Assignment 5: Naive Bayes and Perceptrons

CS4811 — Artificial Intelligence

Due December 13th, 11:59pm

1. (15 points) *Perceptrons and Neural Networks*

   (a) (4 points) Consider a perceptron with two inputs, x_1 and x_2. The perceptron uses a transfer function with a threshold value of 0.5 (any value ≥ 0.5 outputs 1, a value < 0.5 outputs 0). The input and weight vectors are formatted as $\vec{x} = \langle x_1, x_2 \rangle$ and $\vec{w} = \langle w_1, w_2 \rangle$. *Note that there is no intercept term for the weight vector.*

   Assume you have the following data set:

Sample	x_1	x_2	y
1	0	0	1
2	0	1	0
3	1	0	0
4	1	1	0

   Assume the weights are initialized to $\vec{w} = \langle 0.7, 0.3 \rangle$. Apply each data sample to the perceptron in turn (sample 1, sample 2, etc.). For each mis-classified sample, use the perceptron training rule, with a learning rate of 0.2 to update the weights.

   Please provide a table, like below, showing how the weights change as each sample is applied. Include your work (calculations) to show how the new weight values were reached.

   Solution:

		w_1	w_2
	Initial Weights	0.7	0.3
(1)	Weights after observing $\vec{x} = \langle 0, 0 \rangle$	0.70	0.30
(2)	Weights after observing $\vec{x} = \langle 0, 1 \rangle$	0.70	0.30
(3)	Weights after observing $\vec{x} = \langle 1, 0 \rangle$	0.50	0.30
(4)	Weights after observing $\vec{x} = \langle 1, 1 \rangle$	0.30	0.10

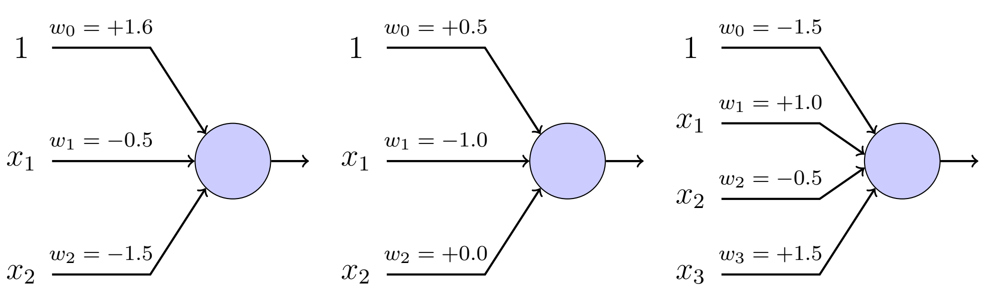
x1	x2	y	
0	0	1	
0	1	0	f(w
1	0	0	f(w
1	1	0	

	x1	x2	w1	w2	f(w*x)	out	y	η	w1'	w2'
1	0	0	0.7	0.3	=SUM((B8*D8)+(C8*E8))	0	1	0.2	=D8+(I8*((H8-G8)*B8))	=E8+(I8*((J8-H8)*C8))
2	0	1	=J8	=K8	=SUM((B9*D9)+(C9*E9))	0	0	0.2	=D9+(I9*((H9-G9)*B9))	=E9+(I9*((H9-G9)*C9))
3	1	0	=J9	=K9	=SUM((B10*D10)+(C10*E10))	1	0	0.2	=D10+(I10*((H10-G10)*B10))	=E10+(I10*((H10-G10)*C10))
4	1	1	=J10	=K10	=SUM((B11*D11)+(C11*E11))	1	0	0.2	=D11+(I11*((H11-G11)*B11))	=E11+(I11*((H11-G11)*C11))

(b) (1 point) If the application of the perceptron training rule continues in a above, would the next cycle through the data have any misclassified samples?

Solution: It should not have any misclassified as all samples have been processed and all posible binary combinations of x1,x2 have been bracketed.

(c) (3 points) Which Boolean functions do the following perceptron units (i), (ii), and (iii) represent?



(i) (ii) (iii)

Assume all inputs take values $\{0, 1\}$ (false, true) and the perceptrons the a transfer function with a threshold value of 0 ($x \geq 0$ outputs 1, $x < 0$ outputs 0).

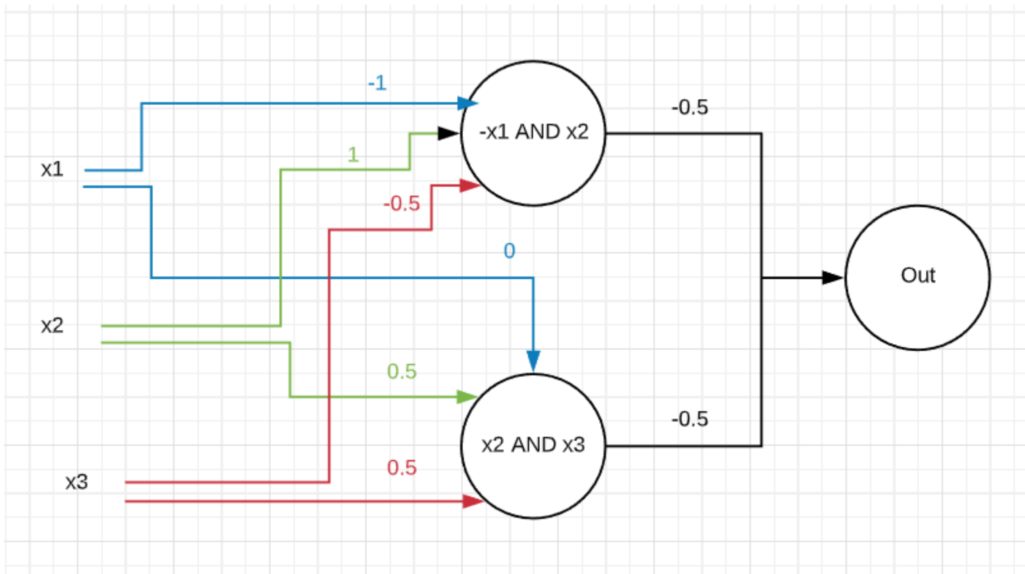
Solution: (i) NOT x1 or x2 (ii) NOT x1 (iii) x1 AND x3

(d) (5 points) Construct a 2-layer network (1 hidden layer and 1 output layer) which recognizes the Boolean expression

$$(\text{not } x_1 \text{ and } x_2) \text{ nor } (x_2 \text{ and } x_3)$$

using binary input values $\{0, 1\}$. Label each edge with its weight. You may include a bias ($x_0 = 1$) if you wish. You are free to use as many perceptrons in the hidden layer as you wish, but 2 hidden perceptrons are sufficient.

Solution:



(e) (2 points) Can the Boolean expression

$$(x_2 \text{ nand } x_3) \text{ or } (x_1 \text{ and not } x_2) \text{ or } (\text{not } x_1 \text{ and } x_2)$$

be learned with a single perceptron? If so, construct the perceptron, labeling each edge with its weight. Otherwise, explain why not.

Solution: No, this is not possible with a single perceptron. This requires at least 2 perceptrons because it calls for both 1 and 0 for x1 out in a single node.

2. (10 points) *Naive Bayes*

Consider the following data set of three Boolean variables: $Weather$, $Roads$, and $Temperature$, and a label $Transport$. This data set describes instances in which your friend drove or walked to campus depending on the conditions outside.

$Weather$	$Roads$	$Temperature$	$Transport$
$Clear$	$Safe$	$Cold$	$Drive$
$Clear$	$Slippery$	$Cold$	$Walk$
$Clear$	$Slippery$	$Frigid$	$Drive$
$Snowing$	$Safe$	$Cold$	$Walk$
$Snowing$	$Slippery$	$Frigid$	$Drive$

(a) (6 points) Calculate the conditional probability distributions for each variable (e.g. $Weather$, $Roads$, $Temperature$) given the label $Transport$:

1. $P(Weather \mid Transport)$
2. $P(Roads \mid Transport)$

3. $P(Temperature \mid Transport)$

Solution:
1. $P(Weather \mid Transport)$
$P(Clear \mid Walk) = 1/2 = 0.5$
$P(Snowing \mid Walk)= 1/2 = 0.5$
$P(Clear \mid Drive) = 2/3 = .666$
$P(Snowing \mid Drive) = 1/3 = .333$

2. $P(Roads \mid Transport)$
$P(Safe \mid Walk) = 1/2 = 0.5$
$P(Slippery \mid Walk) = 1/2 = 0.5$
$P(Safe \mid Drive) = 1/3 = .333$
$P(Slippery \mid Drive) = 2/3 = .666$

3. $P(Temp \mid Transport)$
$P(Cold \mid Walk) = 2/2 = 1.0$
$P(Frigid \mid Walk) = 0/2 = 0$
$P(Cold \mid Drive) = 1/3 = .333$
$P(Frigid \mid Drive) = 2/3 = .666$

(b) (4 points) Suppose today that it is snowing ($Weather = Snowing$), the roads are ($Roads = Safe$), and the temperature is frigid ($Temperature = Frigid$). What are the probabilities associated with the two labels ($Drive$ and $Walk$)? Which classification is preferred? (As the probabilities are simple, you are required to carry out the calculations by hand and show your work)

Solution:
$P(Drive \mid Snowing, Safe, Frigid)$ = 1/3 x 1/3 X 2/3 = 2/27 = 0.0741
$P(Walk \mid Snowing, Safe, Frigid)$ = 1/2 * 1/2 * 0/2 = 0

Drive is the preferred classification. This is due to $P(Frigid \mid Drive) = 2/2 = 1.0$ and $P(Frigid \mid Walk) = 0/2 = 0$.