Characterization of Jet Charge at the Large Hadron Collider

Thomas Dylan Rueter, Krishna Soni

Sample Data:

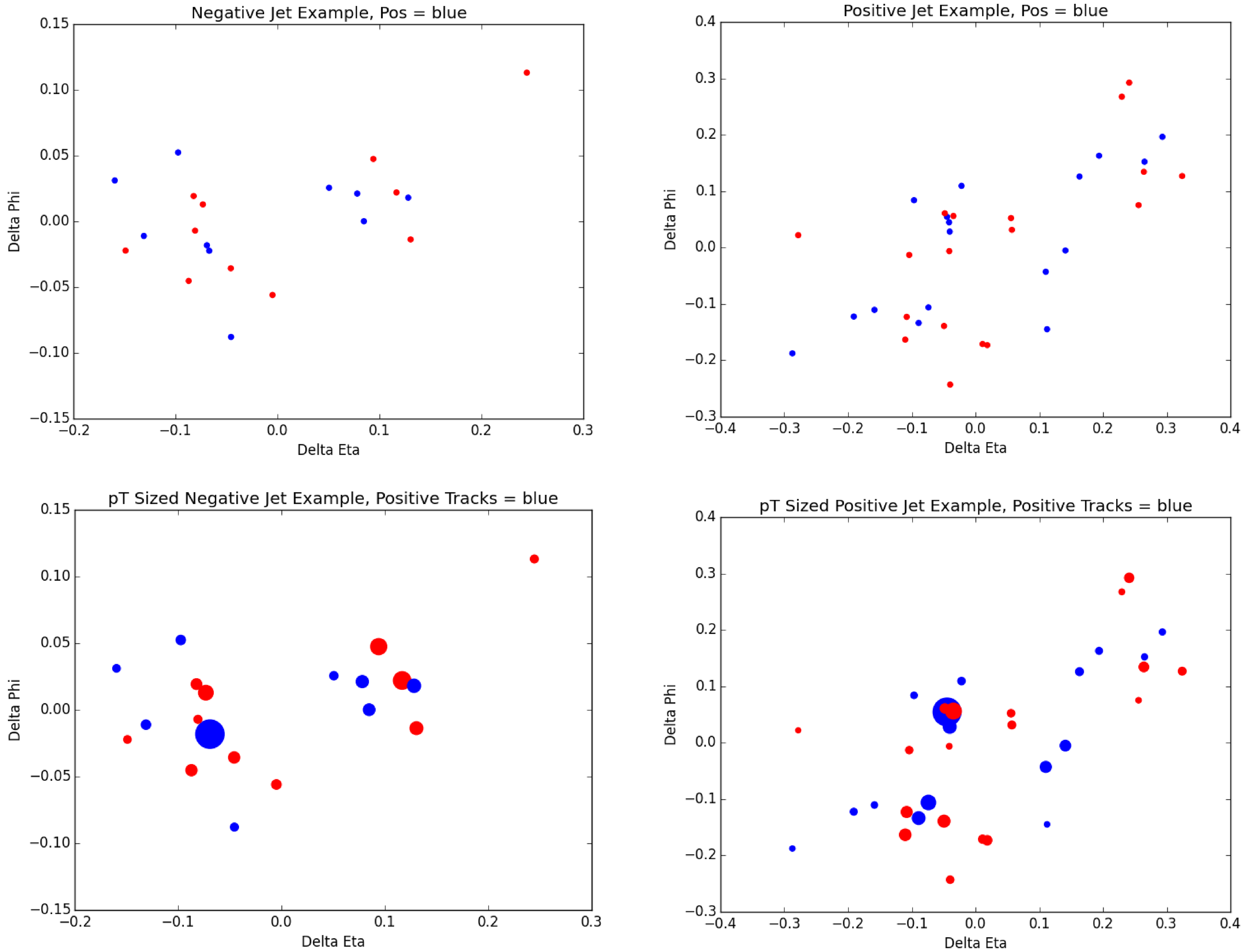

Figure 1: Two sample jets, one positive and one negative. The origin corresponds to the jet's center. The bottom two jets have the size of the particle track proportional to the particle's pT.

The Goal:

Predicting the overall charge of a jet based on the properties of the particles within the jet, with a better Positive Tagging Efficiency (PTE) and Negative Rejection (NR) than the current method of jet charge calculation, which is calculated as:

$$Q_{jet}= \frac{1}{(p_{T\,jet})^{\kappa}} \sum_{i \in Tracks} q_i (p_{Ti})^{\kappa}$$

Here κ is a parameter fixed for the calculation which can take on a range of values $0 \leq \kappa \leq 1$. Different values of κ provide different levels of performance in the jet charge calculation.

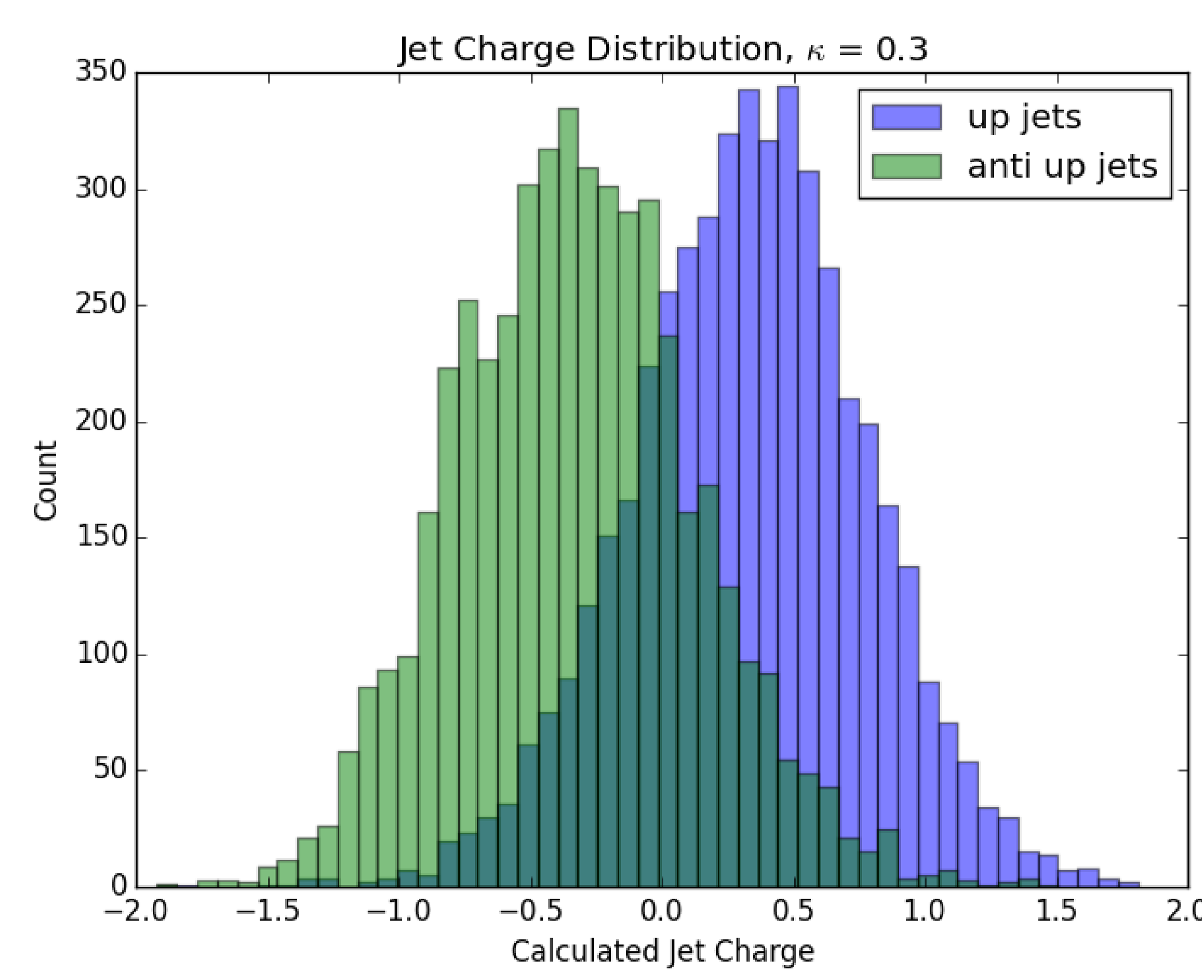

Figure 2: Calculated jet charges for the training set, using κ = 0.3.

Features:

The feature set consists of the particles identified by the ATLAS detector as belonging to the jet. Each particle in the jet has a measured transverse momentum (pT) in GeV, charge, position in the detector, and distance to the center of the jet (dR). Our initial physical intuition, aided by the traditional jet charge calculation, led us to take the pT, charge, and dR for the 5 particles with the largest pT and use these 15 values as our feature set. These should contribute most to the jet charge calculation, and also are indicative of relatively early production in the jet, so perhaps more indicative of initial charge.

The Model:

Our first attempt to characterize the data featured a Support Vector Machine based analysis. This achieved a very low training error and as a result good jet charge classification. This initial attempt (with the training set used to calculate the NR and PTE) fell above the ROC curves for the calculation method, as shown in the figure below. When we performed cross validation with our training set, however, the performance suffered dramatically, with a test error ~50%. This led us to consider our features more carefully. Again drawing from the traditional calculation, we tried adding in the particle charge times the pT as a feature for each particle, and also used a PCA approach to reduce the dimensionality of the feature space. When this did not improve performance, we regularized our pT values, since the SVM is not a scale invariant algorithm. We also tried to use the AdaBoost algorithm on our SVM, which did improve performance somewhat. We attempted to use a boosted decision tree method as well, but that proved fairly fruitless (see Figure 4).

Future Work:

We believe that we will be able to achieve better performance through further feature regularization and selection. Using different kernels may also improve the performance, and is something we intend to investigate. Collecting more training data will also aid our classification attempt.

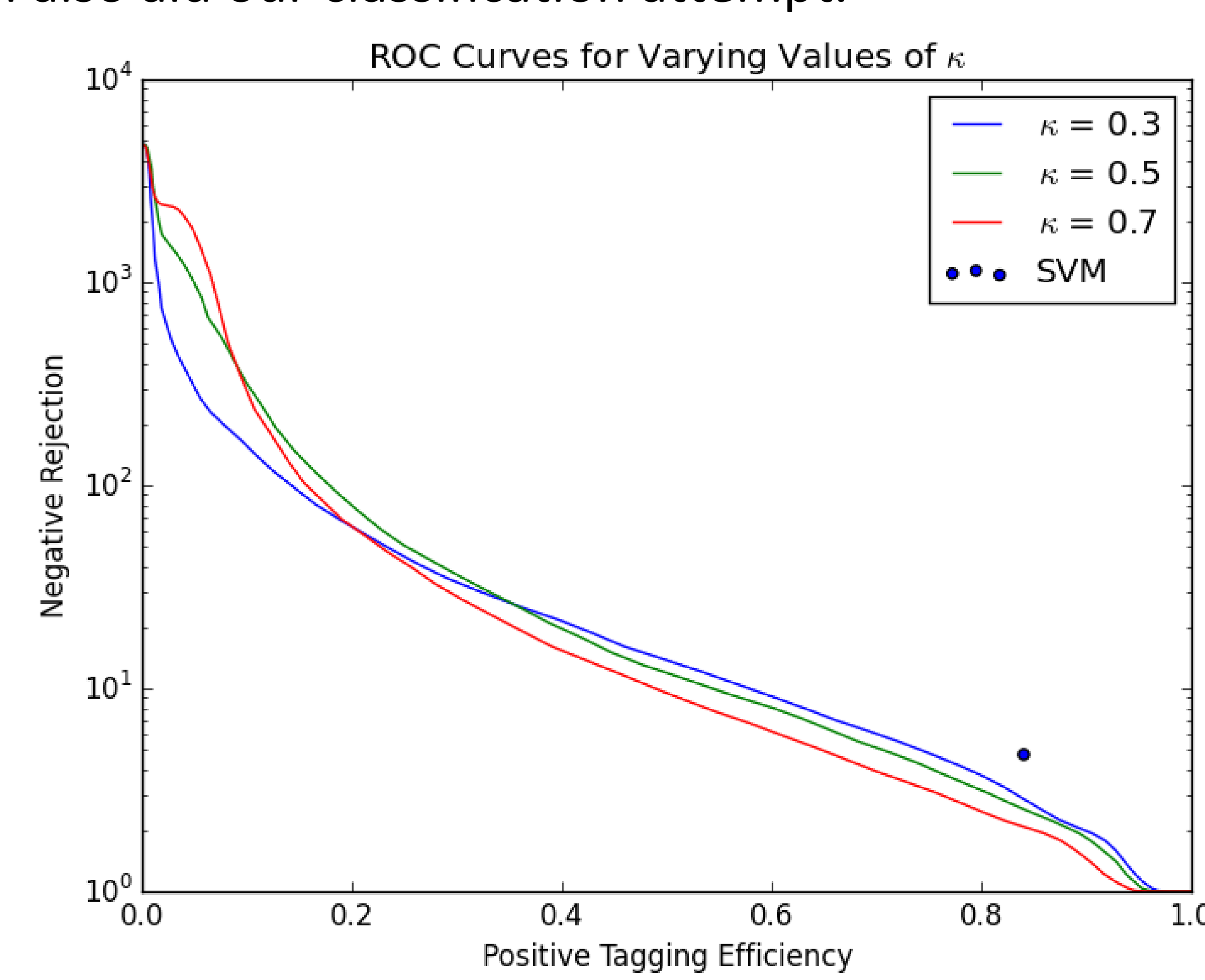

Figure 3: ROC curves calculated from making cuts on the calculated jet charge distribution and classifying all jets above the cut as positive.

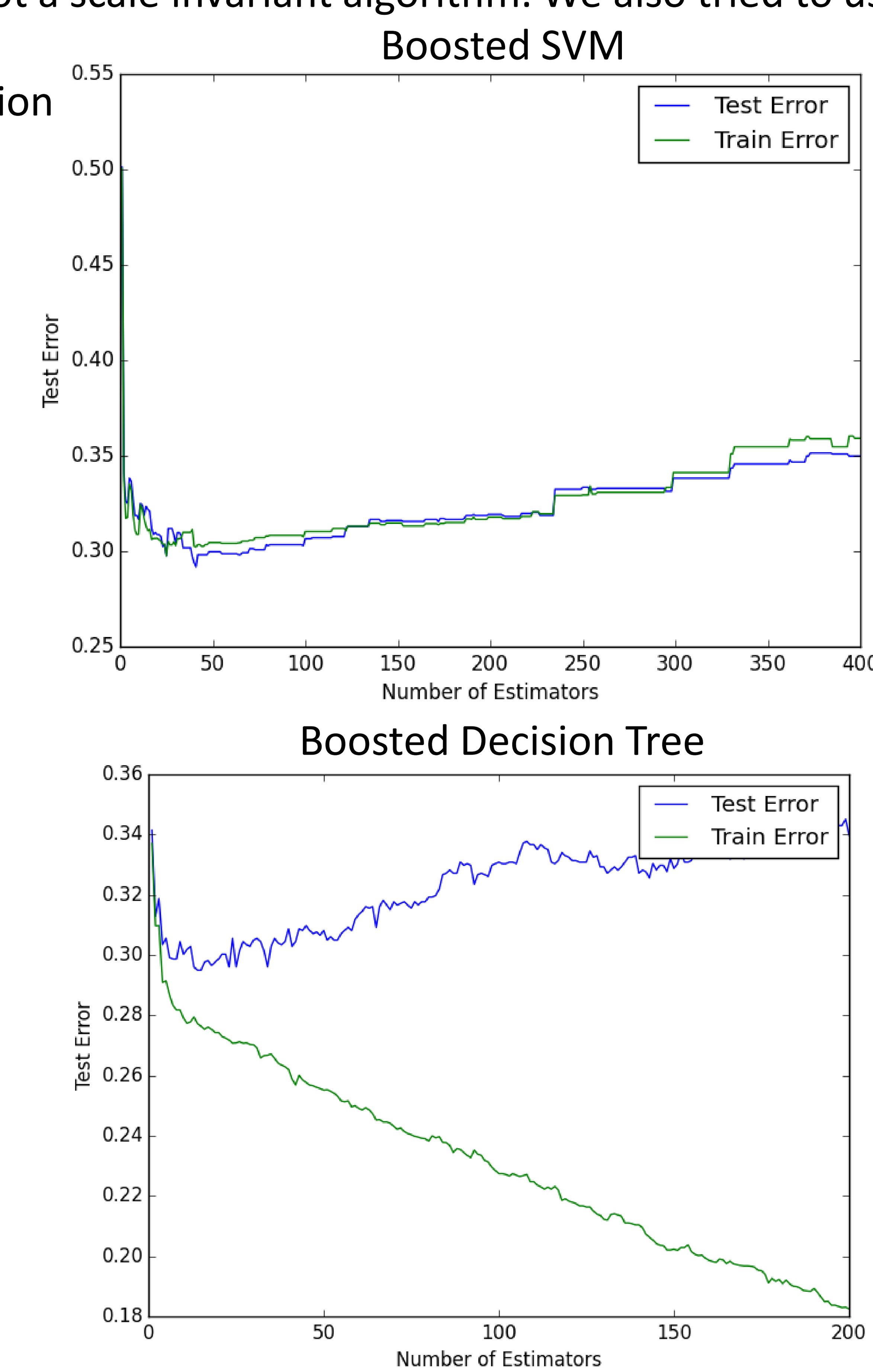

Figure 4: Test Error from cross-validation as a function of the number of estimators for the boosted SVM and boosted decision tree.