Quiz Submissions - Quiz 4

Mike Gao (username: zenghao.gao@mail.mcgill.ca)

Attempt 1

Written: Oct 5, 2020 10:50 AM - Oct 5, 2020 3:00 PM

Submission View

Released: Oct 4, 2020 7:45 PM

Question 1 **1 / 1 point**

Suppose you have a binary classification problem with “m” continuous (real-valued) features. You plan to use Guassian Naive Bayes on the dataset. How many parameters do you have in your model? Select one answer which is true.

- 4m
- ✓ 4m+1
- 2m
- 2m+1

Hide Feedback

To learn the class distribution P(c), since the output is binary, estimating P(c) requires a single parameter (i.e., we need to estimate P(c=1) which gives the estimated prior class probability that c is equal to 1).

We also need to estimate P(x|y=1) and P(x|y=0) for each feature. In this case, we estimate P(x|y=k) as a Gaussian and we need to learn the mean and variance for each feature, which requires 2 parameters to learn. Thus in total we have (2 classes)*(2 parameters to learn the conditional distribution of each feature for each class)*(m features)+(1 parameter to learn the marginal likelihood of the target class) = 4m+1 parameters.

Question 2 **1 / 1 point**

Naive Bayes is a generative classifier based on Bayes rule with a conditional independence assumption on the prior.

- True
- ✓ False

Question 3 **1 / 1 point**

Select **all** the choices that are True.

- ✓ **Naive bayes is a generative classifier and learns the joint distribution of target and features: p(x,y)**
- ✓ **For any two variables x and y having joint distribution p(x, y), we always have H[x, y] ≥ H[x] + H[y] where H is entropy function.**

Question 4 **1 / 1 point**

Suppose we are given data comprising points of several different classes. Each class has a different probability distribution from which the sample points are drawn. We do not have the class labels. We use "k-means" clustering to try to guess the classes. Which of the following circumstances would undermine its effectiveness? Select **all** choices which are True.

- ✔ ☑ Each class has the same mean.
- ✔ ☐ The variance of each distribution is small in all directions.
- ✔ ☑ You choose k = n, the number of sample points.

Question 5 **1 / 1 point**

We want to calculate P(Y | X1, X2) and we don't have any additional conditional independence information. Which of the following sets of distributions are sufficient for calculation? Select **all** the correct answers.

- ✔ ☐ P(X1,X2), P(Y), P(X1|Y), P(X2|Y)
- ✔ ☑ P(X1,X2), P(Y), P(X1,X2|Y)
- ✔ ☐ P(Y), P(X1|Y), P(X2|Y)

▼ Hide Feedback

Using Bayes Rule: P(Y | X1, X2) = P(X1, X2 | Y) P(Y) / P(X1, X2).

Question 6 **1 / 1 point**

We want to calculate P(Y| X1, X2) and now we have additional information that P(X1| Y, X2) = P(X1| Y) for all values of X1, X2 , Y. Which of the following sets of distributions are sufficient for calculation? Select **all** the correct answers.

- ✔ ☑ P(X1,X2), P(Y), P(X1|Y), P(X2|Y)
- ✔ ☑ P(X1,X2), P(Y), P(X1,X2|Y)
- ✔ ☑ P(Y), P(X1|Y), P(X2|Y)

▼ Hide Feedback

With additional information we know that P(X1,X2 | Y) = P(X1 | Y) P(X2 | Y).

If we don't use this additional independence relation.

∑_{y \in Y} P(X1, X2 ,Y)= ∑_{y \in Y} P(X1, X2 | Y) P(Y) = P(X1, X2)

Question 7 **1 / 1 point**

K-mean clustering: Consider this training data set where Examples are A-E, and they have a single real-valued attribute X.

Example	A	B	C	D	E
Attribute X	0.1	0.6	0.8	2.0	3.0

Apply k-Means Clustering to this data set for k=2, i.e., you will produce two data clusters.

You have randomly chosen two data points with which to initialize your two clusters. Randomly, you chose example A to initialize cluster #1 and example B to initialize cluster #2. Now put examples C,D,E into different clusters based on these initial cluster centers. Recompute the cluster centers based on the assignment of data points into 2 clusters. Now redo the cluster assignment for examples A-E based on the new cluster centers. Now select all the examples that would belong to Cluster #1.

- [x] A
- [x] B
- [x] C
- [ ] D
- [ ] E

Hide Feedback

Step 1: Cluster #1 : A, Cluster #2: B,C,D,E

Step 2: Mean Cluster #1 :0.1, Mean of Cluster #2: (0.6+0.8+2.0+3.0)/4 = 1.6

Step 3:Cluster #1 :A,B,C Cluster #2: D,E

Attempt Score: 7 / 7 - 100 %

Overall Grade (highest attempt): 7 / 7 - 100 %

Done