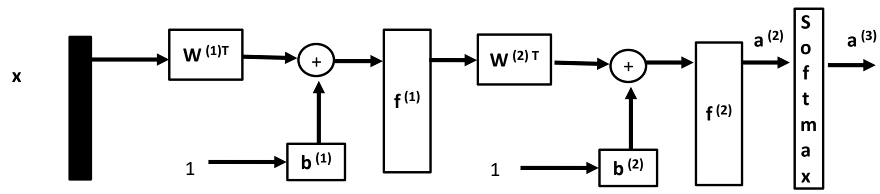
USC ID:

Notes (Please read very carefully):

- Write your name and ID number in the spaces above.
- No books, cell phones or other notes are permitted. Only two letter size cheat sheets (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.
- The exam has 5 questions, 12 pages, and 25 points extra credit.
- In online exams, legible copies SCANNED via phone applications must be submitted in one pdf file, not pictures of answer sheets.
- 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	25	
3	25	
4	25	
5	25	
Total	125	

1. Consider the following MLP



where

$$\mathbf{W}^{(1)} = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 2 & 1 \end{bmatrix}$$
$$\mathbf{W}^{(2)} = \begin{bmatrix} -1 & 2 & -1 \\ 1 & 2 & 0 \end{bmatrix}$$
$$\mathbf{b}^{(1)} = \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}, \ \mathbf{b}^{(2)} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix},$$

Also assume that the components of $\mathbf{f}^{(1)}$, and $\mathbf{f}^{(2)}$ are linear, i.e. is $f_1(x) = f_2(x) = x$. The second layer is followed by a softmax layer, and the output of each neuron in the softmax layer is calculated as:

$$a_j^{(3)} = \frac{e^{a_j^{(2)}}}{e^{a_1^{(2)}} + e^{a_2^{(2)}} + e^{a_3^{(2)}}}$$

where $a_j^{(3)}$ is the output of j^{th} neuron in the third layer and $a_j^{(2)}$ is the output of the j^{th} neuron in the second layer.

Answer the following questions:

(a) How many neurons are there in the first, and second layers, respectively?

(b) How many classes (K) does the classification problem performed by this network involve?

(c) Explain why this network is exactly equivalent to performing multinomial regression using linear logits and write down $p_k(\mathbf{x}) = \Pr(Y = k|\mathbf{X} = \mathbf{x})$ for $k = 1, \ldots, K$.

Solution:

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 $\pi = [0.3\ 0.7]$ shows the initial distribution of C and H, respectively. Calculate the probability $P(O)$ that each of following HMMs have given rise to the observation O= {VL,S} to determine which one is more likely to be fit to the observation. First rows of A_1, B_1, A_2, B_2 represent C and second rows represent H.

(a)
$$\begin{array}{cc} C & H \end{array} \qquad \begin{array}{ccccc} VS & S & M & L & VL \end{array}$$

$$A_1 = \begin{bmatrix} 0.7 & 0.3 \\ 0.1 & 0.9 \end{bmatrix} \quad B_1 = \begin{bmatrix} 0.2 & 0.3 & 0.3 & 0.1 & 0.1 \\ 0.2 & 0.2 & 0.2 & 0.3 & 0.1 \end{bmatrix}$$

(b)
$$\begin{array}{cc} C & H \end{array} \qquad \begin{array}{ccccc} VS & S & M & L & VL \end{array}$$

$$A_2 = \begin{bmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{bmatrix} \quad B_2 = \begin{bmatrix} 0.1 & 0.2 & 0.2 & 0.1 & 0.4 \\ 0.5 & 0.2 & 0.1 & 0.1 & 0.1 \end{bmatrix}$$

Solution:

3. Consider a dataset with 3 points in 1-D:

Index	X	Y
1	0	+
2	-1	-
3	+1	-

(a) Carefully sketch these three training points. Are the classes linearly separable?

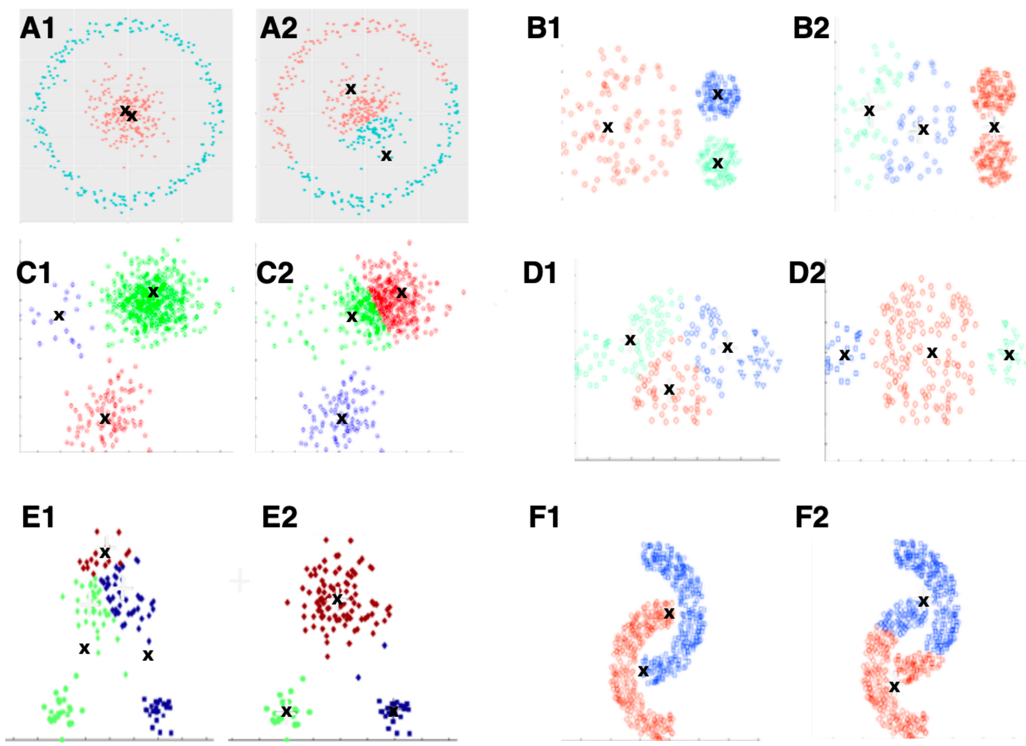
(b) Consider mapping each point to 2-D using new feature vectors $\boldsymbol{\varphi}(x) = [u_1(x), u_2(x)]^T$, in which $u_1(x)$ and $u_2(x)$ are polynomial functions of x. Find a $\boldsymbol{\varphi}(x)$ such that data are linearly separable in the new feature space.

(c) The maximum margin classifier in the new space has the equation $\mathbf{w}^T\boldsymbol{\varphi}(x)+b = 0$. Find $\mathbf{w}$ and b. Determine the decision boundary in the original 1-D space and the class assigned to $x = \frac{1}{3}$ by the classifier.

(d) Find a two layer feedforward neural network whose input-output equation is exactly like the above maximum margin classifier, by determining $\mathbf{W}^{(1)}, \mathbf{b}^{(1)}$ and $\mathbf{W}^{(2)}, \mathbf{b}^{(2)}$, and $\mathbf{f}^{(1)}$ and $\mathbf{f}^{(2)}$.

Hint: in this neural network, elements of $\mathbf{f}^{(1)}$ *are different polynomial functions.*

4. There are 6 different datasets noted as A,B,C,D,E,F. Each dataset is clustered using two different methods, and one of them is K-means. All results are shown in figure below. You are required to determine which result is more likely to be generated by K-means method. (Hint: check the state when K-means converges; Centers for each cluster have been noted as x; Since x and y axis are scaled proportionally, you can determine the distance to centers geometrically). The distance measure used here is the Euclidean distance.



(a) Dataset A (write A1 or A2)

(b) Dataset B (write B1 or B2)

(c) Dataset C (write C1 or C2)

(d) Dataset D (write D1 or D2)

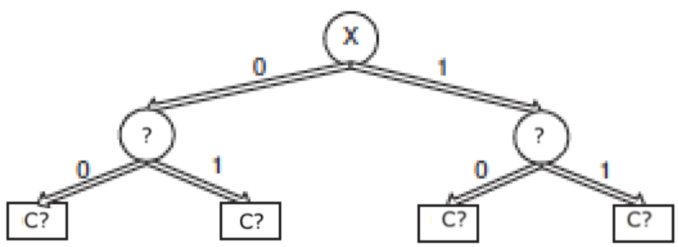
(e) Dataset E (write E1 or E2)

(f) Dataset F (write F1 or F2)

5. Consider the following set of training examples.

X	Y	Z	No. of Class C1 Examples	No. of Class C2 Examples
0	0	0	5	40
0	0	1	0	15
0	1	0	10	5
0	1	1	45	0
1	0	0	10	5
1	0	1	25	0
1	1	0	5	20
1	1	1	0	15

Compute a two-level decision tree, and choose X as the root, as shown in the figure.



Use the classification error rate as the criterion for splitting at the second level and determine whether Y or Z should be used in the second level. To do so, fill in the following tables, assuming that $X = 0$

Y	C1	C2
0		
1		

Z	C1	C2
0		
1		

and the following tables, assuming $X = 1$

Y	C1	C2
0		
1		

Z	C1	C2
0		
1		

The above tables assist in finding the variable splitting which provides better classification error rate for each side of the tree, i.e. for $X = 0$ and $X = 1$. Determine the classes in the leaf nodes. What is the classification error rate of the tree you found on training data?

Solution:

Scratch paper

Name:

USC ID:

Scratch paper

Name:

USC ID: