Answering: Questions 1, 2, 4, 6

**1. Try discretising the numeric attributes in these datasets and treating them as discrete variables in the naive Bayes classifier. You can use a discretisation method of your choice and group the numeric values into any number of levels (but around 3 to 5 levels would probably be a good starting point). Does discretizing the variables improve classification performance, compared to the Gaussian naive Bayes approach? Why or why not?**

**2. Implement a baseline model (e.g., random or 0R) and compare the performance of the naïve Bayes classifier to this baseline on multiple datasets. Discuss why the baseline performance varies across datasets, and to what extent the naive Bayes classifier improves on the baseline performance**

**4. Evaluating the model on the same data that we use to train the model is considered to be a majormistake in Machine Learning. Implement a hold–out or cross–validation evaluation strategy (you should implement this yourself and do not simply call existing implementations from scikit-learn). How does your estimate of effectiveness change, compared to testing on the training data? Explain why. (The result might surprise you!)**

**6. The Gaussian naive Bayes classifier assumes that numeric attributes come from a Gaussian distribution. Is this assumption always true for the numeric attributes in these datasets? Identify some cases where the Gaussian assumption is violated and describe any evidence (or lack thereof) that this has some effect on the NB classifier’s predictions**