CPS 844 Lab 6: Naïve Bayes Classification

The purpose of this is to give you experience with the Naive Bayes Classifiers.

Write a Python script that performs the tasks described below. Submit the .py file on D2L. Please note that if you submit your file in some other format besides .py or (.txt should you meet an issue), then your mark will at most be 60%.

Naïve Bayes Classifier

1. (10 points) Load the dataset weather.csv.
2. (20 points) As you may have noticed, the implementations of many classifiers do not support categorical data. Convert any categorical variable into dummy variables (a dummy variable takes only the value 0 or 1 to indicate the absence or presence of some attribute values). Hint: research about pandas' function get_dummies, and ensure the dummy variables are of type 'float'.
3. (10 points) At this point, you should notice that the target attribute is now split into 2 target attributes. Drop the target attribute 'play_no'.
4. (10 points) Continue with the preprocessing: separate the features attributes from the target attribute.
5. (10 points) Use the sklearn library to construct a Gaussian Naive Bayes classifier, then train this classifier.
6. (20 points) Using the trained classifier, determine if the class label of 'play_yes' is more likely to be '0' or '1', for a "new day", where the new day has the following attributes:

   Outlook = sunny, Temperature = 66, Humidity = 90, Windy = true.
7. (20 points) Given the weather of this "new day", print out the likelihood of play = yes and the likelihood of play = no.