CS251 HW 9 | Mon Apr 22, 2019 | Week 12

Name:

Question 1: Naive Bayes

You are debugging a Naive Bayes model for spam email classification and would like to work out a simple test case to make sure your code is working correctly. You want to figure out by pen-and-paper whether your model should classify a test message containing the words `you`, `won`, `lottery` as **Spam** or **Not Spam**.

For your **likelihood**, you are using normalized counts from your training set (not a Gaussian distribution).

For your **prior**, you are using an internet spam word occurrence dataset and discovered that Not Spam emails are 3 times more likely as Spam emails.

Training counts

Assume that your entire training email data is summarized in the following table:

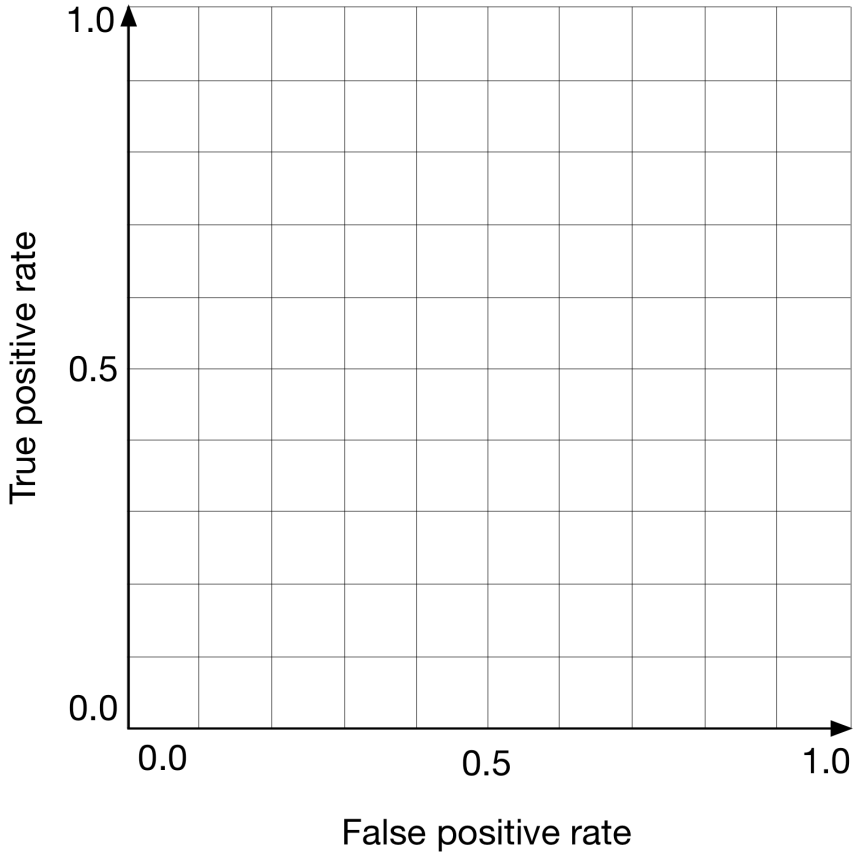
Word	Count when Spam	Count when Not Spam
you	40	10
won	60	40
lottery	100	50

Use Naive Bayes to determine whether the email (with `you`, `won`, `lottery`) is Spam or Not Spam.

Question 2: ROC Curve

A supervised learning model processed a medical database and produced the following confusion matrix:

	Predicted (+)	**Predicted (-)**
Actual (+)	70	30
Actual (-)	80	120



a) Define **true positive rate**.

b) Define **false positive rate**.

c) Define **precision** (positive predictive value).

c) In this example, the true positive rate is _____, the false positive rate is _____, and the precision is _____.

d) In the graph above, place the point you computed from (c).

e) The above confusion matrix was generated with a binary decision threshold of $\gamma = 0.4$ (i.e. the point that you plotted in (d) corresponds to $\gamma = 0.4$).

i. For a different threshold value ($\gamma = 0.2$) estimate the false positive rate if

- Our classifier labels 90% of true positive cases as (+)
- Our classifier labels 120 cases that are actually (-) as (+)

ii. Place the $\gamma = 0.2$ point on the plot.

f) Sketch an ROC curve that includes the $\gamma = 0.4$ and $\gamma = 0.2$ points.

g) Use the "graph paper" to estimate the C-index/AUC value.

h) Is the AUC value "good", "mediocre", or "bad" compared to an all-negative classifier?