CS60075 Quiz 29 Sep 2020

1. Consider the following pieces of text and their classes. You will use the Naïve Bayes model and the assumptions of that model to answer the following questions.

Document	Label
I love the movie.	POS
Boring.	NEG
I hated the movie.	NEG
Really cool movie.	POS

1A. What is the unsmoothed maximum likelihood estimate of **P(POS)** for this data?

[Correct answer : 0.5] (Two POS out of Four intances)

1B. What is the unsmoothed maximum likelihood estimate of P(movie|POS) for this data?

[Correct answer 1/3]

1C. Suppose we are given an unseen input sentence **"the movie"**
What is the joint probability **P(POS, "the movie")**?
~~1/3, 1/8, 1/12, 1/18,~~

Correct answer: 1/49

$$P(POS, \text{“}the\ movie\text{”}) = P(POS) * P(\text{the movie}|POS)$$
$$= P(POS) * P(\text{"the"}|POS) * P(\text{"movie"}|POS) = 1/2 * 1/7 * 2/7 = 1/49$$

2. Consider the following tagged corpus. Suppose you use unsmoothed MLE to estimate A and B.
"You read books. He books a table. I table a bill. Bill the food. Bill books the books"
Consider the following parts of speech along with start and end of sentence <bos> and <eos>
P: Pronoun
N: Noun
V: Verb
D: Determiner or Article
Suppose you have tagged the corpus manually with the most likely tags. Use this to answer the following questions:

3A. Based on this data, find the value of the following:

What is $P(q_{t+1} = N|q_t = V)$ (transition of Verb to Noun) ?

Ans: 1/5 or 0.2

3B. What is the probability of observing "bill" or "Bill" (ignore case) if the tag is Verb, i.e., $P(O_t = \text{bill}|q_t = V)$?

Ans. **1/5 or 0.2**

3C. What is the probability of observing "books" if the tag is Verb, i.e., $P(O_t = \text{books}|q_t = N)$?

1/3 or 0.33

3D. What is the probability of the tag sequence``V N" for the phrase "Bill books." ?

1/375

$$P(V| < s >) * P(Bill|V) * P(N|V) * P(Books|N) = \frac{1}{5} * \frac{1}{5} * \frac{1}{5} * \frac{1}{3}$$

3. True False Question.

(a) [1 point] We can get multiple local optimum solutions if we solve a linear regression problem by minimizing the sum of squared errors using gradient descent.
False
(b) The logarithmic loss function is suitable for classification problems.
True
(c) The Mean Squared Error loss function is a good loss function for 2-class classification.
False
(d) Backpropagation algorithm is based on gradient descent along the error surface.
True

4. Consider the following three datasets:

A

X1	X2	Y
1	1	+
4	2	–
4	5	–
5	5	+

B

X1	X2	Y
1	1	+
5	5	–
4	5	–
5	5	+

C

X1	X2	Y
1	1	+
4	2	–
4	5	–
5	5	+

Which data sets are linearly separable?

(i) None