Assignment: Naive Bayes and Logistic Classifiers

Problem 1 (10 points(2 points each))

Implement a Naive Bayes classifier using e1071 package in R to predict whether it is a good day to play tennis based on the information in the training data. The file "NaiveBayes.csv" provides information about whether previous days were good or bad days for playing tennis. Use the last row in this file for test and all other rows for training.

1. What's the prior distribution for playing/not playing tennis?
2. Determine the likelihood for outlook (sunny/not sunny).
3. Determine the likelihood for temperature (hot/not hot).
4. Determine the likelihood for wind (windy/not windy).
5. What's the posterior probability for the test pattern? Do you predict playing tennis or not for the given weather condition?

Problem 2 (65 points)

In this problem you cannot use an existing Naive Bayes classifier implementation or package.

Implement a Naive Bayes classifier in R to apply it to the task of classifying handwritten digits. Files mnist-train and mnist-test contain training and test digits, together with their ground truth labels (first column). Each row in these files corresponds to a different digit.

Each image is 28x28, hence there are 784 pixel in every image. Columns 2-785 in the data files correspond to the pixel intensity, a value between 0 to 255. Column 1 corresponds to the correct label for each digit.

You should convert the pixel intensities to a single binary indicator feature (F_i) for each pixel. Specifically, if the intensity is smaller than 255/2 map it to a zero, otherwise to a one.

1. (10 points) Estimate the priors $P(class)$ based on the frequencies of different classes in the training set. **Report the values in a table. Round to 3 decimal places.**
2. (15 points) Estimate the likelihoods $P(F_i|class)$ for every pixel location i and for every digit class from 0 to 9. The likelihood estimate is

   $P(F_i = f|class)$ =(Number of times pixel i has value f in training examples from this class) / (Total number of training examples from this class)

   Note that you have to smooth the likelihoods to ensure that there are no zero counts. Laplace smoothing is a very simple method that increases the

observation count of every value f by some constant k. This corresponds to adding k to the numerator above, and k*V to the denominator (where V is the number of possible values the feature can take on). The higher the value of k, the stronger the smoothing. Experiment with different integer values of k from 1 to 5. While you need to find all the likelihoods for k=1 to 5, I'd like you to report the following values in your report: **For k=1 and k=5** $P(F_{682} = 0|class = 5)$ **and** $P(F_{772} = 1|class = 9)$**. Round to 3 decimal places.**

3. (25 points) Perform maximum a posteriori (MAP) classification of test digits according to the learned Naive Bayes modeles. Suppose a test image has feature values f_1, f_2, ... , f_{784}. According to this model, the posterior probability (up to scale) of each class given the digit is given by:

   $$P(class)P(f_1|class)(f_2|class)...P(f_{784}|class)$$

   Note that in order to avoid underflow, you need to work with the log of the above quantity:

   $$\log P(class) + \log P(f_1|class) + \log P(f_2|class) + ... + \log P(f_{784}|class)$$

   Compute the above decision function values for all ten classes for every test image, then use them for MAP classification. **For the first test image, report the log posterior probability of** $P(class = 5|f_1, f_2, ..., f_{784})$ **and** $P(class = 7|f_1, f_2, ..., f_{784})$ **for k=1 and k=5.**

4. (10 points) Use the true class labels of the test images from the mnist_test file to check the correctness of the estimated label for each test digit. **Report your performance in terms of the classification rate (percentage of all test images correctly classified) for each value of k from 1 to 5.**

5. (5 points) **Report your confusion matrix for the best k.** This is a 10x10 matrix whose entry in row r and column c is the percentage of test images from class r that are classified as class c. (Tip: You should be able to achieve at least 70% accuracy on the test set.)

Problem 3 (15 points)

Given the function $f(x) = x^2 + 6x$:

1. Use derivative of $f(x)$ to find the value of x that minimizes this function. (2 points)

2. Use gradient descent to find the value of x that minimizes this function. Compare your answer with the previous part. (13 points)

Problem 4 (35 points)

The Space Shuttle Challenger disaster occurred on January 28, 1986, when it broke apart 73 seconds into its flight, leading to the deaths of its seven crew members. The spacecraft disintegrated over the Atlantic Ocean, off the coast of central Florida at 11:38 EST. Disintegration of the entire vehicle began after an O-ring seal in its right solid rocket booster failed at liftoff. Subsequently, a special commission was appointed to investigate the accident. The commission found that NASA disregarded warnings from engineers about the dangers of launching posed by the low temperatures of that morning, claiming that engineers could not provide a convincing argument against the launch (source: Wikipedia, Applied Probability for Engineers).

File Oring.csv provides data on launch temperature and O-ring failure for the 24-space shuttle launches prior to the Challenger disaster. There are six O-rings used to seal field joints on the rocket motor assembly. A +1 in the O-rings indicates that at least one O-ring failure had occurred on that launch and a 0 indicates that no failure had occurred.

1. Normalize the launch temperature using the expression $\frac{x-\mu}{\sigma}$. (3 points)
2. Create a logistic regression model using the gradient decent technique to predict the probability of O-ring failure based on the launch temperature. Provide the equation for your model. (20 points)
3. Provide a plot of the original data along with your logistic model. (5 points)
4. The actual temperature at the Challenger launch was 31 degrees Fahrenheit. According to your model what was the probability of O-ring failure on the Challenger launch? Could the engineers have used your model to provide a convincing argument to NASA? Elaborate. (7 points)