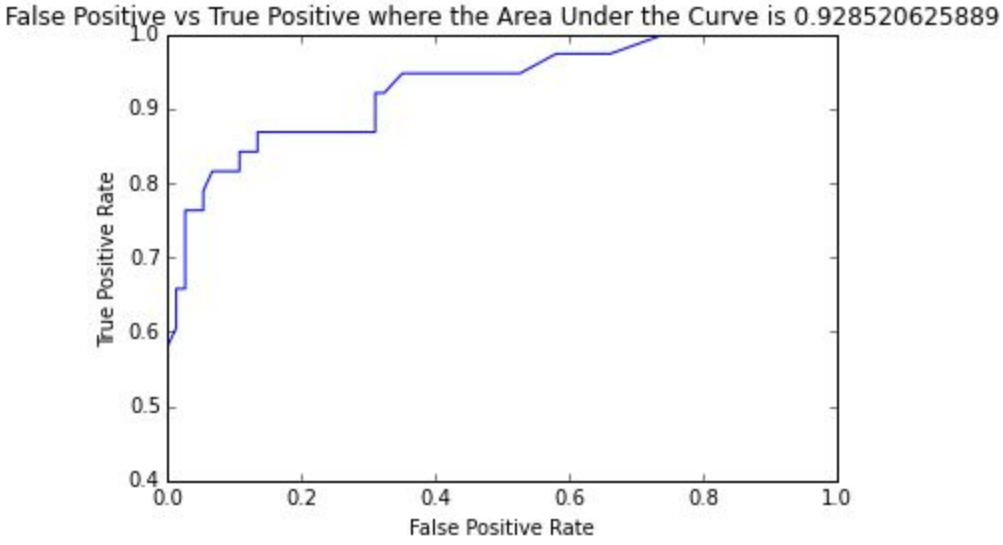
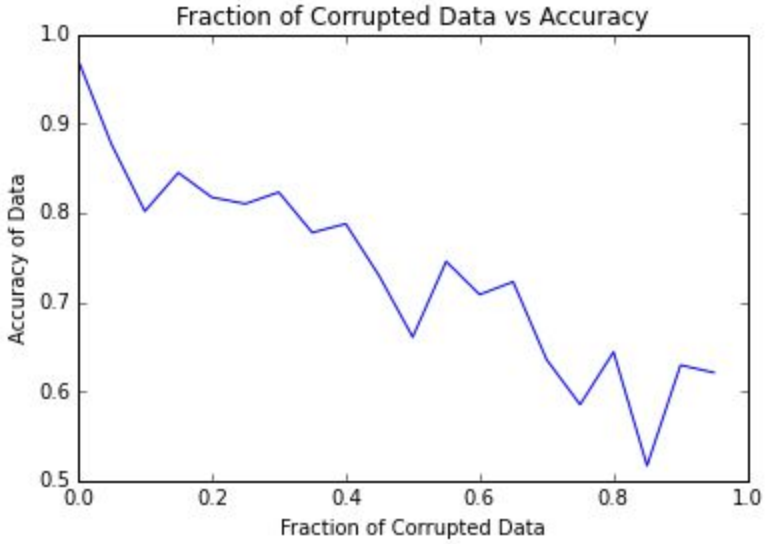
Haley Bates-Tarasewicz





Full data set metrics:

---
full training set accuracy: 0.912472647702
full testing set accuracy: 0.912472647702
full training set sensitivity: 0.844827586207
full testing set sensitivity: 0.815789473684
full training set specificity: 0.95406360424
full testing set specificity: 0.918918918919
full training set ppv: 0.91875
full testing set ppv: 0.837837837838

All of the calculated metrics in the training sets are higher than or stayed the same as the metrics in the testing set. This makes sense, because the model was trained on the training set, so would be better fitted to that set of data rather than the testing set.

reduced data set metrics:

---
reduced training set accuracy: 0.911062906725
reduced testing set accuracy: 0.911062906725
reduced training set sensitivity: 0.810055865922
reduced testing set sensitivity: 0.757575757576
reduced training set specificity: 0.975177304965
reduced testing set specificity: 0.96
reduced training set ppv: 0.953947368421
reduced testing set ppv: 0.892857142857

For the reduced metrics, similarly the metrics for the training sets are higher than or the same as the metrics in the testing set. The model was trained in the exact same way as the full data set, so this makes sense for the same reasons. About half of the metrics in the reduced data set are higher than the metrics in the full data set. I know that the metrics for the reduced data set should be higher than the full data set, and I'm not sure why my code doesn't do this.