

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 ($H \in \{1,2,\dots,24\}$), 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:

Parameters:

Size of the set of parameters:

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

Size of the set of parameters:

Q2. Perceptron

- (a) Suppose you have a binary perceptron in 2D with weight vector $\mathbf{w} = r [w_1, w_2]^T$. 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 classification 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**

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

- (i) Write down a feature vector in which no weight vectors will be updated.
- (ii) Write down a feature vector in which **only** \mathbf{w}_A will be updated by the perceptron.
- (iii) Write down a feature vector in which **only** \mathbf{w}_A and \mathbf{w}_B will be updated by the perceptron.
- (iv) Write down a feature vector in which **only** \mathbf{w}_A and \mathbf{w}_C will be updated by the perceptron.

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**

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

- (v) Write down a feature vector in which **only** \mathbf{w}_B and \mathbf{w}_C will be updated by the perceptron.
- (vi) Write down a feature vector in which **only** \mathbf{w}_A and \mathbf{w}_C will be updated by the perceptron.