REPLACE WITH COVER PAGE

SVM and Logistic Regression Modeling 1

Classification Models 1

Advantages of the Models 2

Interpreting Features 3

Support Vectors Insights 4

# SVM and Logistic Regression Modeling

Classification problems revolve around being able to predict what category a feature falls into within the dataset. The common tools for classification problems are:

* + Percent correction classification (PCC): measures overall accuracy. Every error has the same weight.
  + Confusion matrix: also measures accuracy but distinguished between errors, i.e false positives, false negatives and correct predictions.
  + Area Under the ROC Curve (AUC – ROC): is one of the most widely used metrics for evaluation. Popular because it ranks the positive predictions higher than the negative. Also, ROC curve it is independent of the change in proportion of responders.
  + Lift and Gain charts: both charts measure the effectiveness of a model by calculating the ratio between the results obtained with and without the predictive model. In other words, these metrics examine if using predictive models has any positive effects or not.

Classification Models

The task at hand for this dataset is to use a logistic regression model and support vector machine model for our classification task as well as assessing how each model performs under an 80/20 testing and training data split for the data set.

Advantages of the Models

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Interpreting Features

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliq

Support Vectors Insights

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliq

**SVM and Logistic Regression Modeling**

**• [50 points] Create a logistic regression model and a support vector machine model for the**

**classification task involved with your dataset. Assess how well each model performs (use 80/20 training/testing split for your data). Adjust parameters of the models to make them more accurate. If your dataset size requires the use of stochastic gradient descent, then linear kernel only is fine to use.**

**• [10 points] Discuss the advantages of each model for each classification task. Does one type of model offer superior performance over another in terms of prediction accuracy? In terms of training time or efficiency? Explain in detail.**

**• [30 points] Use the weights from logistic regression to interpret the importance of different features for each classification task. Explain your interpretation in detail. Why do you think some variables are more important?**

* **[10 points] Look at the chosen support vectors for the classification task. Do these provide any insight into the data? Explain.**