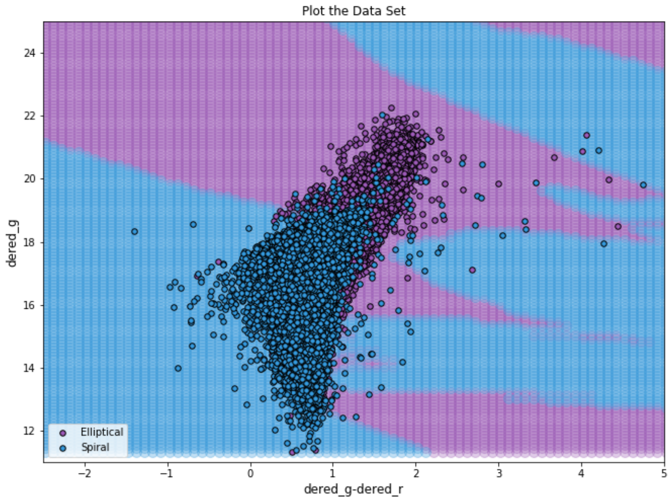
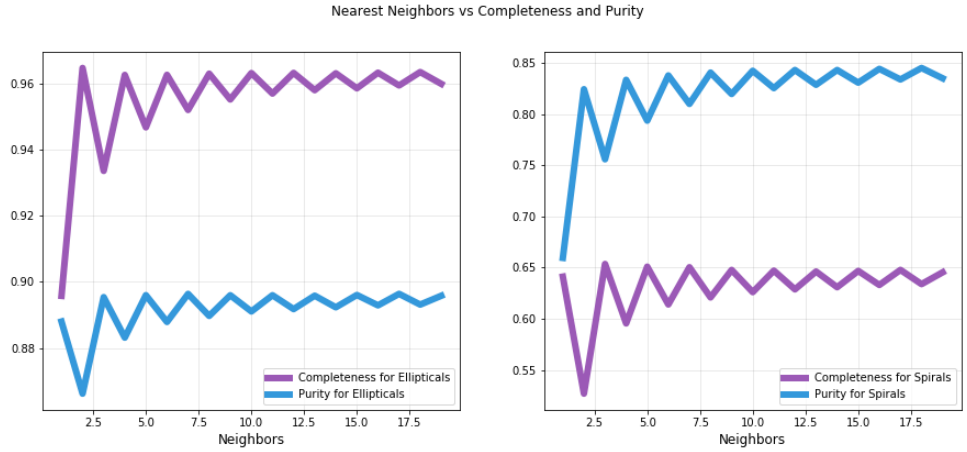


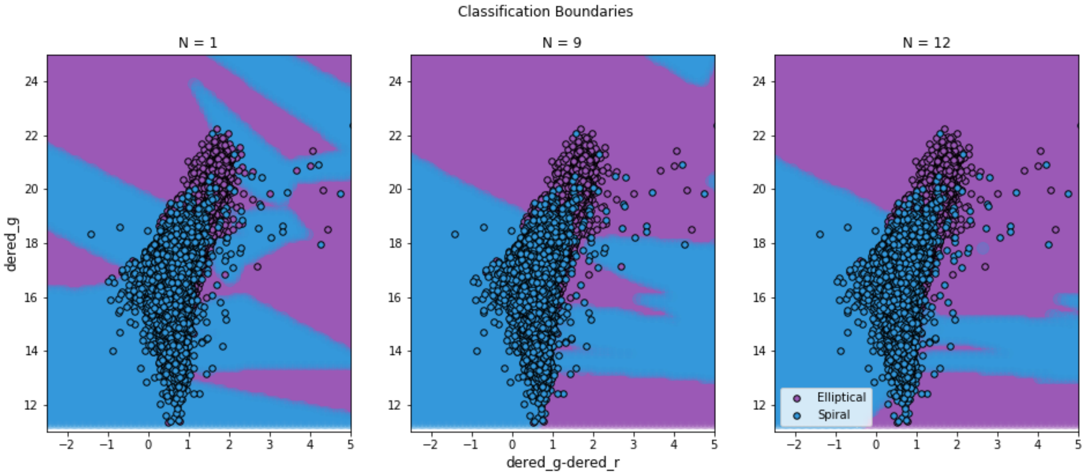
Bethany Ludwig
Machine Learning
Mini Course
April 16th 2019

Assignment 1

To test nearest neighbors as a classification technique in machine learning, I used the galaxy zoo data set combined with SDSS parameters. I looked at color, as a feature to separate elliptical galaxies from spirals. Although not always true, ellipticals tend to be red and spirals tend to be blue. Other interesting features in this parameter space could be things like concentration or texture.



While k = 2 has a peak in completeness for ellipticals, it appears to have an inverted peak for purity and is not a peak for spirals. To choose a more stable value for k, I decided to go with k=9. The classification boundaries are still somewhat sensitive to outliers compared to k=12 but not overly sensitive like k=1.


Classification Boundaries
N = 1
N = 9
N = 12
dered_g
dered_g-dered_r
Elliptical
Spiral