CS 440 Final Exam

Question 1 [20]:

1. Consider a car on a three lane highway. We will model this system as a Markov chain with the following properties.

If the car is in the right lane there is a .7 probability he will remain in the right lane and a .3 probability it will move into the center lane.

If the car is in the center lane there is a .3 probability it will move to the right lane, a .5 probability it will remain in the center lane and a .2 probability it will move into the left lane.

If the car is in the left lane there is a .6 probability it will remain in the left lane and a .4 probability it will move into the center lane.

a) Draw the transition diagram for this system. This diagram should include a node for each state (i.e. for each lane) and a directed edge from each state to each other state indicating the probability of transitioning between the states,

b) If the car is in the right lane at step 0, what is the probability it will be in each lane at step 3.

c) The car is observed to be in the center lane at step 4. What is the probability that it was in each lane at step three given this observation.

Question 2 [20]:

Consider a student, James, who has a choice to buy or not buy a \$200 textbook for a class. If he buys the textbook there is a If he buys the textbook there is a .9 probability he will master the course material. If he doesn't by the book there is a .7 probability he will master the material.

$p(m|b) = .9$
$p(m|\neg b) - .7$

Where b is a Boolean variable indicating if James decided to buy the textbook and m is a Boolean variable indicating if James has mastered the material.

The course's grade will be determined by an open book take home final and the probability that James will pass is given by the following.

$p(p|m,b) = .95$
$p(p|m,\neg b) = .8$
$p(p|\neg m,b) = .6$
$p(p|\neg m,\neg b) = .3$

Where p is a Boolean variable indicating if James passes the class.
The utility of passing the class is \$2500.

a) What is the expected reward for buying the textbook vs. not buying the textbook. Should James but it?

b) Consider a variation of this problem where James also has the option to buy the book after the class but before the final. Again, if James chooses to buy the book at the beginning of the class the probability that he will master the material is .9. If he does not choose to buy the book at the beginning of the class there is a .7 probability he will master the material (regardless if the chooses to buy the book after the class).

   p($m|b$) = .9
   p($m|\neg b$) - .7

   However, if James decided to defer the decision to buy the textbook he will know if he has master the material and be able to use this information to decide if he should buy the textbook for the final. Again the probability of him passing the final is as follows.

   ($p|m,b$) = .95
   p($p|m,\neg b$) = .8
   p($p|\neg m,b$) = .6
   p($p|\neg m,\neg b$) = .3

   What is the expected reward of deferring the decision to buy the book and what is the optimal policy for this problem?

Question 3 [20]:

Consider the chart below:

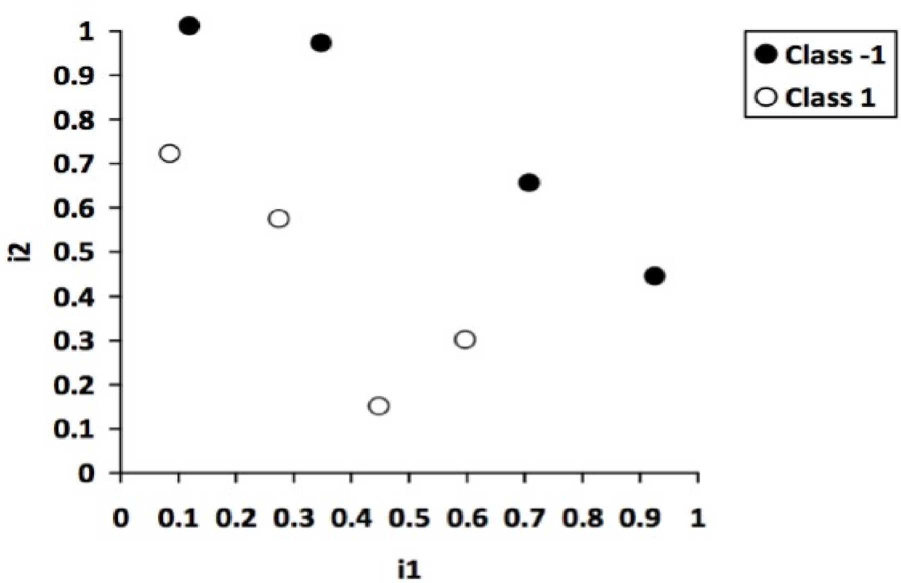

Figure 1: A set of two-dimensional input samples from two classes.

a) Notice that for the above graph there is a perfect linear separator for the two input classes. Therefore, a single perceptron should be able to learn this classification task perfectly. Your task is to replicate the learning process, starting with a random perceptron with weights $w_0 = 0.2$, $w_1 = 1$, and $w_2 = -1$, where the weight w_0 corresponds to the constant offset i0 = 1. For the inputs, just estimate their coordinates from the chart. Start by adding the perceptron's initial line of separation to the chart. Compute then how many samples are misclassified? Then, select an arbitrary misclassified sample and describe the computation of the weight update (you can choose $\eta = 1$ or any other value; if you like you can experiment a bit to find a value that leads to efficient learning). Illustrate the perceptron's new line of division in the same chart or a different one, and give the number of misclassified samples. Repeat this process four more times so that you have a total of six lines (or fewer if your perceptron achieves perfect classification earlier). Have in mind that the classification may not be perfect after these four steps.

b) If your perceptron did not reach perfect classification, determine a set of weights that would achieve perfect classification, and draw the separating line for those weights.

c) Now assume that less information were available about the samples. For instance, consider we only know the value for i1 for each sample, which means that our perceptron has only two weights to classify the input as best as possible, i.e., it has weights w0 and w1, where w0 is once again the weight for the constant offset i0 = 1. Draw a diagram that visualizes this one-dimensional classification task, and determine weights for a perceptron that does the task as best as possible (minimum error, i.e., minimum proportion of misclassified samples). Where does it separate the input space, and what is its error?

Question 4 [15]:

We would like to use regression to fit a sinusoidal curve to a set points.

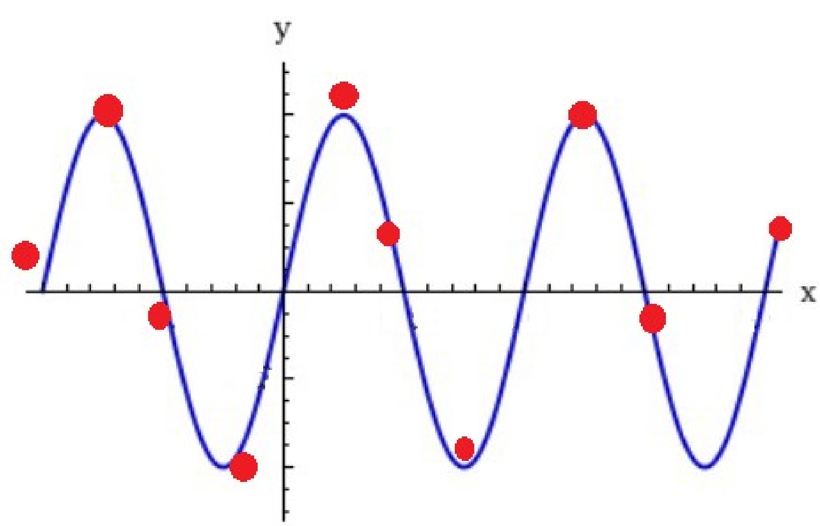

Figure 2: Fitting a 2d sinusoidal function to a set of data points

a) In linear regression we fitted the function $y = w_1x + w_2$ to the data. What would the sinusoidal function look like and what weight parameters would it include. Hint: Think about the frequency, amplitude, horizontal translation and vertical translation of the wave.

b) We define the cost/error of a a fitting to be the square of the difference between the y values in the data points and the y values returned by the model.

$\mathrm{Cost}_W(D) = \Sigma_{(x,y)\in D}(y-f_W(x))_2$

Where D is a set of data points (e.g. the red dots in figure 2) and $f_W(x)$ is the function that is being fit (in this case a sinusoidal function). To perform regression we step the w values in the direction of the gradient of the cost function until the function converges. What is the gradient of the cost function for the sinusoidal function you specified in a.

Hint: $\nabla\mathrm{Cost}_W(D) = \{d\mathrm{Cost}_W(D)/dw_0, d\mathrm{Cost}_W(D)/dw_1, ...\}$

Extra Credit (5 pts): What are some applications where it would make sense to fit a sinusoidal curve.

Question 5 [10]:

In class we discussed how to build neurons that mimic the behavior of AND, OR and NOT gates. These neurons expected an input of 1s and 0s and outputted a 1 or 0. No consider an XOR gate that takes a set of inputs of 1s and 0s and returns 1 if exactly one of those inputs is 1 (and returns 0 otherwise). What would a neuron that models this behavior look like. In particular, what would its input and activation functions be.

Question 6 [15]:

Consider a neural network that consists of n layers of m neurons with each neuron taking as input an edge from each neuron in the previous layer. This network takes a set of inputs $(x_1,x_2,..x_i)$ and returns a set of output values $(y_1,y_2,..y_j)$.

a. What is the big O time complexity of querying this network (i.e. for an input $(x_1,x_2,..x_i)$ what is the time required to compute $(y_1,y_2,..y_j)$. You may assume that both i and j are less then m.

b. What is the big O space complexity of this network?

Hint: The complexity of this network scales with the number of edges.