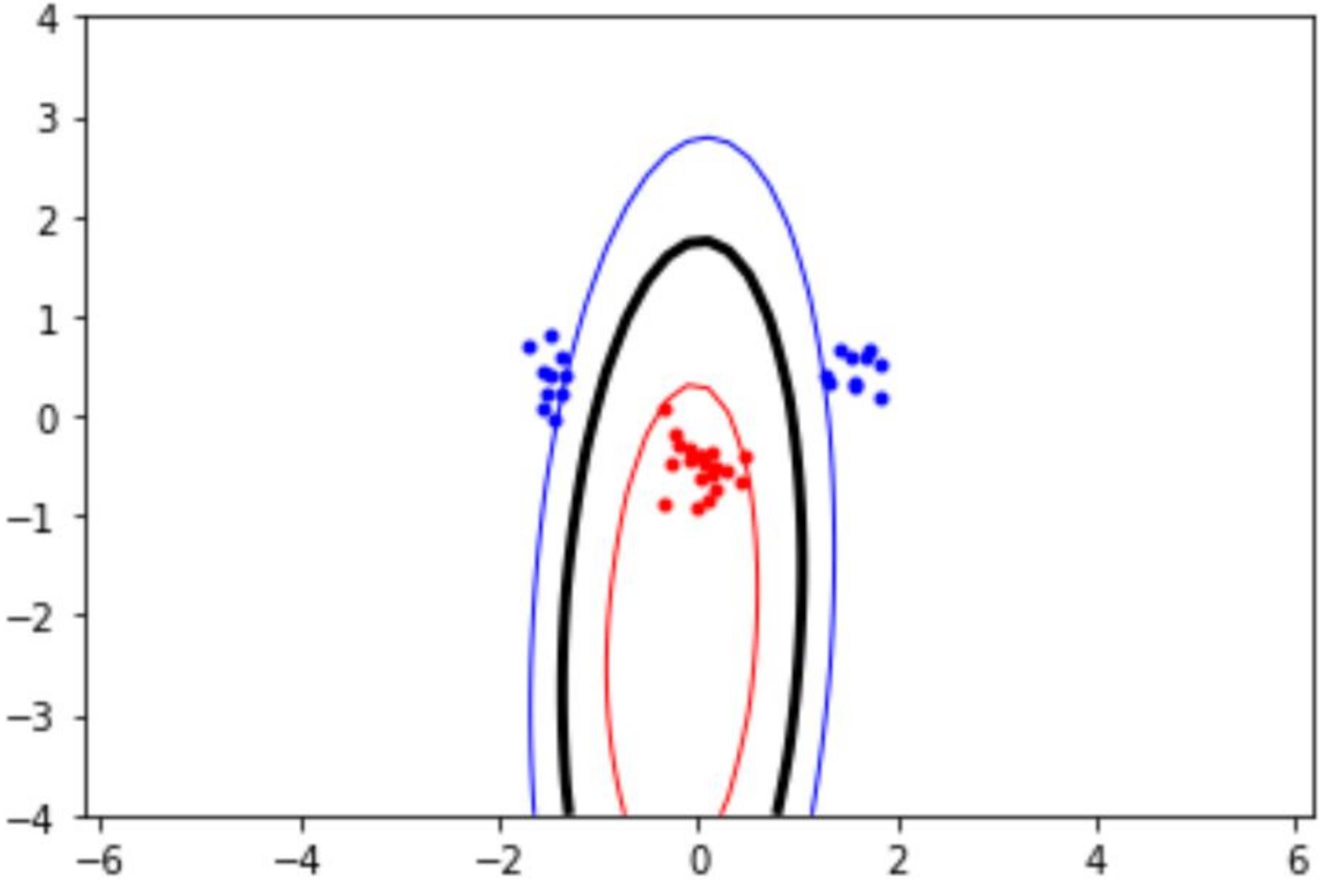
Machine Learning Lab 2

Qiao Ren, Catherine Weldon

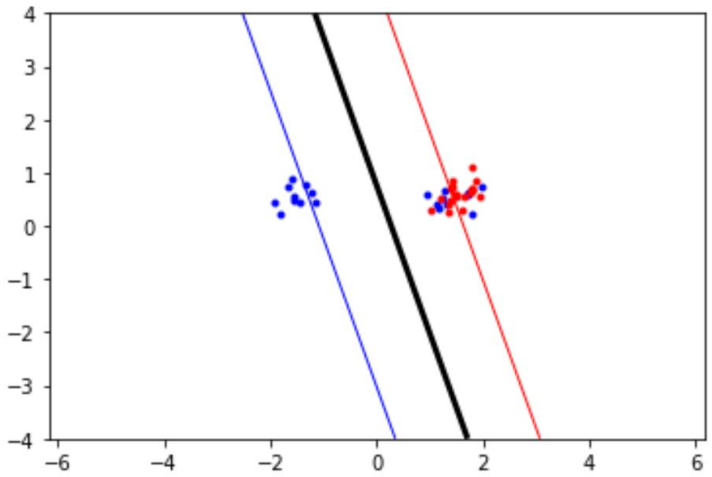
Assignment 0

- Graph using the poly kernel:

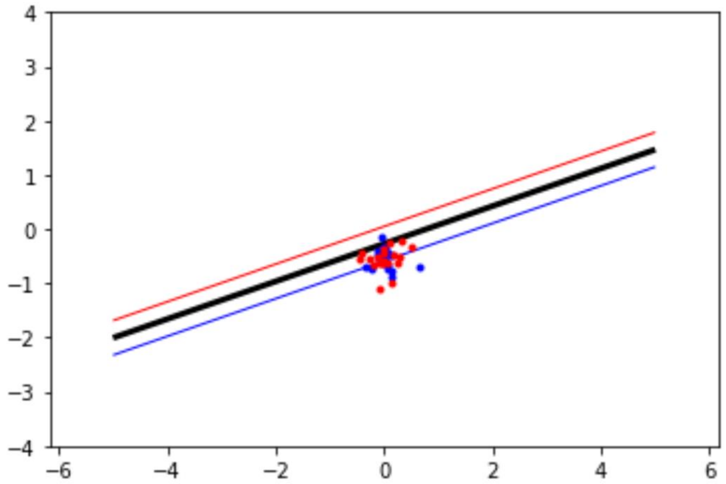


Assignment 1

- Move the clusters around and change their sizes to make it easier or harder for the classifier to find a decent boundary. Pay attention to when the optimizer (minimize function) is not able to find a solution at all.



linear C=0.1

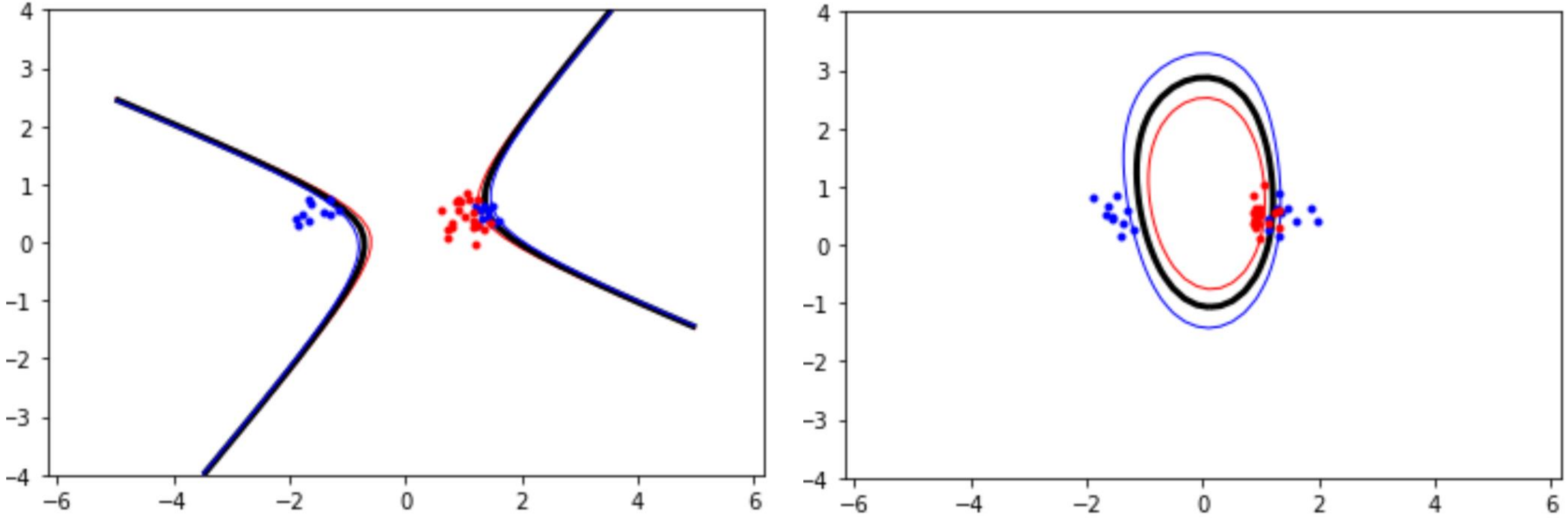


linear C=1

- Linear Kernel does not allow for good separation when clusters overlap
- if cost C is too large (1000), then slack is low, → not able to find a solution

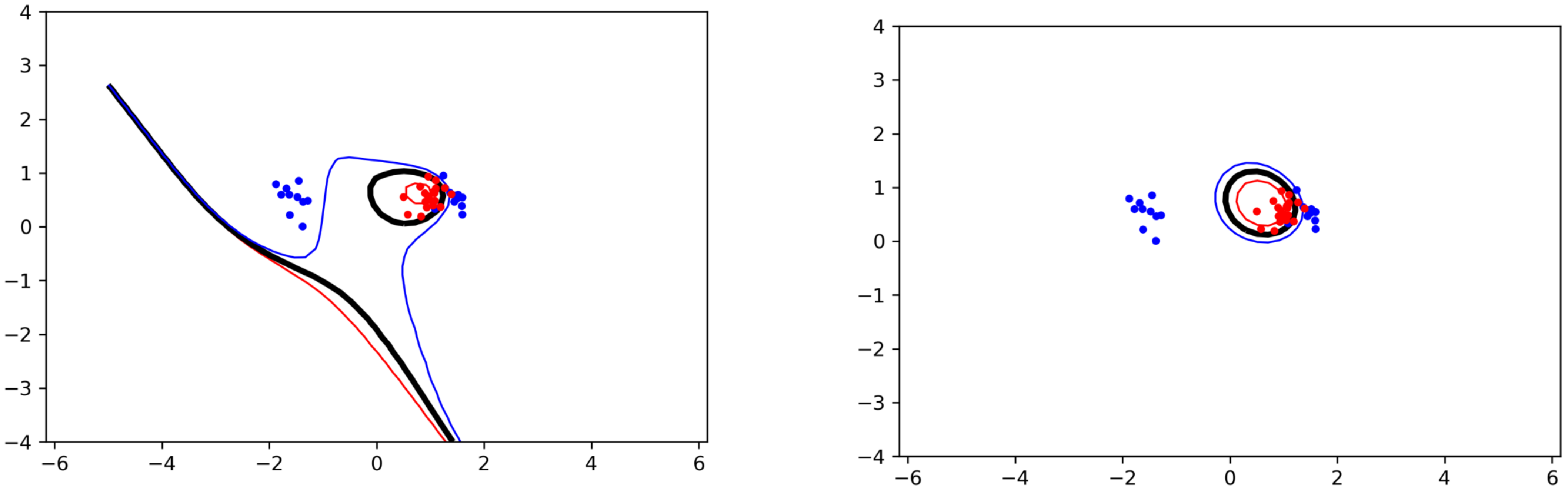
Assignment 2

- Implement the two non-linear kernels. You should be able to classify very hard data sets with these.



Poly: order 3

RBF: Sigma = 1



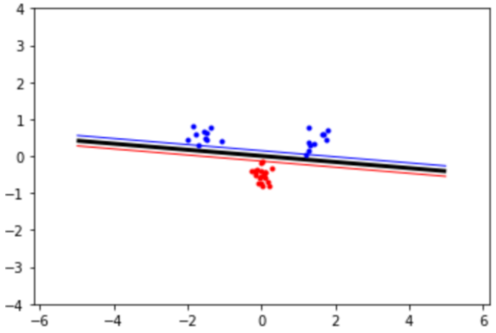
4
3
2
1
0
−1
−2
−3
−4
−6
−4
−2
0
2
4
6
4
3
2
1
0
−1
−2
−3
−4
−6
−4
−2
0
2
4
6

Assignment 3

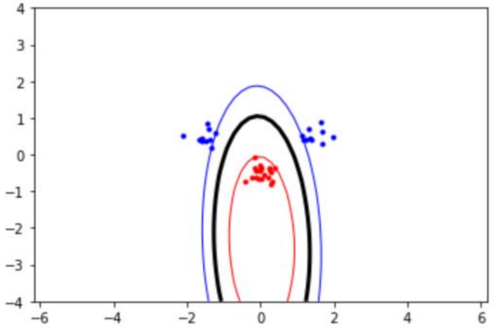
- The non-linear kernels have parameters; explore how they influence the decision boundary. Reason about this in terms of the bias/variance trade-off.

polynomial C=10

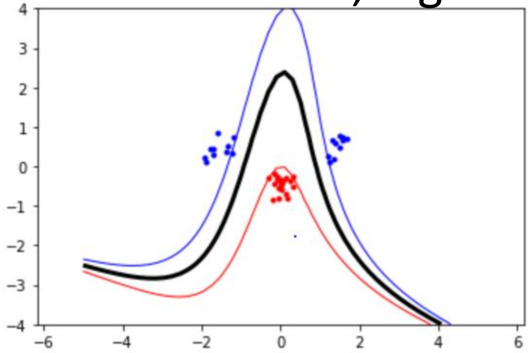
power =1



power=2

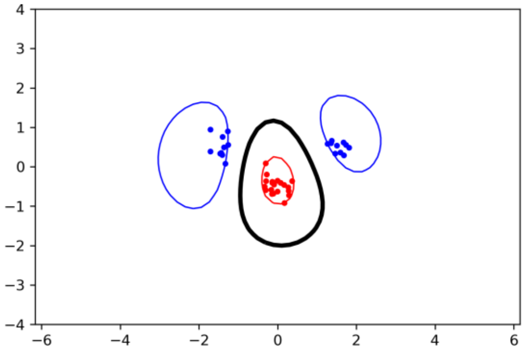


power=3 low bias, high variance

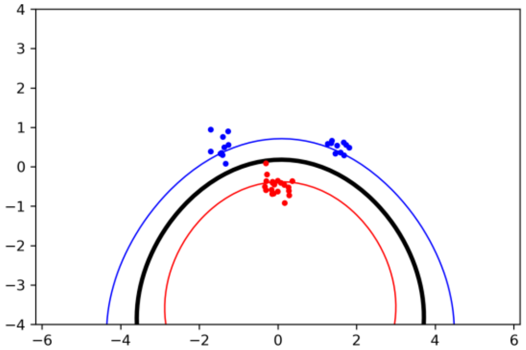


RBF C=10

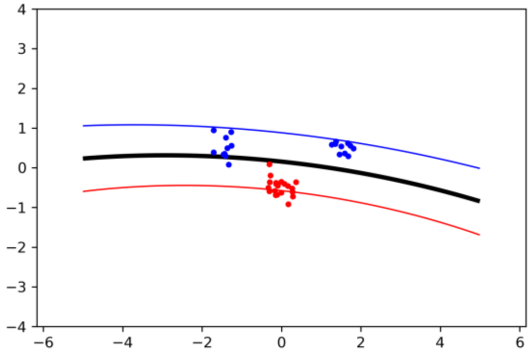
sigma =1 low bias, high variance



sigma=4

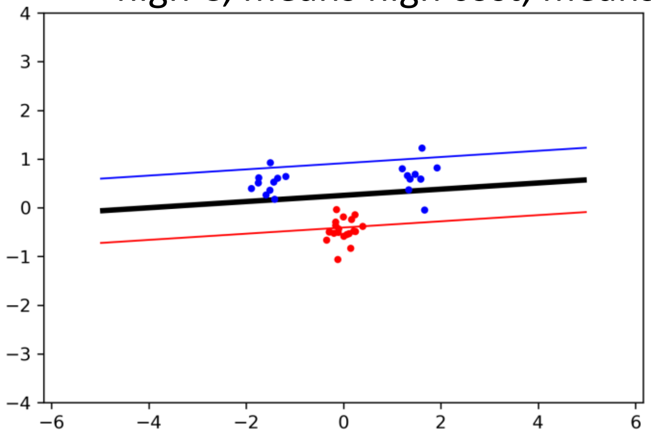


sigma=10

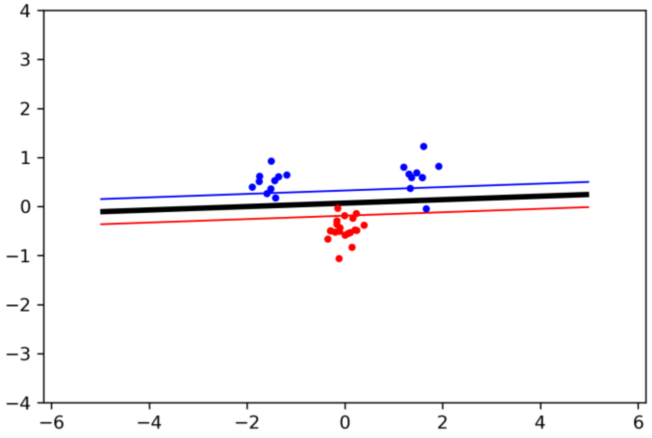


Assignment 4

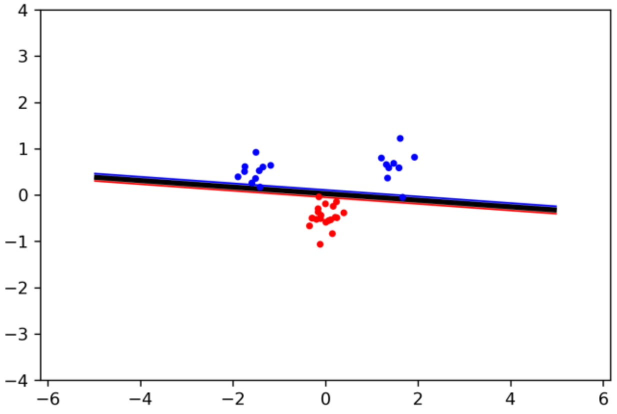
- Explore the role of the slack parameter C. What happens for very large/small values?
  - C: cost of misclassification
  - slack: how many data points are allowed to be misclassified
  - high C, means high cost, means low slack, could be overfitting



c=0.1



c=10



c=100

Assignment 5

- Imagine that you are given data that is not easily separable. When should you opt for more slack rather than going for a more complex model (kernel) and vice versa?

when the dataset is too noisy (or too many outliers), better to use a simple kernel. if we choose a complex kernel, then we need to low down the cost in order to allow more slack. therefore, overfitting could be avoided.