Naive Bayes Algorithms

We believe that you have learned both theoritical and practical knowledge on Naive Bayes classification algorithm through your assignment.

So let's test your knowledge here. This will help you to be prepared for interviews too!

Best with Quest

1. Why is naive Bayes so 'naive' ?

```
#Comment

naive Bayes is so 'naive' because it assumes that all of the features in a data set are equally important and independent. As we know, these assumption are rarely true in real world scenario.
```

2. Explain prior probability, likelihood and marginal likelihood in context of naiveBayes algorithm?

```
#comment

Prior probability is nothing but, the proportion of dependent (binary) variable in the data set. It is the closest guess you can make about a class, without any further information. For example: In a data set, the dependent variable is binary (1 and 0). The proportion of 1 (spam) is 70% and 0 (not spam) is 30%. Hence, we can estimate that there are 70% chances that any new email would be classified as spam.

Likelihood is the probability of classifying a given observation as 1 in presence of some other variable. For example: The probability that the word 'FREE' is used in previous spam message is likelihood. Marginal likelihood is, the probability that the word 'FREE' is used in any message.
```

3. Explain the difference between Type 1 and Type 2 errors?

```
#Comment

Type 1 error is a false positive error that 'claims' that an incident has occurred when, in fact, nothing has occurred. The best example of a false positive error is a false fire alarm – the alarm starts ringing when there's no fire.  Contrary to this, a Type 2 error is a false negative error that 'claims' nothing has occurred when something has definitely happened.  It would be a Type 2 error to tell a pregnant lady that she isn't carrying a baby.
```

4. What is meant by F1 score?

```
#comment
```

```
In simple terms, the F1 score is a measure of a model's performance – an average of the
Precision and Recall of a model, with results nearing to 1 being the best and those
nearing to 0 being the worst. The F1 score can be used in classification tests that do not
place importance on true negatives.
```

5. Brief on Variants of Naive Bayes Classifier?

```
#comment

There are several types of Naive Bayes classifiers.  Which one you use will depend on the
features you are working with.  The different types are:

Gaussian NB – use when you have continuous feature values.  This classifier assumes each
class is normally distributed.
MultiNomial NB – good for text classification.  This classifier treats each occurrence of
a word as an event.
Bernoulli NB – use when you have multiple features that are assumed to be binary.  This
classifier can be used for text classification but the features must be binary.  For text
classification, the features can be set as a word is in the document or not in the
document.
```

6. What are advantage of Naive Bayes?

```
#comment

1. Can successfully train on small data set
2. Good for text classification, good for multiclass classification
3. Quick and simple calculation since it is naive
```

7. What are disadvantages of Naive Bayes?

```
#comment

1. Can't learn the relationship among the features because assumes feature independence
2. Continous feature data is assumed to be normally distributed
```

8. What does a ROC curve tell you?

```
#comment

AUC – ROC curve is a performance measurement for the classification problems at various
threshold settings. ROC is a probability curve, and AUC represents the degree or measure
of separability. It tells how much model is capable of distinguishing between classes.
Higher the AUC, better the model is at predicting 0s as 0s and 1s as 1s. By analogy,
Higher the AUC, better the model is at distinguishing between patients with the disease
and no disease.

The ROC curve is plotted with TPR against the FPR where TPR is on the y-axis and FPR is on
the x-axis.
```

9. What is CountVectorizer which you used in your assignment?

```
#comment
```

CountVectorizer is a great tool provided by the scikit-learn library in Python. It is used to transform a given text into a vector on the basis of the frequency (count) of each word that occurs in the entire text. This is helpful when we have multiple such texts, and we wish to convert each word in each text into vectors (for using in further text analysis).

10. What is Bayes' Theorem? How is it useful in a machine learning context?

#comment

Bayes' Theorem gives you the posterior probability of an event given what is known as prior knowledge.

Mathematically, it's expressed as the true positive rate of a condition sample divided by the sum of the false positive rate of the population and the true positive rate of a condition. Say you had a 60% chance of actually having the flu after a flu test, but out of people who had the flu, the test will be false 50% of the time, and the overall population only has a 5% chance of having the flu. Would you actually have a 60% chance of having the flu after having a positive test?

Bayes' Theorem says no. It says that you have a (.6 * 0.05) (True Positive Rate of a Condition Sample) / (.6*0.05)(True Positive Rate of a Condition Sample) + (.5*0.95) (False Positive Rate of a Population) = 0.0594 or 5.94% chance of getting a flu.

Bayes' Theorem is the basis behind a branch of machine learning that most notably includes the Naive Bayes classifier. That's something important to consider when you're faced with machine learning interview questions.