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 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, 10 pages, and 20 points extra credit. However, your grade cannot exceed 100/100.
- In online exams, legible copies that are SCANNED via phone applications or created using tablets must be submitted, 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	20	
4	25	
5	25	
Total	120	

1. We have a dataset whose points have two features for binary classification: $\left\{\mathbf{x}_1 = \begin{bmatrix}0\\0\end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix}0\\1\end{bmatrix}\right\}$, with $y = 1$ and $\left\{\mathbf{x}_3 = \begin{bmatrix}2\\0\end{bmatrix}, \mathbf{x}_4 = \begin{bmatrix}2\\2\end{bmatrix}\right\}$ with $y = -1$. Using the initial weights $\mathbf{w}^T = [0\ 0]$ and initial bias $b = 0$, run the standard perceptron algorithm ($\alpha = 0.5$) until it converges and find the weights and bias that perfectly classify the training set. Always present the data to the perceptron in the following order: $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4$.

$f(\underline{w}^T \underline{x}_1 + b) = f([0\ \ 0]\begin{bmatrix}0\\0\end{bmatrix} + 0) = f(0) = 1$ 2.5

$y = 1 \quad e = 0 \rightarrow$ no change

$f(w^T \underline{x}_2 + b) = f([[0\ \ 0]\begin{bmatrix}0\\1\end{bmatrix} + 0] = f(0) = 1$

$y = 1 \quad e = 0 \longrightarrow$ no change 2.5

$f(\underline{w}^T \underline{x}_3 + b) = f([0\ \ 0]\begin{bmatrix}2\\0\end{bmatrix} + 0) = 1$

$y = -1 \quad e = -2$

2.5

$w = \begin{bmatrix}0\\0\end{bmatrix} + (.5)(-2)\begin{bmatrix}2\\0\end{bmatrix} = \begin{bmatrix}-2\\0\end{bmatrix}$

$b = 0 + .5(-2) = -1$

$f(\underline{w}^T \underline{x}_4 + b) = f([-2 \;\; 0]\begin{bmatrix}2\\2\end{bmatrix} - 1) = -1$

$y = -1$ $\quad e = 0 \quad$ no change $\quad$ 2.5

$f(\underline{w}^T \underline{x}_1 + b) = f([-2 \;\; 0]\begin{bmatrix}0\\0\end{bmatrix} - 1) = -1$

$y = 1$ $\quad e = +2$ $\quad$ 2.5

$\underline{x}_1 = 0$ no change in $\underline{w}$

$b = -1 + .5(2) = 0$

$f(\underline{w}^T \underline{x}_2 + b) = f([-2 \;\; 0]\begin{bmatrix}0\\1\end{bmatrix} + 0) = 1$

$y = 1$ no change $\quad$ 2.5

$f(\underline{w}^T \underline{x}_3 + b) = f([-2 \;\; 0]\begin{bmatrix}2\\1\end{bmatrix} + 0) = -1$

$y = -1$ no change $\quad$ 2.5

$$f(\underline{w}^T \underline{x}_4 + b) = f\left(\begin{bmatrix}-2 & 0\end{bmatrix}\begin{bmatrix}3\\2\end{bmatrix} + 0\right)$$

$$= -1$$

$y = -1$ no change 2.5

$$f(\underline{w}^T \underline{x}_1 + b) = f(6) = 1$$

$y = 1$ no change. 2.5

So $\underline{w} = \begin{bmatrix}-2\\0\end{bmatrix}$ $b = 0$

2.5

Be lenient on numerical mistakes. The above is just a guideline

2. Consider the Hidden Markov Model $\lambda = (A, B, \pi)$ with hidden states D, E, F and possible observation symbols α, β, γ, where $\pi = [1\ 0\ 0]$ and

$$A = \begin{array}{c} \begin{array}{ccc} D & E & F \end{array} \\ \begin{bmatrix} 0.2 & 0.8 & 0.0 \\ 0.0 & 0.8 & 0.2 \\ 0.4 & 0.0 & 0.6 \end{bmatrix} \end{array} \qquad B = \begin{array}{c} \begin{array}{ccc} \alpha & \beta & \gamma \end{array} \\ \begin{bmatrix} 0.8 & 0.2 & 0.0 \\ 0.0 & 0.6 & 0.4 \\ 0.2 & 0.0 & 0.8 \end{bmatrix} \end{array}$$

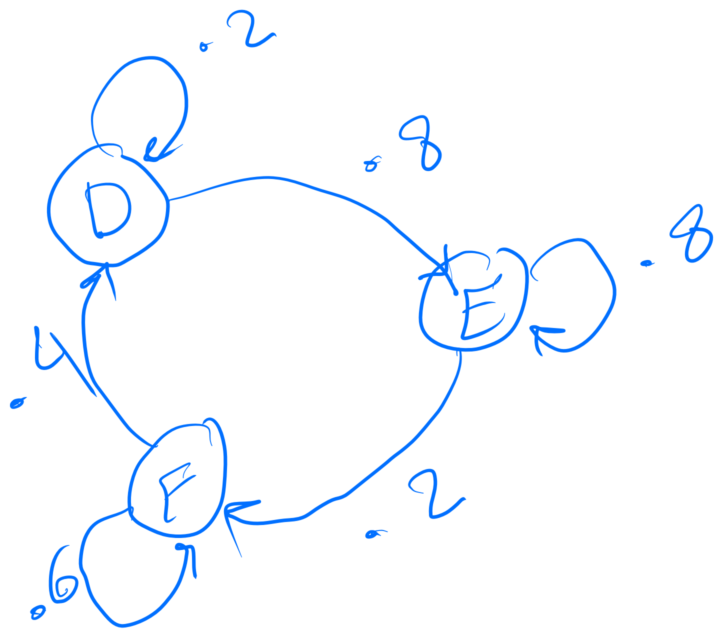
(a) Draw the state diagram of this HMM and show the transition probabilities.

(b) For the sequence of observations $O = (\alpha, \beta, \gamma, \alpha)$, give all sequences of states $X = (X_1, X_2, X_3, X_4)$ for which $P(X, O) \neq 0$. There are only four such sequences!

(c) Calculate $P(O)$.

(d) What is the most likely $X = (X_1, X_2, X_3, X_4)$?

(a)



5pts

(b) Possible state transitions

~~DDDD~~

~~DDDE~~

~~DDEE~~

✓ DDEF

~~DEEE~~

✓ DEEF

✓ DEFF

✓ DEFD

Solution:

$$A = \begin{array}{c} \begin{array}{ccc} D & E & F \end{array} \\ \begin{bmatrix} 0.2 & 0.8 & 0.0 \\ 0.0 & 0.8 & 0.2 \\ 0.4 & 0.0 & 0.6 \end{bmatrix} \end{array} \quad B = \begin{array}{c} \begin{array}{ccc} \alpha & \beta & \gamma \end{array} \\ \begin{bmatrix} 0.8 & 0.2 & 0.0 \\ 0.0 & 0.6 & 0.4 \\ 0.2 & 0.0 & 0.8 \end{bmatrix} \end{array}$$

$P(DDEF, O) = (1)(.8)(.2)(.2)(.8)(.4)(.2)(.2)$

$= 4 \times 10^{-4}$

$P(DEEF, O) = (1)(.8)(.2)(.8)(.6)(.8)(.2)(.2)$

$= 25 \times 10^{-4}$

10 pts

$P(DEFF, O) = (1)(.8)(.8)(.6)(.2)(.8)(.6)(.2)$

$= 74 \times 10^{-4}$

$P(DEFD, O) = (1)(.8)(.8)(.6)(.2)(.8)(.4)(.8)$

$= 197 \times 10^{-4}$

(c) $P(O) = \sum_{X} P(X, O) = (197 + 74 + 25 + 4) \times 10^{-4}$

$= .03$

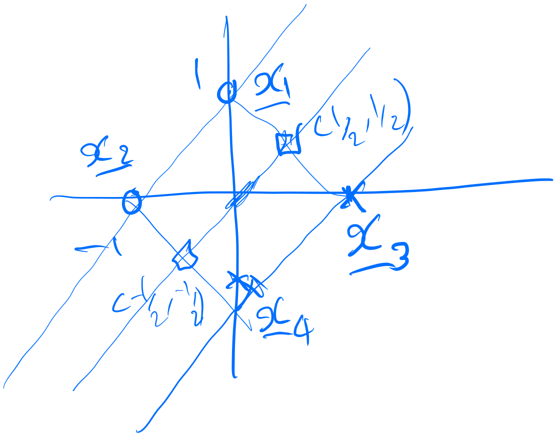
5 pts

(d) DEFD is the most probable path

5 pts

3. We wish to classify the following two dimensional dataset using a Maximum Margin Classifier:
$\left\{\mathbf{x}_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} -1 \\ 0 \end{bmatrix}\right\}$, with $y = 1$ and $\left\{\mathbf{x}_3 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \mathbf{x}_4 = \begin{bmatrix} 0 \\ -1 \end{bmatrix}\right\}$ with $y = -1$.
   (a) Determine the equation of the linear decision boundary $f(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$.
   (b) Determine the set of support vectors $\mathcal{S}$.

   Hint: you do not need to solve an optimization problem to answer this question.



(a) $x_1 - x_2 = 0$

10 pts

(b) $\mathcal{S} = \{\underline{x}_1, \underline{x}_2, \underline{x}_3, \underline{x}_4\}$

10 pts

4. We have a dataset whose points have two features: $\{\mathbf{x}_1 = (0,0), \mathbf{x}_2 = (0,1), \mathbf{x}_3 = (2,0), \mathbf{x}_4 = (2,2)\}$.
   (a) Using the K-means algorithm, cluster the data into two groups. Initially, assign $\mathbf{x}_1$ and $\mathbf{x}_4$ to cluster one and $\mathbf{x}_2$ and $\mathbf{x}_3$ to cluster 2.
   (b) Using Euclidean distance and single linkage, perform hierarchical clustering on the dataset and show the results by a dendrogram.

(a)

$\underline{x}_1, x_4$ cluster 1 $\longrightarrow$ centroid : $\frac{1}{2}\left((0,0)+(2,2)\right) = (1,1)$

$\underline{x}_2, \underline{x}_3$ cluster 2 $\rightarrow$ centroid $\frac{1}{2}\left((0,1)+(2,0)\right)$
$= (1, \frac{1}{2})$

Recluster: $d(\underline{x}_1, \underline{c}_1) = \sqrt{1^2+1^2} = \sqrt{2}$ $\longrightarrow$ x_1 cluster 1
$d(\underline{x}_1, \underline{c}_2) = \sqrt{1^2 + \frac{1}{2}^2} = \sqrt{\frac{5}{4}}$

$d(\underline{x}_2, \underline{c}_1) = 1$ $\longrightarrow$ x_2 cluster 1
$d(\underline{x}_2, \underline{c}_2) = \sqrt{1^2 + \frac{1}{2}^2} = \sqrt{\frac{5}{4}}$

12 pts

$d(x_3, \underline{c}_1) = \sqrt{2}$ $\longrightarrow$ x_3 cluster 2
$d(\underline{x}_3, \underline{c}_2) = \sqrt{\frac{5}{4}}$

$d(\underline{x}_4, \underline{c}_1) = \sqrt{2}$ $\longrightarrow$ x_4 cluster 2
$d(\underline{x}_4, \underline{c}_2) = \sqrt{\frac{5}{4}}$

New centroids: $\underline{c}_1 = (0, \frac{1}{2})$
$\underline{c}_2 = (2, 1)$

$d(\underline{x}_1, \underline{c}_1) = \frac{1}{2}$ → cluster 1

$d(\underline{x}_2, \underline{c}_2) = \sqrt{5}$

$d(\underline{x}_2, \underline{c}_1) = \frac{1}{2}$ → cluster 1

$d(\underline{x}_2, \underline{c}_2) = 2$

$d(\underline{x}_3, \underline{c}_1) = \sqrt{2^2 + \left(\frac{1}{2}\right)^2}$

$d(\underline{x}_3, \underline{c}_2) = 1$ → Cluster 2

$d(\underline{x}_4, \underline{c}_1) = \sqrt{2^2 + \left(\frac{7}{4}\right)^2}$

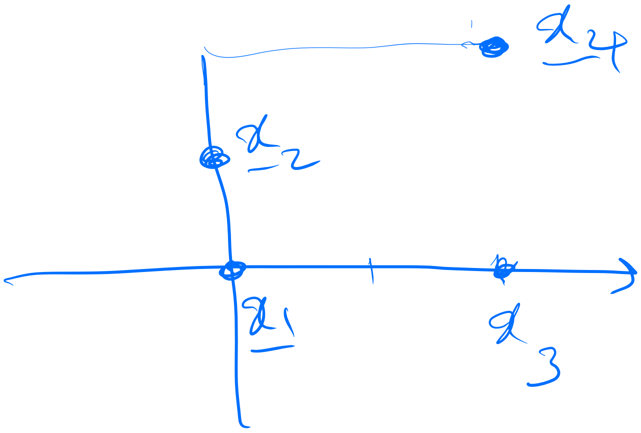
$d(\underline{x}_4, \underline{c}_2) = 1$ → Cluster 2

cluster's don't change

So cluster 1 = $\{\underline{x}_1, \underline{x}_2\}$

cluster 2 = $\{\underline{x}_3, \underline{x}_4\}$

(b)



13 pts

Hierarchical

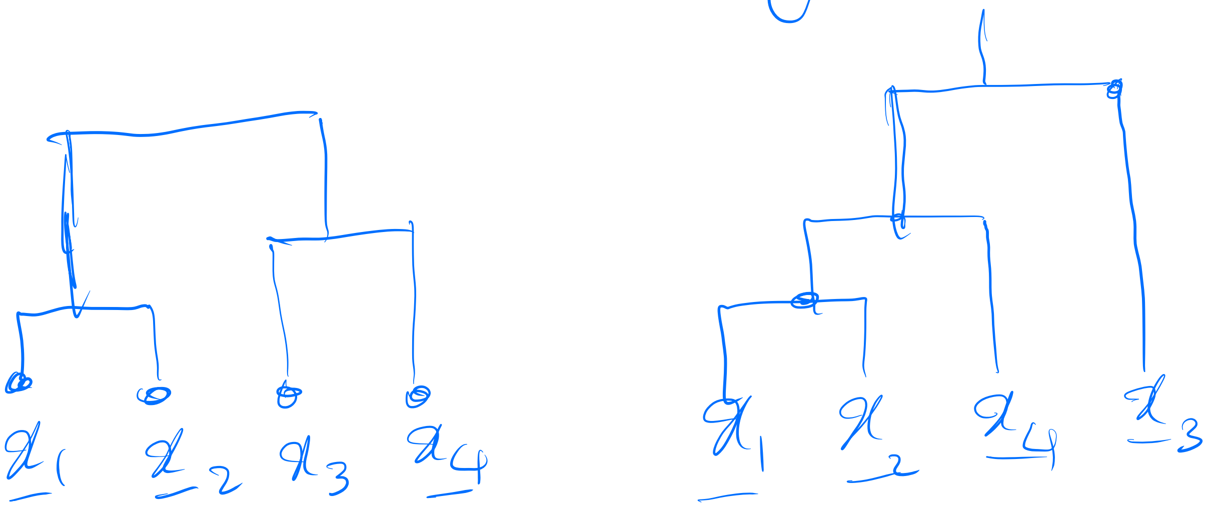
$\{\underline{x}_1\}, \{\underline{x}_2\}, \{\underline{x}_3\}, \{\underline{x}_4\}$

$\longrightarrow \underset{C_1}{\{\underline{x}_1, \underline{x}_2\}}, \underset{C_2}{\{\underline{x}_3\}}\underset{C_3}{\{\underline{x}_4\}}$

$$\left.\begin{array}{l} d(C_1, C_2) = \sqrt{5} \\ d(C_1, C_3) = 2 \\ d(C_2, C_3) = 2 \end{array}\right\} \begin{array}{l} \longrightarrow \{\underline{x}_1, \underline{x}_2, x_4\}, \{\underline{x}_3\} \\ \text{either} \quad \text{or} \\ \longrightarrow \{\underline{x}_1, \underline{x}_2\}, \{\underline{x}_3, \underline{x}_4\} \end{array}$$

$\longrightarrow \{\underline{x}_1, \underline{x}_2, \underline{x}_3, x_4\}$

Two Dendrograms



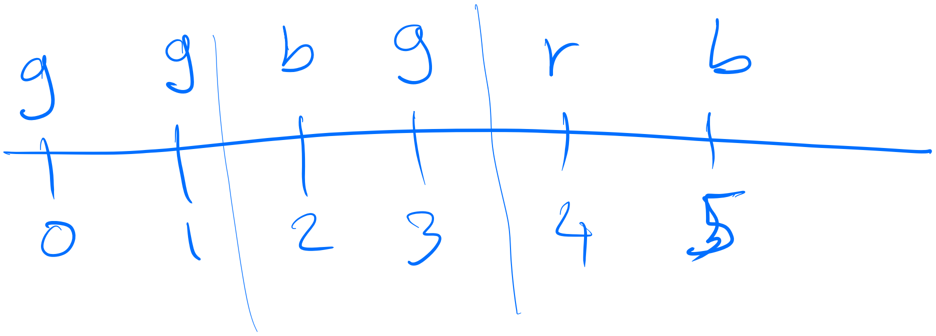
One of them is enough

5. Assume the dataset with one feature and three classes:

   $\{(0, green), (1, green), (2, blue), (3, green), (4, red), (5, blue)\}$.

   Build a decision tree with three leaves using recursive binary splitting and the Gini index whose formula for region R_m is $\sum_{k=1}^{K} \hat{p}_{mk}(1 - \hat{p}_{mk})$, where $\hat{p}_{mk}$ is the etsimated probability of class k in region R_m, where $k = 1, 2, \cdots, K$ and K is the number of classes. Assume that the feature space can only be splitted at 1.5, 2.5, and 3.5. Sketch the diagram of the decision tree. Do NOT use weighted Gini index over the regions. Simply add the Gini indices of the regions of the decision tree.

   Hint: How to break possible ties? If you ever encountered ties, consider $green = 100, blue = 200, red = 300$. Break the tie in favor of the smaller number. So the tie between green and red would be broken in favor of green.



Th : 1.5 Gini = $0 + \frac{2}{4}\left(1-\frac{2}{4}\right) + \frac{1}{4}\left(\frac{3}{4}\right) + \frac{1}{4}\left(\frac{3}{4}\right)$

$= \frac{1}{4} + \frac{3}{16} + \frac{3}{16} = \frac{10}{16}$

16pts first split

Th = 2.5 Gini = $\frac{2}{3} \times \frac{1}{3} + \frac{1}{3} \times \frac{2}{3} + \frac{1}{3} \times \frac{2}{3} + \frac{1}{3} \times \frac{2}{3} + \frac{1}{3} \times \frac{2}{3}$

$= \frac{10}{9}$

Th = 3.5 Gini = $\frac{3}{4} \times \frac{1}{4} + \frac{1}{4} \times \frac{3}{4} + \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2}$

Solution:

$$= \frac{3}{16} + \frac{3}{16} + \frac{1}{4} + \frac{1}{4} = \frac{14}{16}$$

Th = 1.5 has the lowest Gini

Second split [10pts]

$$Th = 2.5 \quad Gini = 0 + 0 + 3 \times \frac{1}{3} \times \frac{2}{3} = \frac{2}{3}$$

$$Th = 3.5 \quad Gini = 0 + \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2} = 1$$

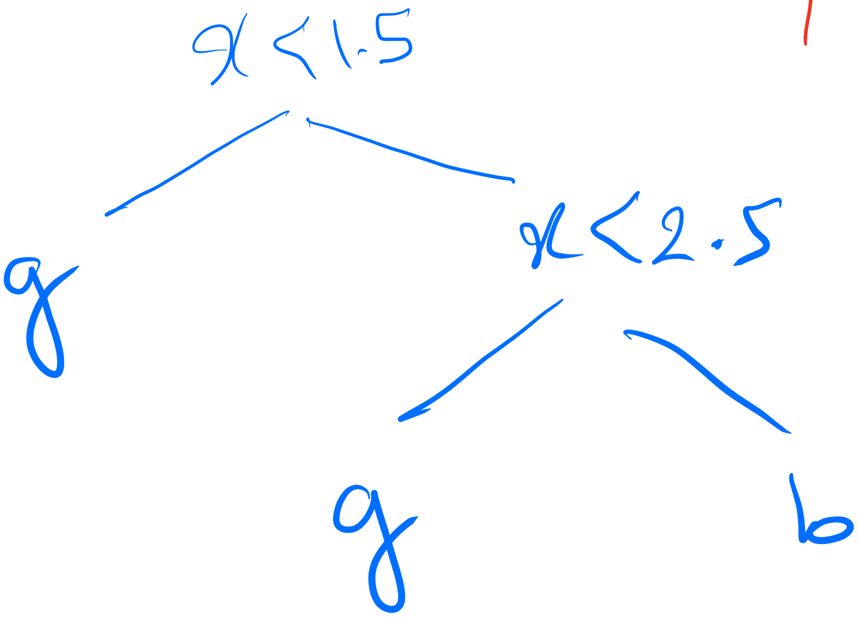
Th = 2.5 has the lowest Gini

Scratch paper

Name:

USC ID:

5 pts



Scratch paper

Name:

USC ID: