Untitled

Support Vector Machines

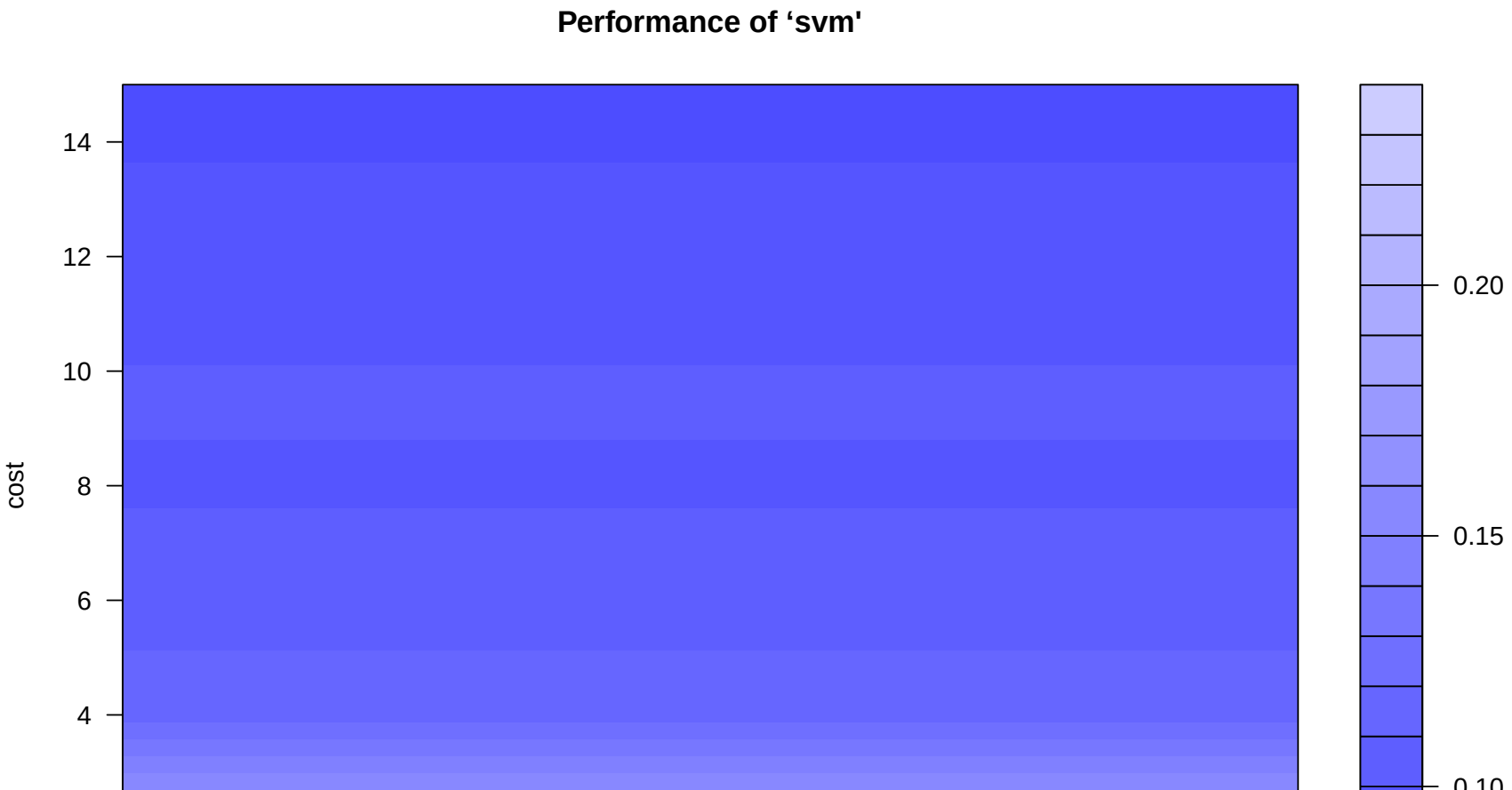
Support Vector Machines, or SVM, is a supervised machine learning algorithm that utilizes a seperating hyperplane to analyze data for regressin and classification analysis. We used SVM on the Bumpus data to figure out how to use the qualitative predictors in order to predict Survival (Dead or Alive) - e.g., as a classfication technique.

Model Parameters

Two important parameters are cost and epsilon. Cost is the amount of penalty given to samples within the margins (reducing variance at the expense of increasing bias). A higher cost penalizes the amount within the margins more. Epsilon controls the amount of error allowed to be admitted in the solution. A larger episolon meants that there is less penalty per error.

Model Parameters

Using 6-fold cross validation, we determined the cost and epsilon values needed to reduce the test error rate (darker means more accurate). Epsilon seems like it doesn't matter on the whole, whereas cost seems to matter but then plateus. An epsilon of 0 (errors penalized significantly) and a cost of 10 (more penalty to samples within the margin, but not by much) seemed to be the best.



Training the SVM Model

We used 6-fold cross validation when training and testing the model, and the resulting training error rate was 27.9%. Here is a table of our result:

```
##
##          Alive Dead
##    Alive    25    9
##    Dead     10   24
```

Testing the SVM Model

Our test error rate was 29.4% (so a percent correct of 70.6%, slightly lower then the training error rate, as expected, but not by much). Overall, it is better then random chance. Here is a table of our result:

```
##
## pred.test Alive Dead
##     Alive    30   13
##     Dead      7   18
```