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.
- The exam has 5 questions, 9 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. The input image matrix of a convolutional neural network is $[1\ 0\ 2]$ and the filter (kernel) matrix is $[1\ -1]$, which has a stride of 1. The resulting feature map is passed through a ReLU layer to yield a rectified feature map. No max pooling layer is used. The output (which is a row vector) is then flattened to become a column vector (this can be simply done by transposing this output, i.e.the row vector, to become a column vector). The flattened vector is then passed through a layer with a 2×2 weight matrix $\mathbf{W}$ whose all elements are 0's and 1's. The activation functions of this layer are linear $f(x) = x$ and it bias terms are 0's. Determine $\mathbf{W}$ such that the input is classified in class 2, encoded as $[0\ 1]^T$. This network does not have a softmax layer after the linear layer, although the answer would not change if it had a softmax layer. Hint: this problem has more than one solution.

2. Consider the following data set: In class 1, we have $[1\ 1]^T, [1\ 2]^T, [2\ 2]^T$. In class 2, we have $[1.5\ 1.5]^T$.

   (a) Sketch the data set and determine whether or not it is linearly separable.

   (b) Regardless of the answer to 2a, find a quadratic feature $X_3 = f(X_1, X_2) = aX_1^2 + bX_2^2 + cX_1X_2 + dX_1 + eX_2 + f$, that makes the data linearly separable; that is, $X_3 \geq 0$ for members of class 1, and $X_3 < 0$ for members of class 2. Find the maximum margin classifier only based on X_3. Hint: The equation of the maximum margin classifier based on only one feature is $X_3 = \beta_0$ and you should determine β_0.

   (c) By solving $X_3 = f(X_1, X_2) = \beta_0$ for X_2, find the equation of the decision boundary in the original feature space and sketch it. Show the regions in the feature space that are classified as class 1 and class 2. You do not need to be very precise.

3. Consider the following unlabeled dataset:

$$\left\{\mathbf{x}_1 = \begin{bmatrix}1\\1\end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix}-1\\-1\end{bmatrix}\right\}, \left\{\mathbf{x}_3 = \begin{bmatrix}1\\-1\end{bmatrix}, \mathbf{x}_4 = \begin{bmatrix}-1\\1\end{bmatrix}\right\}.$$

   (a) Determine the (loading) vector φ_1, $\|\varphi_1\|_2 = 1$ that represents the first principal direction of the data.

   (b) Determine the (loading) vector φ_2, $\|\varphi_2\|_2 = 1$ that represents the first principal direction of the data.

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

4. We have a dataset whose points have one features and are in two classes: $\{-4, 2, 3\}$ are in the negative class and $\{-3, -2, 1\}$ in the positive class.

   (a) Disregard the labels and cluster the data in two groups using K-means. Label all data in each cluster using the label of the center of the cluster. What is the misclassification error rate (%) of the classifier you built?

   (b) Assume that we have a simple threshold classifier that classifies a data point x to the positive class if $x \geq -1.5$ and in the negative class otherwise. Assume that we are implementing *active learning* with uncertainty sampling, so we ask the "oracle" to give us the label of the most uncertain data point and update the threshold of our classifier only if it misclassifies the new point. We only use the threshold among $\{-3.5, -2.5, -1.5, -0.5, 0.5, 1.5, 2.5\}$ that is the closest to the newly labeled data point and fixes its misclassification. Run the algorithm three iterations (i.e. by asking the labels of three data points from the oracle) and provide the final classifier.

5. Assume the following dataset with one feature and one quantitative (continuous) output, i.e. data are in the form (x_i, y_i):

   $\{(0,1),(1,2),(2,-1),(3,2),(4,0),(5,4)\}$.

   Build a decision tree with three leaves using recursive binary splitting and the Sum of Absolute Errors criterion whose formula for region R_m is $\sum_{j=1}^{|R_m|} |y_j - \hat{y}_{R_m}|$, where $\hat{y}_{R_m}$ is the average response in region R_m, and $|R_m|$ is the number of data points in region R_m. Assume that the feature space can only be splitted at 1.5, 2.5, and 3.5. Sketch the diagram of the decision tree.

Solution:

Scratch paper

Name:

USC ID:

Scratch paper

Name:

USC ID: