VE 492 Homework8

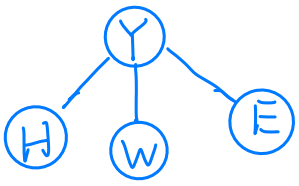
Due: 23:59, July 28

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:



$Y \in \{spam, ham\}$.

Parameters: $P(Y)$, $P(H|Y)$, $P(W|Y)$, $P(E|Y)$.

Size of the set of parameters:

$2 + 24\times2 + 2\times2 + 3\times2 = 60$

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.

(b) $P(Y=spam)=\frac{1}{3}$, $P(Y=ham)=\frac{2}{3}$, $P(H=3|Y=spam)=1$, $P(H=14|Y=ham)=\frac{1}{2}$, $P(H=15|Y=ham)=\frac{1}{2}$,
$P(W=yes|Y=spam)=1$, $P(W=No|Y=ham)=1$, $P(E=S|Y=spam)=1$, $P(E=K|Y=ham)=1$, All others are 0

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

No such 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.

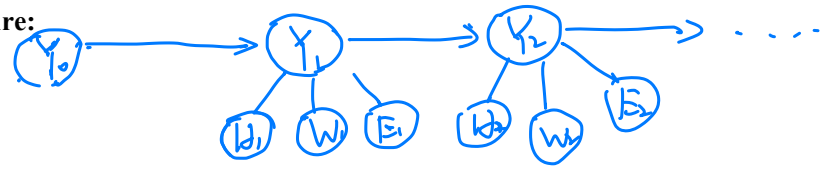
(d) $P(Y=spam)=\frac{3}{7}$, $P(Y=ham)=\frac{4}{7}$, $P(H=3|Y=spam)=\frac{3}{48}$, $P(H'|Y=spam)=\frac{2}{48}$, $H' \in \{1,2,4,\cdots,24\}$
$P(W=yes|Y=spam)=\frac{3}{5}$, $P(W=no|Y=spam)=\frac{2}{5}$, $P(E=S|Y=spam)=\frac{3}{5}$, $P(E=K|Y=spam)=\frac{2}{5}$, $P(E=U|Y=spam)=\frac{2}{7}$
$P(H=14|Y=ham)=\frac{3}{50}$, $P(H=15|ham)=\frac{3}{50}$, $P(H_1|Y=ham)=\frac{1}{25}$, $H_1 \in \{1,2,\cdots,13,16,\cdots,24\}$, $P(W=no|Y=ham)=\frac{2}{3}$,
$P(W=Yes|Y=ham)=\frac{1}{3}$, $P(E=K|Y=ham)=\frac{1}{2}$, $P(E=S|Y=ham)=\frac{1}{6}$, $P(E=U|Y=ham)=\frac{1}{6}$

(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:

$P(Y_0)$, $P(Y_t|Y_{t-1})$, $P(H_t|Y_t)$, $P(W_t|Y_t)$, $P(E_t|Y_t)$. $t \geq 1$.

Size of the set of parameters:

6044 = 64.

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 w**$_A$ will be updated by the perceptron. Not possible.

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

(iv) Write down a feature vector in which **only w**$_A$ and **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 w***B* and **w***C* will be updated by the perceptron. Not possible.

(vi) Write down a feature vector in which **only w***A* and **w***C* will be updated by the perceptron.

f(x) = [-1, 1, 0]^T