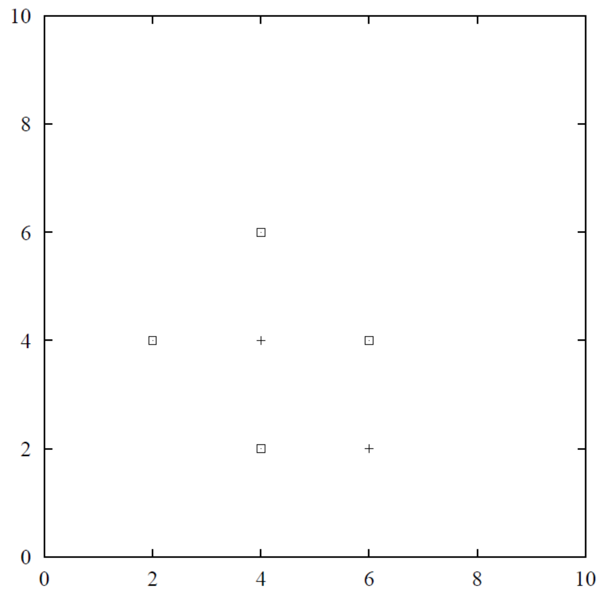
CS 6375 Homework Assignment 2 - Due Monday February 24, 2020

a. (4 points) In our class, we filled the table of criteria for comparing the different learning methods - the type of classifier (linear vs non-linear), hypothesis space (fixed vs flexible), type of learning algorithm (lazy vs eager), online vs batch and robustness to outliers. Fill in the table for Nearest neighbors, Neural Networks and Naive Bayes. Justify in NO MORE than 1 sentence for each choice.

b. (3 points) Consider the set of training examples shown in the diagram below. Draw the decision



boundaries for the nearest neighbor algorithm. Assume Euclidean distance to compute the nearest neighbors. Plus indicates positive example and the small square indicates negative example. How will points (8,1) and (8,8) be classified?

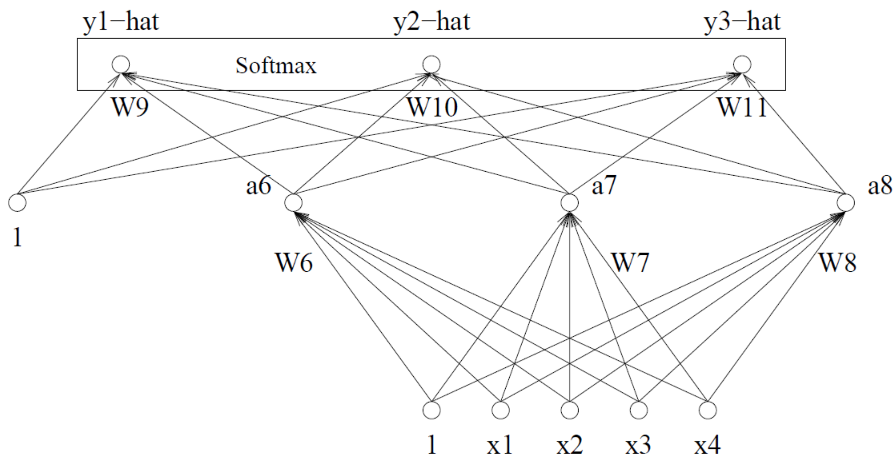
c. (4 points) Consider the set of training examples below. What feature will be used to split at the root of the decision tree if we use mutual information as the splitting criterion? Show your work.

x_1	x_2	y
0	1	0
1	1	0
0	0	1
1	0	0
0	1	0
0	1	1

d. (4 points) For the same data set (presented above) learn a Naive Bayes Classifier. Estimate the probabilities as fractional counts. Make predictions for $(x_1 = 0, x_2 = 0)$ and $(x_1 = 1, x_2 = 0)$.

e. (10 points) Consider a Neural network that uses a softmax activation function for its output layer. Its outputs can be interpreted as posterior probabilities $P(y|x)$ for a categorical target variable y. Consider the following neural network with three output units. The softmax activation function is defined as $\hat{y_i} = \frac{exp(x_i)}{\sum_{j=1} 3exp(x_j)}$ (as against what we saw in class $\hat{y_i} = \frac{1}{1+exp(-x_i)}$), where x_i is the net input for the activation function of the output node i. Note that using this function will enable us to get $\hat{y_1} + \hat{y_2} + \hat{y_3} = 1$, making this a valid posterior distribution. In this problem, you are required to compute the derivatives needed for the backpropagation algorithm for this kind of network.

- Write down the log likelihood objective function $J(\mathbf{w})$ for this network, where $\mathbf{w}$ is the concatenation of W6, W7, W8, W9, W10, W11. Assume that each training example has the form



$(\mathbf{x}, y)$ where $\mathbf{x} = (1, x_1, x_2, x_3, x_4)$ and $y = (y_1, y_2, y_3)$. Note that there are only three y values: $y = (1, 0, 0)$,$y = (0, 1, 0)$ and $y = (0, 0, 1)$.

- Compute $\frac{\partial j}{\partial 9,6}$