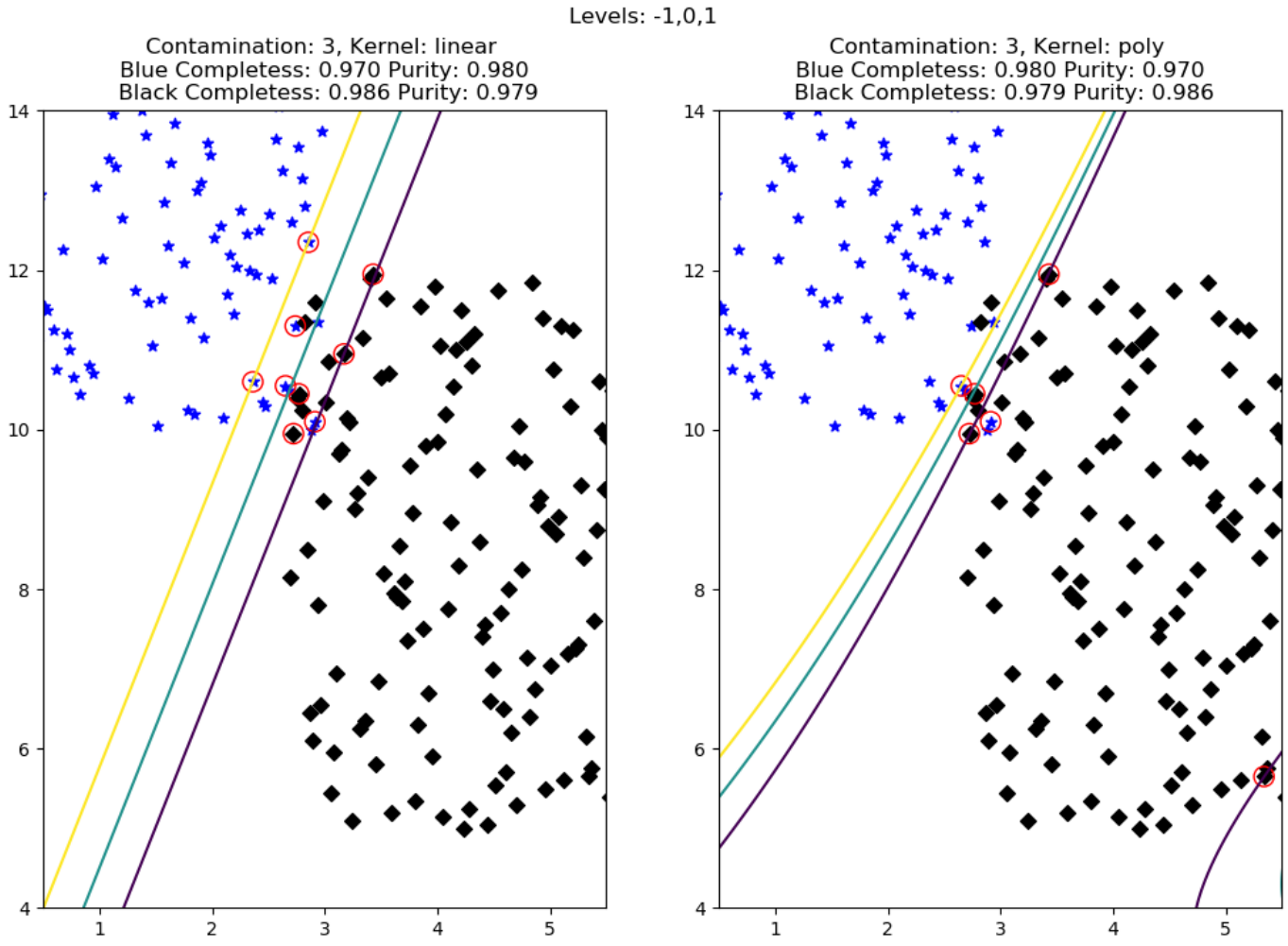
Bethany Ludwig

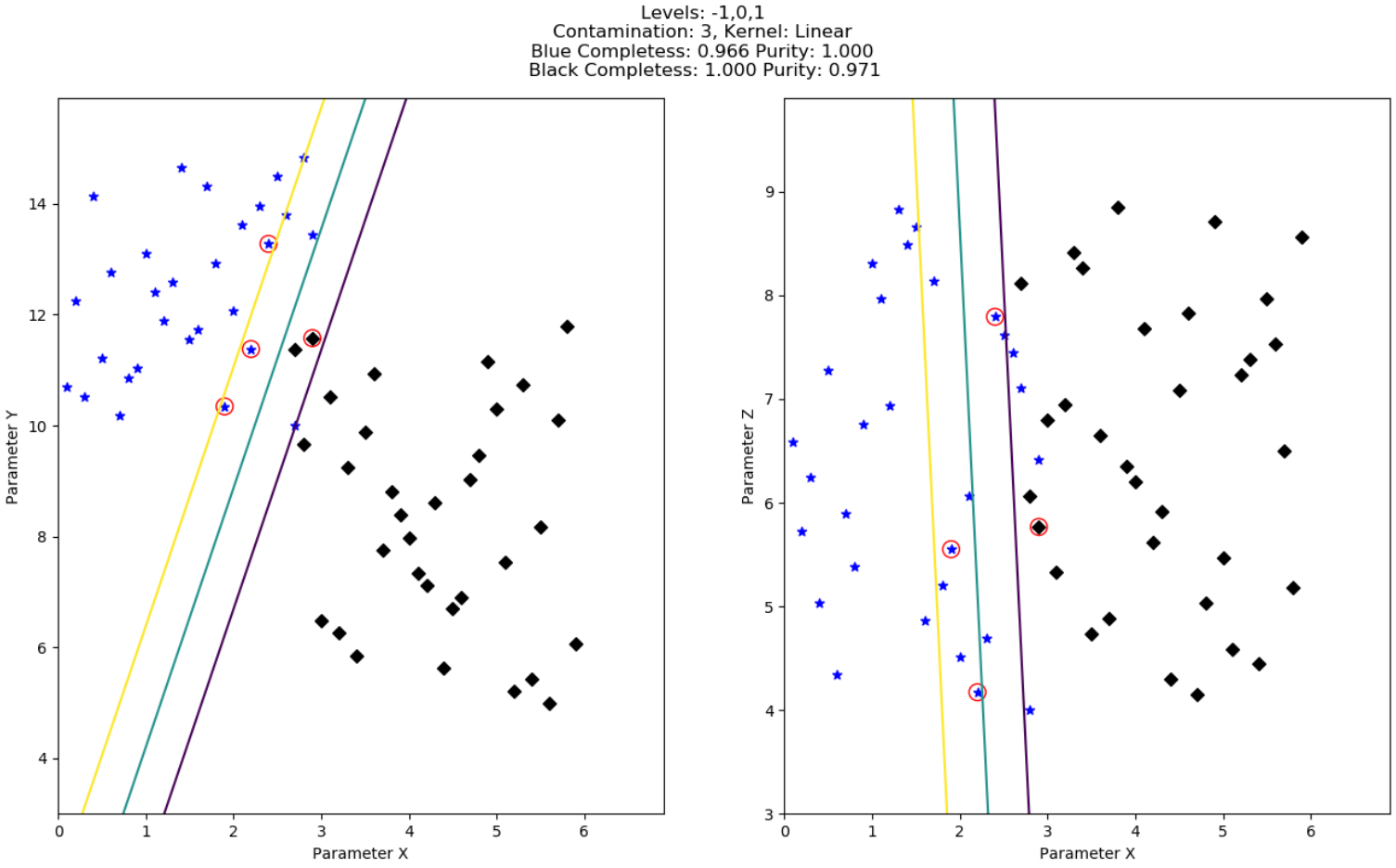
For this assignment I simulated fake data. Last assignment I used real data but this was such a large dataset that trying to accomplish the homework proved time consuming. The parameters also don't separate very cleanly from each other in the galaxy zoo/sdss data set so for visualization I thought it would be best to use a simulated set.

Plot the data set with decision boundary, margins, and label the support vectors from SVM.



Support vectors are circled in red.

Plot the data set again in two 2D plots, showing the third feature in relation to the first two.



- Write a few sentences about the what you expected to happen to the completeness and purity with an additional feature. Is this what happened? Why or why not?

Since the data would be more constrained by a third feature, I would expect the completeness and purity to be higher than the data with 2 features. This appears to both be true and not true.

Blue purity and black completeness are higher but blue completeness and black purity is lower.

It could be the case that you would expect more extremes since purity and completeness seem to be inverted from each other and that does seem to be the case.

It could also be the case that something went horribly wrong as evidenced by the stray support vector point as a black diamond outside of the contours.