ML 4375 – Intro to Machine Learning – Summer 2018 – Mazidi

Homework 4: Naïve Bayes and SVM

Objective: In this homework you will run various ML algorithms on the Auto data set.

**Turn in:** an “.Rmd” file to eLearning

**Important**: Use prominent headings for the steps like “# 1” so the TA can find your code.

1. Set up the Auto data:
   1. Load the ISLR package and the Auto data
   2. Determine the median value for mpg
   3. Use the median to create a new column in the data set named mpglevel, which is 1 if mpg>median and otherwise is 0. Make sure this variable is a factor. We will use mpglevel as the target (response) variable for the algorithms.
   4. Use the names() function to verify that your new column is in Auto
   5. create a 75-25 train/test split, using seed 1234 but do not include columns ‘name’ or ‘mpg’ in either train or test
2. Plots
   1. Set up a 2x2 graph grid and plot the following pairs of plots
   2. Plot pair 1: plot horsepower~mpg and weight~mpg, setting colors according to the factor mpglevel, ex: col=(Auto$mpglevel)
   3. Plot pair 2: plot horsepower~mpglevel and weight~mpglevel
3. Build a Naïve Bayes model
   1. build the model on the train set
   2. use the predict() function on the test set
   3. create a table comparing predicted to actual values for mpglevel
   4. calculate the mean accuracy
4. SVM linear kernel
   1. use the tune() function to perform cross-validation to determine the best value for cost
   2. use the parameter(s) from the previous step to build an svm model with a linear kernel on the train set
   3. use the predict() function on the test set
   4. create a table comparing predicted to actual values for mpglevel
   5. calculate the mean accuracy
5. SVM polynomial kernel
   1. use the tune() function to perform cross-validation to determine the best values for cost and degree
   2. use the parameter(s) from the previous step to build an svm model with a polynomial kernel on the train set
   3. use the predict() function on the test set
   4. create a table comparing predicted to actual values for mpglevel
   5. calculate the mean accuracy
6. SVM radial kernel
   1. use the tune() function to perform cross-validation to determine the best values for cost and gamma
   2. use the parameter(s) from the previous step to build an svm model with a radial kernel on the train set
   3. use the predict() function on the test set
   4. create a table comparing predicted to actual values for mpglevel
   5. calculate the mean accuracy
7. Questions.
   1. Compare the accuracy results for the 4 models
   2. Discuss the advantages and disadvantages of Naïve Bayes versus svm

|  |  |
| --- | --- |
| Element | Points |
| R/Rmd script runs without errors | 30 |
| Appropriate comments and white space | 10 |
| Steps 1-6 | 10 points each |