[Quiz] Naive Bayes

- Due 30 Mar at 23:59
- Points 11
- Questions 11
- Time limit None
- Allowed attempts 2

This quiz is no longer available as the course has been concluded.

Attempt history

	Attempt	Time	Score
LATEST	Attempt 1	16 minutes	11 out of 11

Answers will be shown after your last attempt

Score for this attempt: 11 out of 11
Submitted 30 Mar at 17:29
This attempt took 16 minutes.

Question 1
1 / 1 pts

When two random events X and Y are independent, we can write the joint probability $P(X, Y)$ as which of the following?

- [x] $P(Y)P(X)$
- [ ] $P(X \mid Y)$
- [ ] $P(Y \mid X)P(X)$
- [ ] $P(Y \mid X)$

Question 2
1 / 1 pts

Recall that *maximum likelihood (ML)* is one way to compute the label for a given class. We do so by selecting the label for the class with the highest likelihood. Mathematically this is written as follows:

$$\begin{aligned}\hat{y}_{ML} &= \arg\max_{k \in K} P(\mathbf{x} \mid y_k) \\ &= \arg\max_{k \in K} \prod_{i=0}^{n} P(x_i | y_k)\end{aligned}$$

Select ALL of the following that are true about ML.

- [x] ML and MAP perform relatively the same when given enough data. Thus, it can be more efficient to simply use ML as we don't need to compute the prior.
- [x] ML ignores the prior and evidence producing a simplified equation.
- [ ] ML indclues the prior which allows us to inject our prior knowledge into the equation.
- [ ] ML includes the evidence as a normalizing term.

Question 3
1 / 1 pts

Recall that *maximum a posteriori (MAP)* is one way to compute the label for a given class. We do so by taking selecting the label for the class with the highest joint likelihood. Mathematically this is written as follows:

$$\begin{aligned}\hat{y}_{MAP} &= \arg\max_{k \in K} P(\mathbf{x} \mid y_k)P(y_k) \\ &= \arg\max_{k \in K} P(y_k) \prod_{i=0}^{n} P(x_i | y_k)\end{aligned}$$

Select ALL of the following that are true about MAP.

- [ ] MAP includes the evidence as a normalizing term.
- [x] MAP tends to perform better given when there is little training data
- [x] MAP and ML are roughly equivalent if the priors for each class are equal (i.e., a uniform distribution across priors).
- [x] MAP indclues the prior which allows us to inject our prior knowledge into the equation.

Question 4
1 / 1 pts

Fill in the blank.

In order to turn the products in the MAP and ML equations into sums to prevent underflow and make the equations more efficient to compute, we apply the _____ to the equations.

- [ ] argmax
- [ ] exponential
- [x] log
- [ ] normalizing term

Question 5
1 / 1 pts

Fill in the blank.

Naive Bayes makes the naive assumptions of _____ which allows computing the likelihood $P(\mathbf{x} \mid y_k)$ to be simplified to the following:

$$\begin{aligned}P(\mathbf{x}|y_k) &= P(x_1|y_k)P(x_2|y_k)\ldots P(x_n|y_k) \\ &= \prod_{i=0}^{n} P(x_i|y_k)\end{aligned}$$

- [ ] dependence
- [ ] independence
- [ ] product/chain rule
- [x] conditional independence

Question 6
1 / 1 pts

True or false.

When maximizing MAP, we need to compute **ONLY** the likelihood $P(\mathbf{x} \mid y_k)$.

- [ ] True
- [x] False

Question 7
1 / 1 pts

Select all that apply.

To derive **Bayes rule** we use which of the following probability theory terms and rules?

- [x] joint probability
- [x] product/chain rule
- [ ] conditional independnece
- [x] conditional probability
- [ ] independence

Question 8
1 / 1 pts

True or false.

Typically we drop the evidence $P(X)$ when computing either MAP or ML. However, we add the evidence $P(X)$ back when we want to compute the probability that a data sample belongs to each class. We do so by dividing the joint likelihood $P(X \mid Y)P(Y)$ (or just the likelihood $P(X \mid Y)$ when using ML) by the evidence $P(X)$.

- [x] True

- False

Question 9

1 / 1 pts

Typically, Naive Bayes (also called Categorical Naive Bayes) computes the likelihood $P(\mathbf{x} \mid y_k)$ for each feature in a class by counting the frequency of the unique values (i.e., categories) and dividing by the total number of values the said feature has.

Which of the following is the issue with this method of computing the likelihood?

- It doesn't work for discrete feature values
- It doesn't work for continuous feature values
- There is no issue.
- It doesn't work for either discrete or continuous features.

Question 10

1 / 1 pts

Choose the BEST answer.

When using Gaussian Naive Bayes, we have to compute the mean and std for ______.

- all the classes
- for all the continuous features, for each class
- all the continuous features

Question 11

1 / 1 pts

If our data has 4 features and 2 classes, how many mean and standard deviation (std) parameters will need to be computed when using Gaussian Naive Bayes?

- 6
- 8
- 2
- 4

Quiz score: 11 out of 11