Week 1

Theory：

1. How is probability theory and machine learning theory related?

There are some direct equivalents between the probability theory and machine learning. A random variable from probability maps to a feature value in machine learning. Random variables in probability theory and features in machine learning are categorical or continuous.

1. What is the difference between prior probability and conditional probability? Explain your answer using two random variables X and Y.

Prior probability is the number of times we observed divided by the total number of observations. But conditional probability is to calculate the probability of an event given we already know outcome of some other event.

For example, the observations are

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Time | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| number | 3 | 3 | 2 | 4 | 6 | 3 | 2 | 1 | 4 | 3 |

X: number is 3

Y: number is 2

P(X) = 4/10 = 2/5

P(Y|X) = 4/10\*2/10 = 2/25

1. What is meant by event independence? What is P(X | Y) if X and Y are independent?

If a random variable X has no effect on another random variable Y, then X and Y are said to be independent.

And if X and Y are independent, P(X | Y)= P(X) \* P(Y)

1. What is meant by conditional independence? What is P(X, Y | Z) if X and Y are conditionally independent with Z?

At least one of our features influences the target feature while others may be fully independent. This means conditional independence.

P(X,Y|Z) = P(X|Z) \* P(Y|Z)

1. What is meant by the term “overfitting”?

Overfitting means that there is a risk that if the distribution of training data does not sufficiently match the distribution of the real-world data, then the model will be overly biased towards the training distribution and will not perform well for classification or predictions of unseen data from the real-world.

1. Why is the Naïve Bayes model considered naïve? Does this matter in practice and why?

Because we simplify assumptions of conditional independence between features. It is really important in practice. It perfroms very well in practice.

1. What does the “accuracy metric” measure in binomial classification problems?

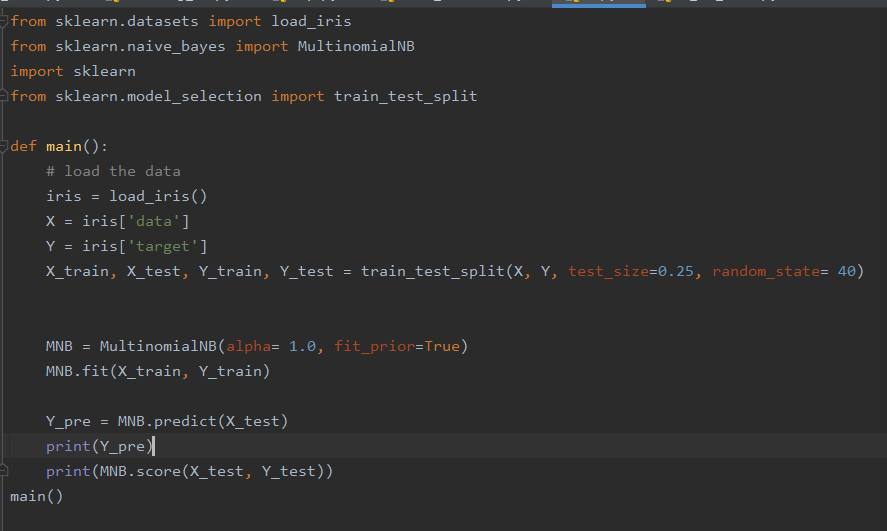
Concepts of true positives, true negatives, false positives and false negatives.

1. What is the difference between “term-frequency” and “intra-document frequency”?

The ratio of the number of occurrences of each object to the total is the “term-frequency”.

“Intra-document frequency” is an evaluation of the importance of a word to a document in a document set or corpus.

Practice:

1. I already have my Anaconda environment installed from my lab top. And I can run my python code.
2. According to watch the tutorial videos. I can run the examples as described as the tutorial videos.
3. In this practice, I know the general meaning of the source code, and then I change the value of “gram” in the “nlp.py” to 0.01, then the accuracy becomes higher. And then I don’t know the function of these codes. So I delete these, then the accuracy becomes higher.
4. in the last, I use the built-in sklean MultinomialNB model to classify the iris dataset. According to the result, the accuracy is better then one I have implemented in the previous step. As you can see the accuracy is :
5. 