CS 4100/5100: Foundations of Artificial Intelligence (Fall 2015)	Robert Platt
Student(s):	Dec. 3, 2015

Naive Bayes and Perceptrons

1 Move prediction

You want to predict if movies will be profitable based on their screenplays. You hire two critics A and B to read a script you have and rate it on a scale of 1 to 5 (so the critic ratings are the features to be used and that we are trying to predict profitability). The critics are not perfect; here are five data points including the critics scores and the performance of the movie:

Movie Name	A	B	*Profit*?
Meet Pac Man	3	2	Yes
Pixels	1	1	No
The Ghostly Adventures	4	6	No
Pac Baby	2	4	Yes
Pac is Back	3	4	Yes

1 First, you would like to examine the linear separability of the data. Plot the data on the 2D plane. Label profitable movies with + and non-profitable movies with and determine if the data are linearly separable.

2 Now you first decide to use a perceptron to classify your data. This problem will use the multi-class formulation even though there are only two classes. Suppose you directly use the scores given above as features, together with a bias feature. That is $f_0 = 1$, f_1 = score given by A and f_2 = score given by B. You want to train the perceptron on the training data in the table below.

Profit	Weights	Weights after 1st update
Yes	[-1, 0, 0]	
NO	[+1, 0, 0]	

- Which is the first training instance at which you update your weights? Why?

- Write the updated weights after the first update.

3 More generally, irrespective of the training data, you want to know if your features are powerful enough to allow you to handle a range of scenarios. Some scenarios are given on the next page. Circle those scenarios for which a perceptron using the features above can indeed perfectly classify the data:

(i) Your reviewers are art critics. Your movie will succeed if and only if each reviewer gives either a score of 2 or a score of 3.

(ii) Your reviewers are awesome: if the total of their scores is more than 8, then the movie will definitely be a success and otherwise it will fail.

(iii) Your reviewers have weird but different tastes. Your movie will succeed if and only if both reviewers agree.

You decide to use a different set of features. Consider the following feature space:

$f_0 = 1$ (The bias feature)

$f_{1A} = 1$ if score given by A is 1, 0 otherwise

$f_{1B} = 1$ if score given by B is 1, 0 otherwise

$f_{2A} = 1$ if score given by A is 2, 0 otherwise

$f_{2B} = 1$ if score given by B is 2, 0 otherwise

...

$f_{5B} = 1$ if score given by B is 5, 0 otherwise

4 Consider again the three scenarios in part 2. Using a perceptron with the new features, which of the three scenarios can be perfectly classified? Circle your answer(s) below:

(i) Your reviewers are art critics. Your movie will succeed if and only if each reviewer gives either a score of 2 or a score of 3.

(ii) Your reviewers have weird but different tastes. Your movie will succeed if and only if both reviewers agree.

(iii) Your reviewers are awesome: if the total of their scores is more than 8, then the movie will definitely be a success, and otherwise it will fail.

5 You have just heard of naive Bayes and you want to use a naive Bayes classifier. You use the scores given by the reviewers as the features of the naive Bayes classifier, i.e., the random variables in your naive Bayes model are A and B, each with a domain of $\{1, 2, \cdots, 5\}$, and Profit with a domain of Yes and No. Draw the Bayes net corresponding to the naive Bayes model.

6 List the types of the conditional probability tables you need to estimate along with their sizes (e.g., $P(X|Y)$ has 24 entries)

7 Your friend is taking the CS5100 class at NEU. He claims that the naive Bayes classifier you just built is actually a linear classifier in the feature space used for question 3. In other words, the decision boundary of the naive Bayes classifier is a hyper-plane in this feature space. For the positive class, what is the weight of the feature f_{3B} in terms of the parameters of the naive Bayes model? You can answer in symbols, but be precise. (Hint: Consider the log of the probability.)