CS 188
Spring 2019

Final Review ML

Q1. Naive Bayes: Pacman or Ghost?

You are standing by an exit as either Pacmen or ghosts come out of it. Every time someone comes out, you get two observations: a visual one and an auditory one, denoted by the random variables X_v and X_a, respectively. The visual observation informs you that the individual is either a Pacman ($X_v = 1$) or a ghost ($X_v = 0$). The auditory observation X_a is defined analogously. Your observations are a noisy measurement of the individual's true type, which is denoted by Y. After the indiviual comes out, you find out what they really are: either a Pacman ($Y = 1$) or a ghost ($Y = 0$). You have logged your observations and the true types of the first 20 individuals:

individual i	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
first observation $X_v^{(i)}$	0	0	1	0	1	0	0	1	1	1	0	1	1	0	1	1	1	0	0	0
second observation $X_a^{(i)}$	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0
individual's type $Y^{(i)}$	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0	0

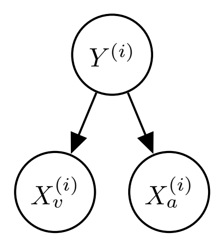
The superscript (i) denotes that the datum is the ith one. Now, the individual with $i = 20$ comes out, and you want to predict the individual's type $Y^{(20)}$ given that you observed $X_v^{(20)} = 1$ and $X_a^{(20)} = 1$.

(a) Assume that the types are independent, and that the observations are independent conditioned on the type. You can model this using naïve Bayes, with $X_v^{(i)}$ and $X_a^{(i)}$ as the features and $Y^{(i)}$ as the labels. Assume the probability distributions take on the following form:

$$P(X_v^{(i)} = x_v | Y^{(i)} = y) = \begin{cases} p_v & \text{if } x_v = y \\ 1 - p_v & \text{if } x_v \neq y \end{cases}$$

$$P(X_a^{(i)} = x_a | Y^{(i)} = y) = \begin{cases} p_a & \text{if } x_a = y \\ 1 - p_a & \text{if } x_a \neq y \end{cases}$$

$$P(Y^{(i)} = 1) = q$$



for $p_v, p_a, q \in [0, 1]$ and $i \in \mathbb{N}$.

(i) What's the maximum likelihood estimate of p_v, p_a and q?

$p_v =$ ______________________ $p_a =$ ______________________ $q =$ ______________________

(ii) What is the probability that the next individual is Pacman given your observations? Express your answer in terms of the parameters p_v, p_a and q (you might not need all of them).

$P(Y^{(20)} = 1 | X_v^{(20)} = 1, X_a^{(20)} = 1) =$ ______________________

Now, assume that you are given additional information: you are told that the individuals are actually coming out of a bus that just arrived, and each bus carries *exactly* 9 individuals. Unlike before, the types of every 9 consecutive individuals are *conditionally* independent given the bus type, which is denoted by Z. Only after all of the 9 individuals have walked out, you find out the bus type: one that carries mostly Pacmans ($Z = 1$) or one that carries mostly ghosts ($Z = 0$). Thus, you only know the bus type in which the first 18 individuals came in:

individual i	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
first observation $X_v^{(i)}$	0	0	1	0	1	0	0	1	1	1	0	1	1	0	1	1	1	0	0	0
second observation $X_a^{(i)}$	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0
individual's type $Y^{(i)}$	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0	0

bus j	0	1	
bus type $Z^{(j)}$	0	1	

(b) You can model this using a variant of naïve bayes, where now 9 consecutive labels $Y^{(i)}, \ldots, Y^{(i+8)}$ are *conditionally* independent given the bus type $Z^{(j)}$, for bus j and individual $i = 9j$. Assume the probability distributions take on the following form:

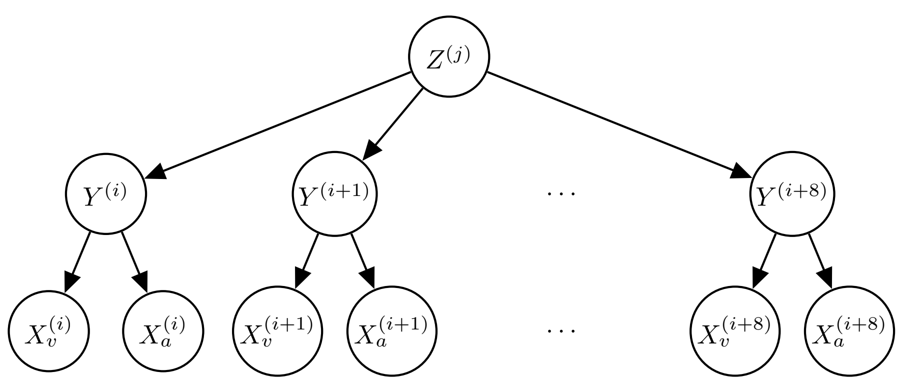
$$P(X_v^{(i)} = x_v | Y^{(i)} = y) = \begin{cases} p_v & \text{if } x_v = y \\ 1 - p_v & \text{if } x_v \neq y \end{cases}$$

$$P(X_a^{(i)} = x_a | Y^{(i)} = y) = \begin{cases} p_a & \text{if } x_a = y \\ 1 - p_a & \text{if } x_a \neq y \end{cases}$$

$$P(Y^{(i)} = 1 | Z^{(j)} = z) = \begin{cases} q_0 & \text{if } z = 0 \\ q_1 & \text{if } z = 1 \end{cases}$$

$$P(Z^{(j)} = 1) = r$$

for $p, q_0, q_1, r \in [0, 1]$ and $i, j \in \mathbb{N}$.



(i) What's the maximum likelihood estimate of q_0, q_1 and r?

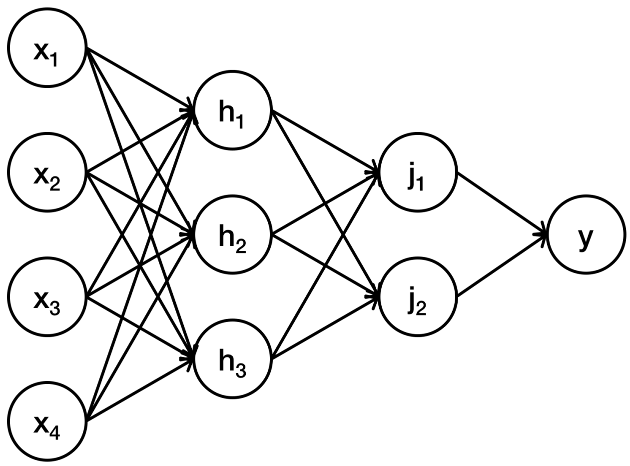
$q_0 =$ ______________________ $q_1 =$ ______________________ $r =$ ______________________

(ii) Compute the following joint probability. Simplify your answer as much as possible and express it in terms of the parameters p_v, p_a, q_0, q_1 and r (you might not need all of them).

$P(Y^{(20)} = 1, X_v^{(20)} = 1, X_a^{(20)} = 1, Y^{(19)} = 1, Y^{(18)} = 1) =$ ______________________

Q2. Neural Network Data Sufficiency

The next few problems use the below neural network as a reference. Neurons h_{1-3} and j_{1-2} all use ReLU activation functions. Neuron y uses the identity activation function: $f(x) = x$. In the questions below, let $w_{a,b}$ denote the weight that connects neurons a and b. Also, let o_a denote the value that neuron a outputs to its next layer.



Given this network, in the following few problems, you have to decide whether the data given are sufficient for answering the question.

(a) Given the above neural network, what is the value of o_y?

Data item 1: the values of all weights in the network and the values o_{h_1}, o_{h_2}, o_{h_3}
Data item 2: the values of all weights in the network and the values o_{j_1}, o_{j_2}

- ◯ Data item (1) alone is sufficient, but data item (2) alone is not sufficient to answer the question.
- ◯ Data item (2) alone is sufficient, but data item (1) alone is not sufficient to answer the question.
- ◯ Both statements taken together are sufficient, but neither data item alone is sufficient.
- ◯ Each data item alone is sufficient to answer the question.
- ◯ Statements (1) and (2) together are not sufficient, and additional data is needed to answer the question.

(b) Given the above neural network, what is the value of o_{h_1}?

Data item 1: the neuron input values, i.e., o_{x_1} through o_{x_4}
Data item 2: the values o_{j_1}, o_{j_2}

- ◯ Data item (1) alone is sufficient, but data item (2) alone is not sufficient to answer the question.
- ◯ Data item (2) alone is sufficient, but data item (1) alone is not sufficient to answer the question.
- ◯ Both statements taken together are sufficient, but neither data item alone is sufficient.
- ◯ Each data item alone is sufficient to answer the question.
- ◯ Statements (1) and (2) together are not sufficient, and additional data is needed to answer the question.

(c) Given the above neural network, what is the value of o_{j_1}?

Data item 1: the values of all weights connecting neurons h_1, h_2, h_3 to j_1, j_2
Data item 2: the values o_{h_1}, o_{h_2}, o_{h_3}

- ◯ Data item (1) alone is sufficient, but data item (2) alone is not sufficient to answer the question.
- ◯ Data item (2) alone is sufficient, but data item (1) alone is not sufficient to answer the question.
- ◯ Both statements taken together are sufficient, but neither data item alone is sufficient.

◯ Each data item alone is sufficient to answer the question.
◯ Statements (1) and (2) together are not sufficient, and additional data is needed to answer the question.

(d) Given the above neural network, what is the value of $\partial o_y / \partial w_{j_2,y}$?

Data item 1: the value of o_{j_2}
Data item 2: all weights in the network and the neuron input values, i.e., o_{x_1} through o_{x_4}

◯ Data item (1) alone is sufficient, but data item (2) alone is not sufficient to answer the question.
◯ Data item (2) alone is sufficient, but data item (1) alone is not sufficient to answer the question.
◯ Both statements taken together are sufficient, but neither data item alone is sufficient.
◯ Each data item alone is sufficient to answer the question.
◯ Statements (1) and (2) together are not sufficient, and additional data is needed to answer the question.

(e) Given the above neural network, what is the value of $\partial o_y / \partial w_{h_2,j_2}$?

Data item 1: the value of $w_{j_2,y}$
Data item 2: the value of $\partial o_{j_2} / \partial w_{h_2,j_2}$

◯ Data item (1) alone is sufficient, but data item (2) alone is not sufficient to answer the question.
◯ Data item (2) alone is sufficient, but data item (1) alone is not sufficient to answer the question.
◯ Both statements taken together are sufficient, but neither data item alone is sufficient.
◯ Each data item alone is sufficient to answer the question.
◯ Statements (1) and (2) together are not sufficient, and additional data is needed to answer the question.

(f) Given the above neural network, what is the value of $\partial o_y / \partial w_{x_1,h_3}$?

Data item 1: the value of all weights in the network and the neuron input values, i.e., o_{x_1} through o_{x_4}
Data item 2: the value of w_{x_1,h_3}

◯ Data item (1) alone is sufficient, but data item (2) alone is not sufficient to answer the question.
◯ Data item (2) alone is sufficient, but data item (1) alone is not sufficient to answer the question.
◯ Both statements taken together are sufficient, but neither data item alone is sufficient.
◯ Each data item alone is sufficient to answer the question.
◯ Statements (1) and (2) together are not sufficient, and additional data is needed to answer the question.