VE 492 Homework9

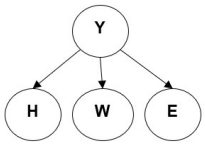
Due: 23:59, July. 22th

Q1. Naive Bayes

Your friend claims that he can write an effective Naive Bayes spam detector with only three features: the hour of the day that the email was received ($H \in \{1,2,...,24\}$), whether it contains the word 'viagra' ($W \in \{yes,no\}$), and whether the email address of the sender is Known in his address book, Seen before in his inbox, or Unseen before ($E \in \{K,S,U\}$).

(a) Flesh out the following information about this Bayes net:

Graph structure:



Here $Y \in$ *{spam, ham}*

Parameters: *P(Y=spam), P(H′ | Y), P(W=yes | Y), P(E′ | Y), where H′ ∈ {1,2,...,23}, E′ ∈ {K,S}*

Size of the set of parameters: 1 + 23·2 + 2 + 2·2 = 53

Suppose now that you labeled three of the emails in your mailbox to test this idea:

spam or ham?	H	W	E
spam	3	yes	S
ham	14	no	K
ham	15	no	K

(b) Use the three instances to estimate the maximum likelihood parameters.
P(Y=spam) = 1/3, P(H=3 | Y=spam) = 1, P(H=14 | Y=ham) = 1/2, P(H=15 | Y=ham) = 1/2, P(W=yes | Y=spam) = 1, P(E=S | Y=spam) = 1, P(E=K | Y=ham) = 1,
All the other parameters defined in (a) are 0

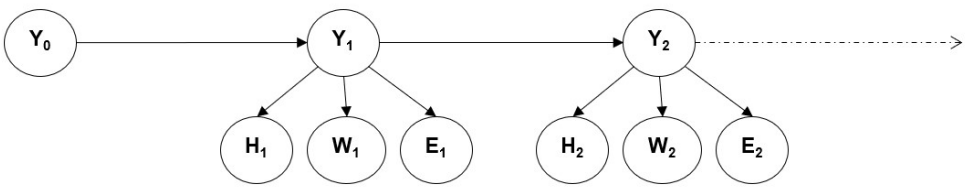
(c) Using the maximum likelihood parameters, find the predicted class of a new datapoint with $H = 3, W = no, E = U$. No predicted class

(d) Now use the three to estimate the parameters using Laplace smoothing and $k = 2$. Do not forget to smooth both the class prior parameters and the feature values parameters.
P(Y=spam) = 3/7, P(H=3 | Y=spam) = 3/49, all the other P(H' | Y = spam) = 2/49, P(H=14 | Y=ham) = 3/50, P(H=15 | Y=ham) = 3/50, all the other P(H' | Y = ham) = 1/25, P(W=yes | Y=spam) = 3/5, P(W=yes | Y=ham) = 1/3, P(E=S | Y=spam) = 3/7, P(E=K | Y=spam) = 2/7, P(E=S | Y=ham) = 1/4, P(E=K | Y=ham) = 1/2

(e) Using the parameters obtained with Laplace smoothing, find the predicted class of a new datapoint with $H = 3, W = no, E = U$. ham

(f) You observe that you tend to receive spam emails in batches. In particular, if you receive one spam message, the next message is more likely to be a spam message as well. Explain a new graphical model which most naturally captures this phenomena.

Graph structure:



Parameters:

$P(Y_0=spam)$, $P(Y_t=spam \mid Y_{t-1})$, $P(H'_t \mid Y_t)$, $P(W_t=yes \mid Y_t)$, $P(E'_t \mid Y_t)$, where $H'_t \in \{1,2,...,23\}$, $E'_t \in \{K,S\}$

Size of the set of parameters: 53 + 2 = 55

Q2. Perceptron

(a) Suppose you have a binary perceptron in 2D with weight vector $\boldsymbol{w} = r\,[w_1, w_2]^T$. You are given w_1 and w_2, and are given that r > 0, but otherwise not told what r is. *Assume that ties are broken as positive*. Can you determine the perceptron's classifification of a new example x with known feature vector f(x)?

A. Always

B. Sometimes

C. Never

(b) Now you are learning a multi-class perceptron between 4 classes. The weight vectors are currently $[1,0]^T$,$[0,1]^T$,$[-1,0]^T$,$[0,-1]^T$ for the classes A, B, C, and D. The next training example *x* has a **label of A** and feature vector *f*(*x*).

For the following questions, *do not make any assumptions about tie-breaking.* (Do not write down a solution that creates a tie.)

If the answer does not exist, write down **Not possible**

$f(x) = \begin{bmatrix} \quad \\ \quad \end{bmatrix}$ ○ Not possible

(i) Write down a feature vector in which no weight vectors will be updated. $f(x) = [1,0]^T$

(ii) Write down a feature vector in which **only** $\mathbf{w}_A$ will be updated by the perceptron. Not possible

(iii) Write down a feature vector in which **only** $\mathbf{w}_A$ and $\mathbf{w}_B$ will be updated by the perceptron. $f(x) = [0,1]^T$

(iv) Write down a feature vector in which **only** $\mathbf{w}_A$ and $\mathbf{w}_C$ will be updated by the perceptron. $f(x) = [-1,0]^T$

The weight vectors are the same as before, but now there is a bias feature with value of 1 for all *x* and the weight of this bias feature is 0, -2, 1, - 1 for classes A, B, C, and D respectively. As before, the next training example *x* has a **label of A** and a feature vector *f*(*x*). The always "1" bias feature is the first entry in *f*(*x*).

If the answer does not exist, write down **Not possible**

$f(x) = \begin{bmatrix} 1 \\ \quad \end{bmatrix}$ ○ Not possible

(v) Write down a feature vector in which **only** $\mathbf{w}B$ and $\mathbf{w}C$ will be updated by the perceptron. Not possible

(vi) Write down a feature vector in which **only** $\mathbf{w}A$ and $\mathbf{w}C$ will be updated by the perceptron. $f(x) = [1,-1,0]^T$