COMP 135 – Machine Learning – Fall 2016

Homework Assignment 2

Due date: Wednesday 10/5 (hardcopy in class)

1. The *uneven* factory produced a special die as follows. As usual the die has 6 sides. But the values printed on these sides and their probability of occurring when tossed follows the following characteristics: (2,0.5),(4,0.2),(8,0.1),(16,0.1),(32,0.08),(64,0.02) where each pair represents a value and its probability.

   (i) Let X be a random variable capturing the value of the die when tossed. Calculate the expectation, variance, and standard deviation of X. Please make sure to show and explain the steps in your computation.

   (ii) The die is tossed 10 times independently to produce values $Y_1, Y_2, \ldots, Y_{10}$. Let the sum be $S = \sum_{i=1}^{10} Y_i$. Calculate the expectation, variance, and standard deviation of S. Please make sure to show and explain the steps in your computation.

2. A 4-way coin with sizes A,B,C,D is tossed 20 times independently to produce the sequence of observations: ADBDAAACBDBCDACABBBA. Describe the probability model and calculate the maximum likelihood estimate for its parameters. Please make sure to show and explain the steps in your computation.

3. A Poisson random variable X takes natural numbers as values where $p(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$ and where λ is the parameter of the distribution. You observe 10 independent samples from X with values: 4,9,1,22,7,3,6,4,9,5. Calculate the maximum likelihood estimate for λ. Please make sure to show and explain the steps in your computation.

4. Ideally a good learning algorithm will be robust against simple manipulation of features. In this question we consider the Naive Bayes algorithm. In particular consider a dataset D_1 with discrete features (let's say binary for simplicity) and consider a variant D_2 where one of the features has been duplicated 100 times. Is the classifier produced by Naive Bayes identical when it learns on D_1 versus D_2? If you answer Yes please explain your answer clearly. If you answer No explain and give an example where the prediction of the two classifiers will be different.