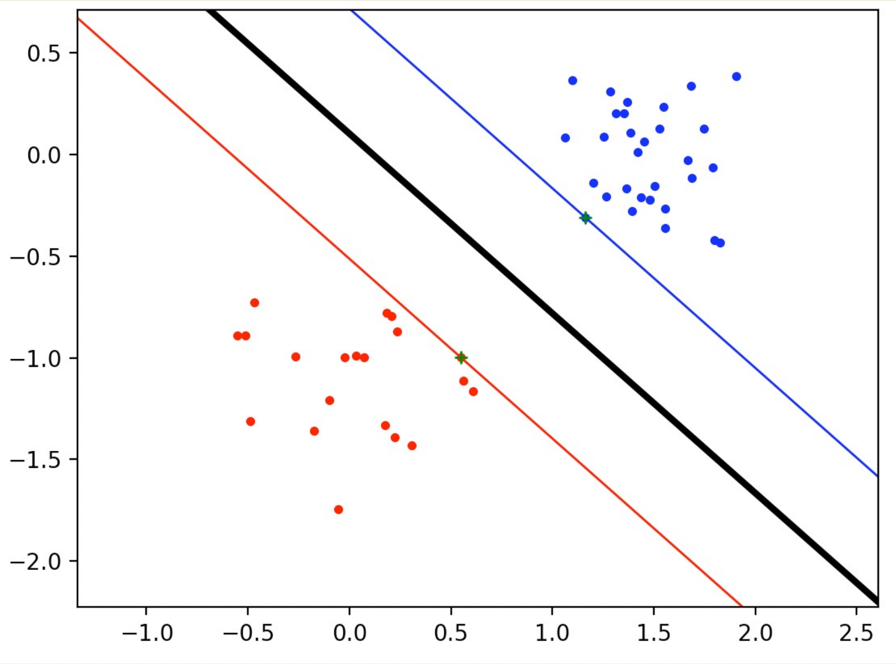
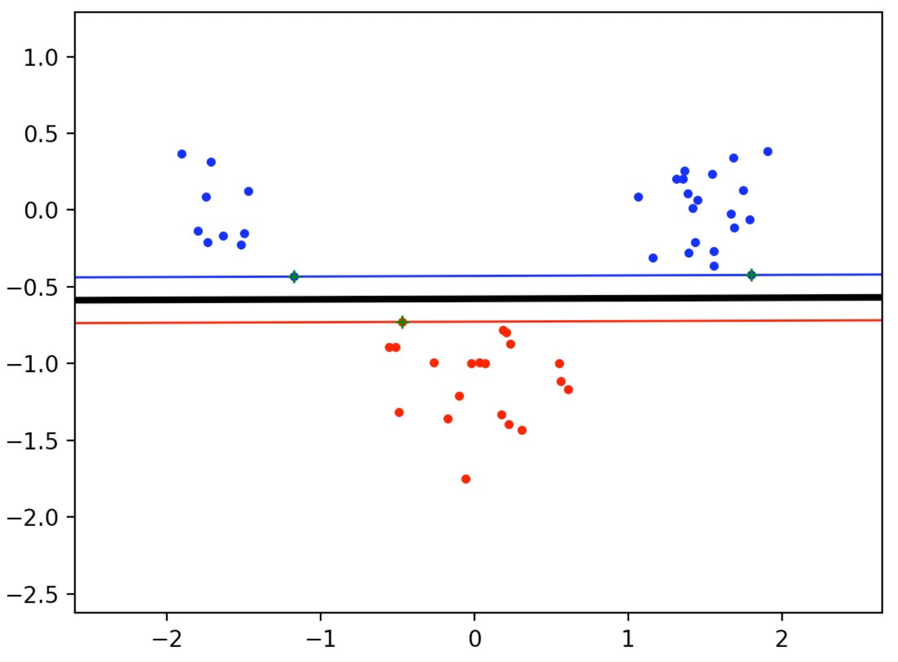
SUPPORT VECTOR MACHINES

Lab 02 – Machine Learning

Ivan & Simon

Linear kernel





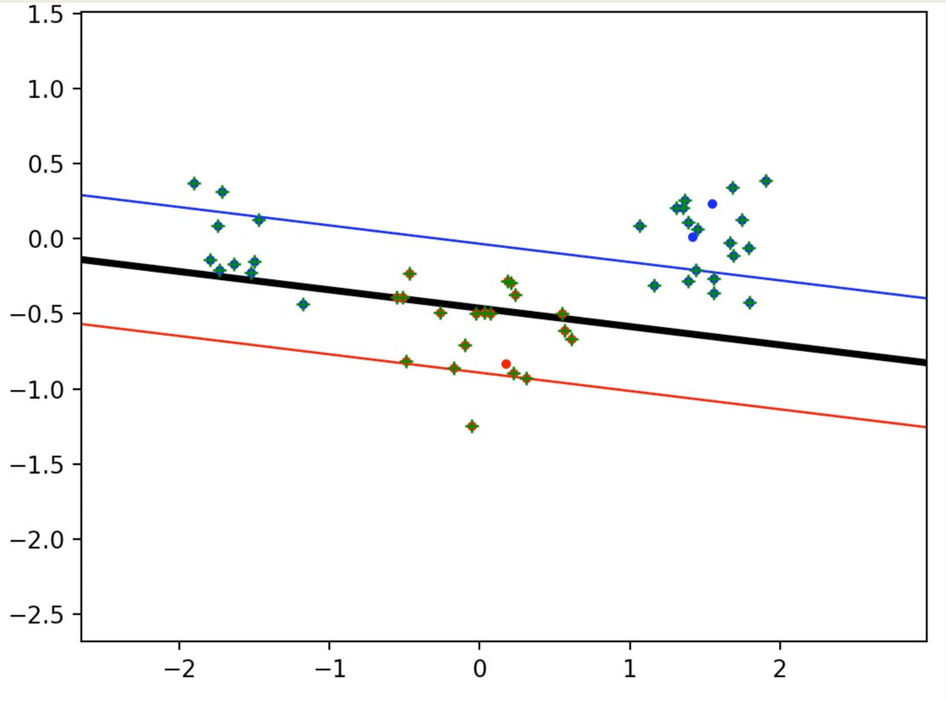
”easiest” case

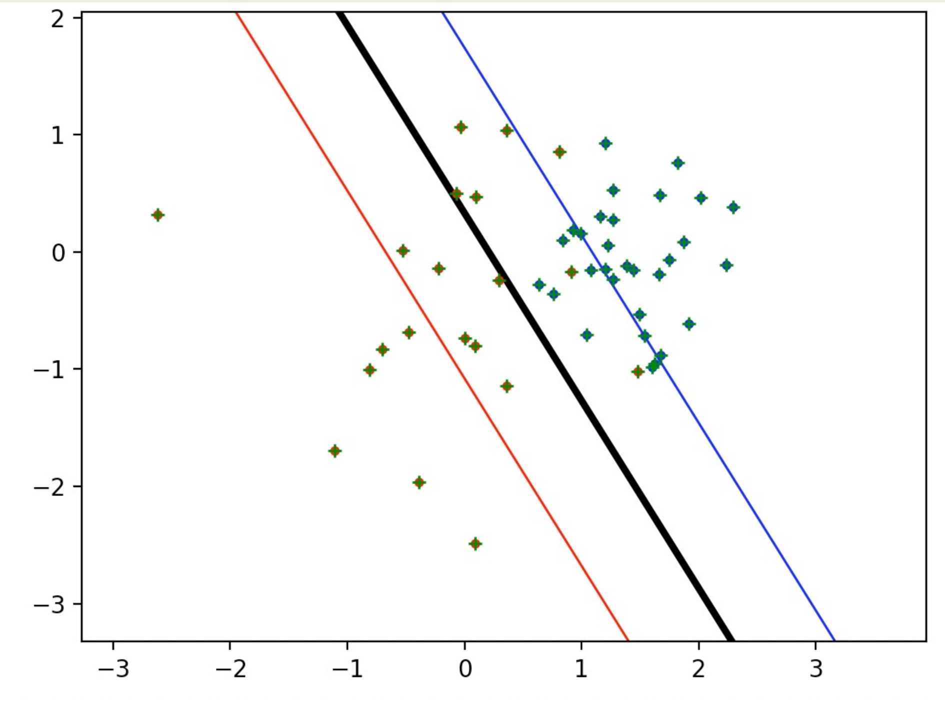
- two data clusters
- clearly linearly separable

Harder case

- three data clusters
- clusters reasonably close but not trivial

Linear kernel – problem cases





No overlap, but not linearly separable
- ➢ New kernel functions

Overlapping data
- ➢ Introduction of slack

Non-Linear Kernels

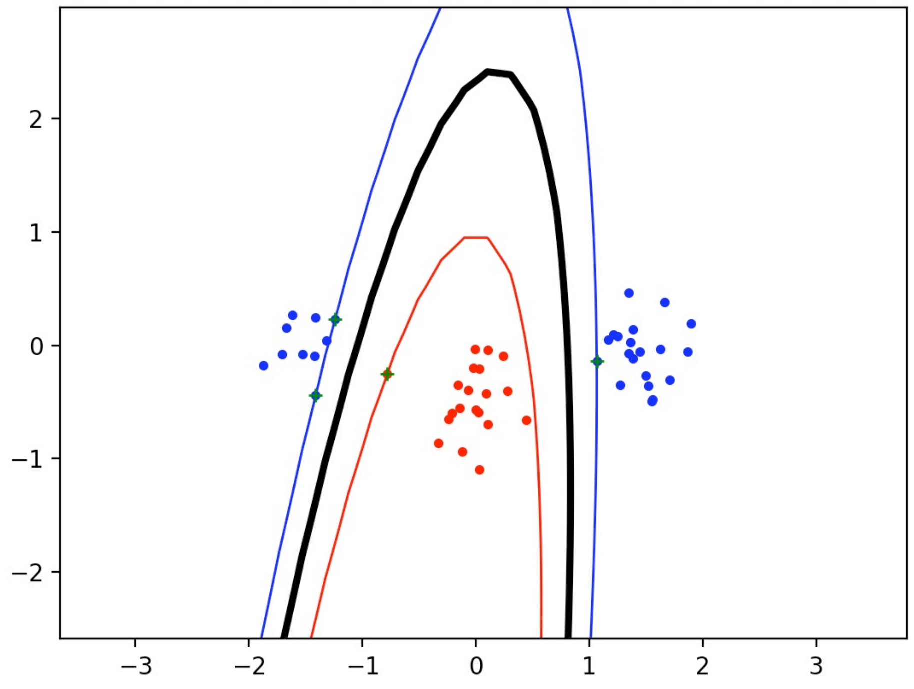
Polynomial Kernel

$$\mathcal{K}(\vec{x}, \vec{y}) = (\vec{x}^T \cdot \vec{y} + 1)^p$$

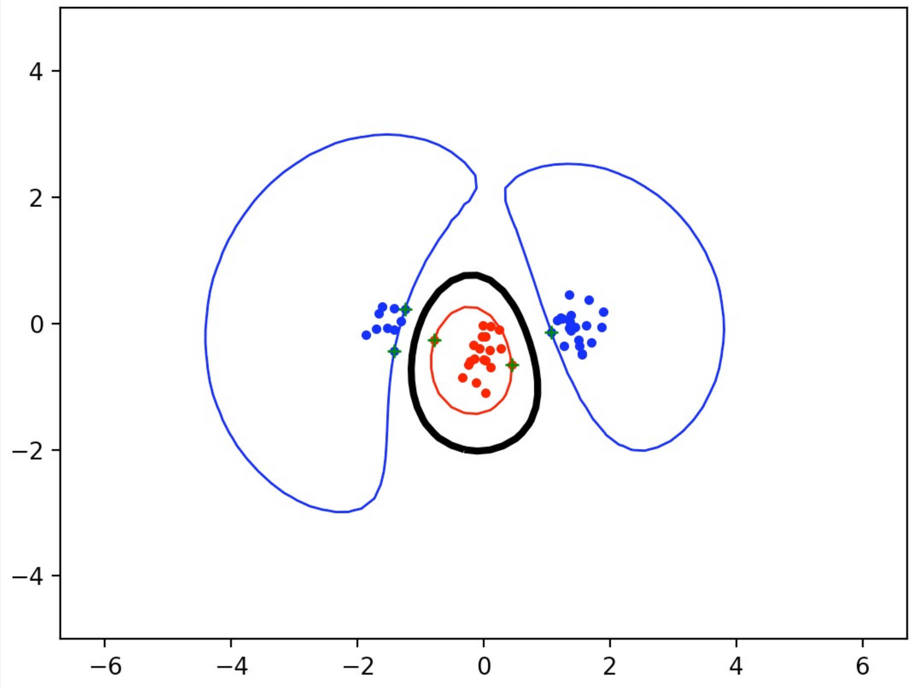
Radial Basis Function Kernel

$$\mathcal{K}(\vec{x}, \vec{y}) = e^{-\frac{||\vec{x} - \vec{y}||^2}{2\sigma^2}}$$

Non-linear kernels

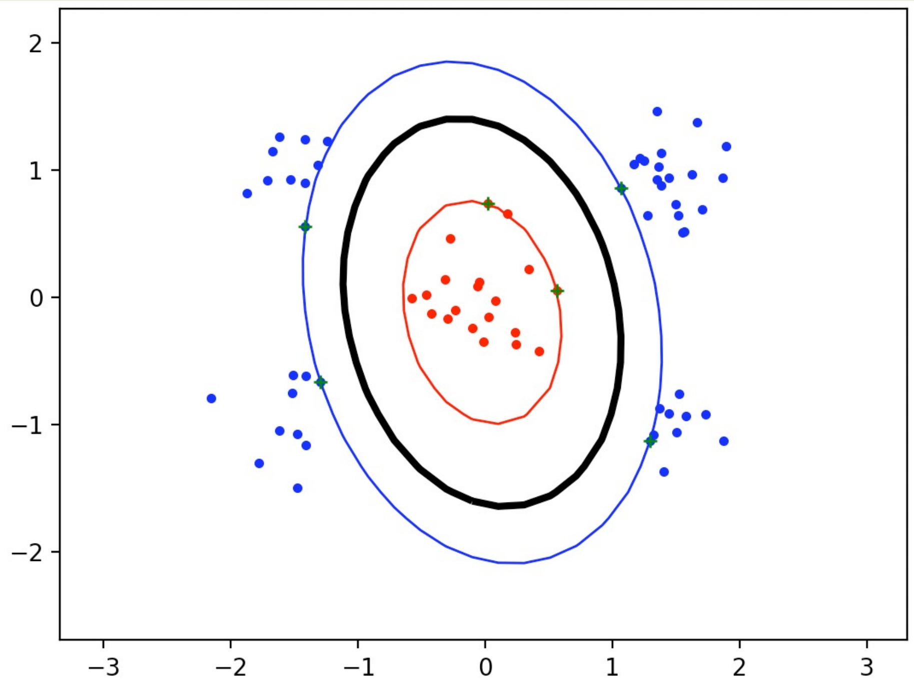


Polynomial kernel with order 2

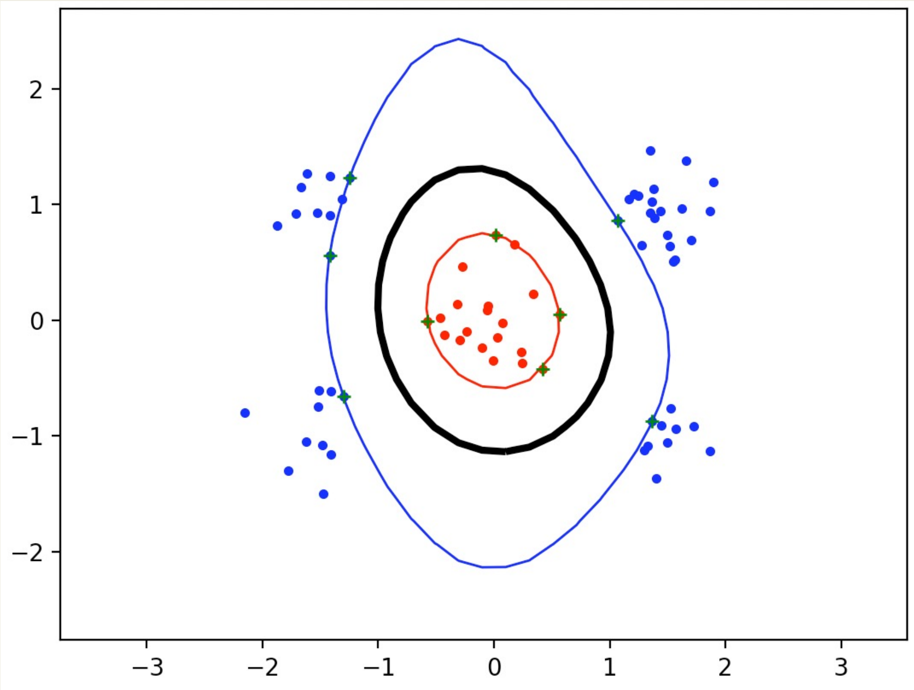


RBF kernel with sigma = 1.0

Non-linear kernels

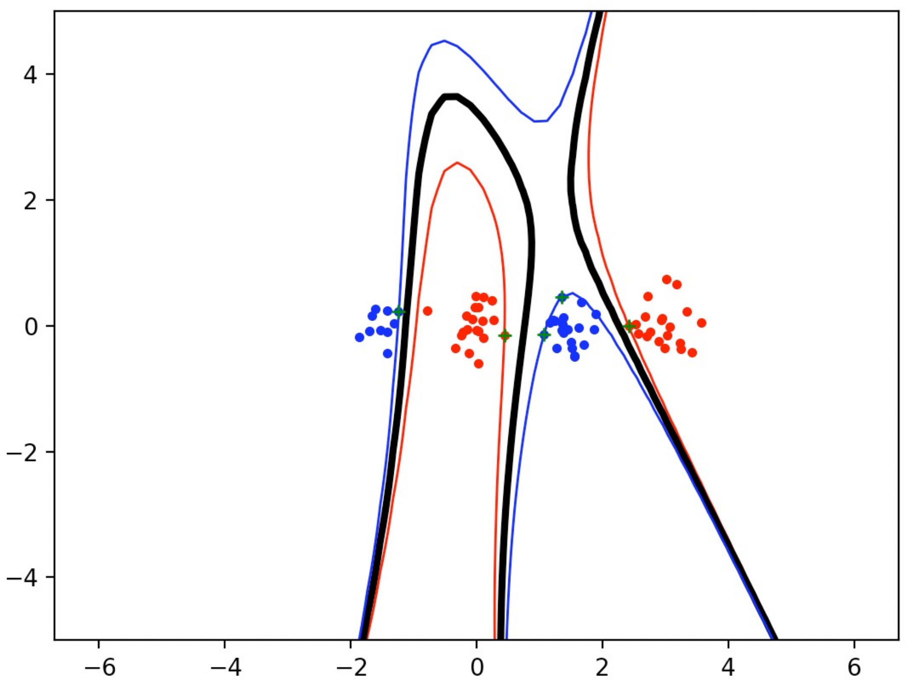


Polynomial kernel with order 2

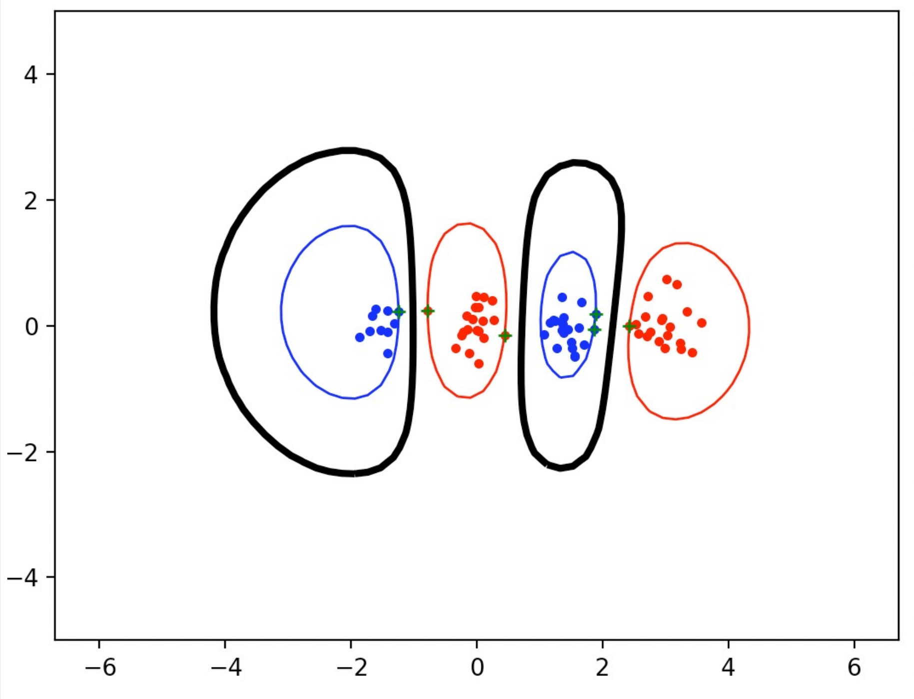


RBF kernel with sigma = 1.0

Non-linear kernels



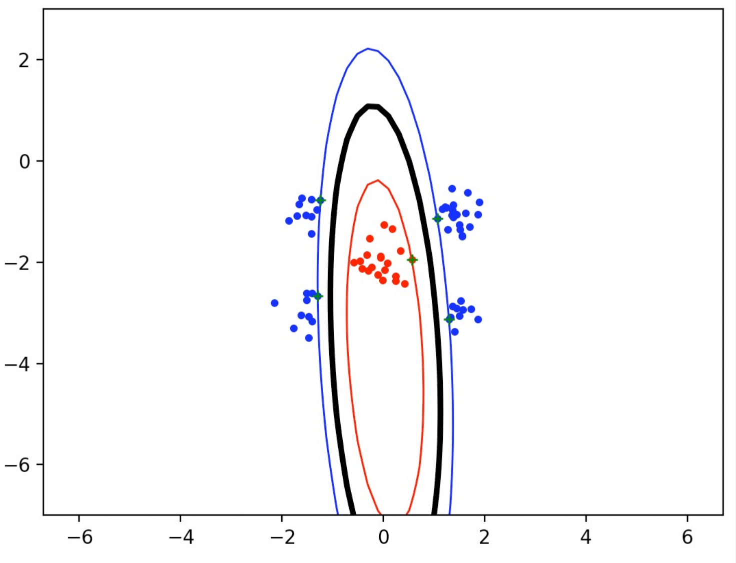
Polynomial kernel with order 3
- Order 2 was not “enough” to solve the problem



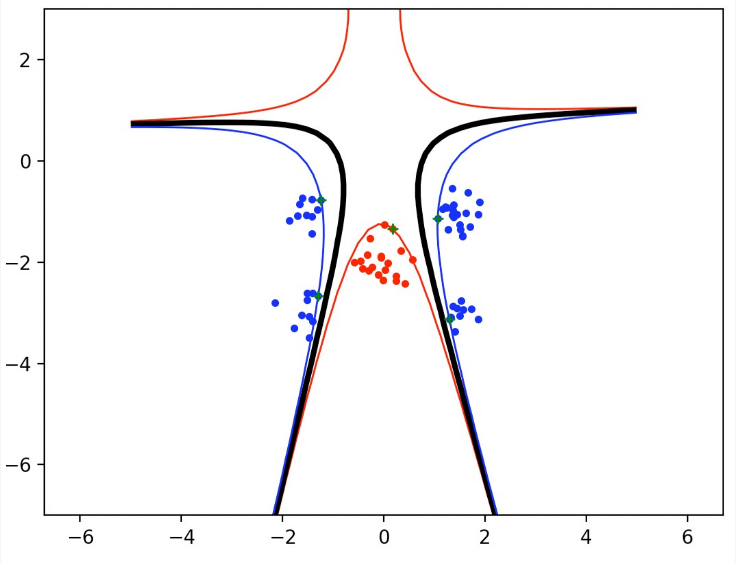
RBF kernel with sigma = 1.0

Bias-variance tradeoff

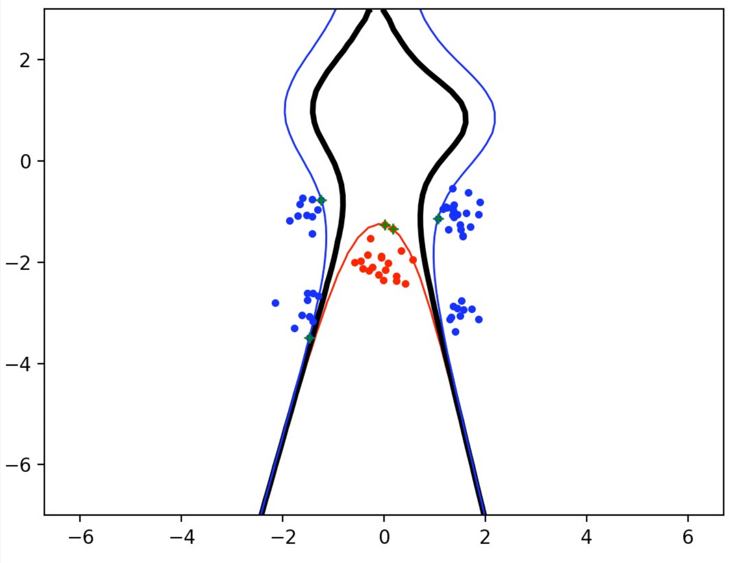
Polynomial kernel - order



2



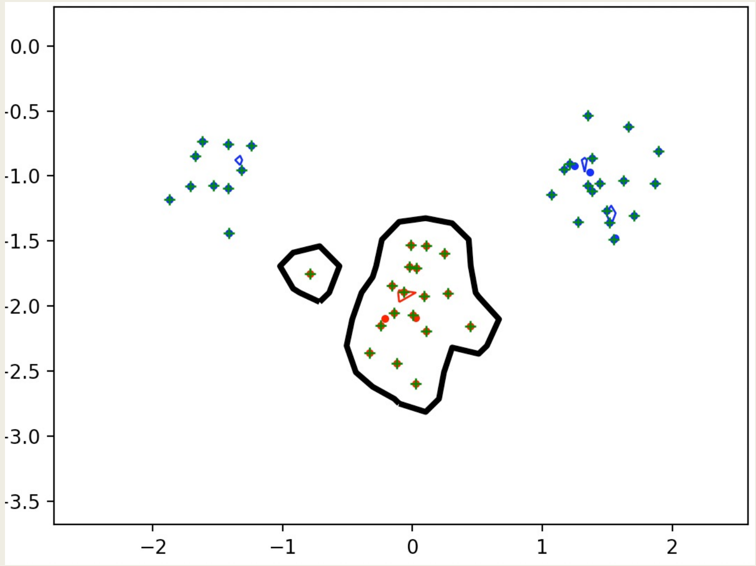
3



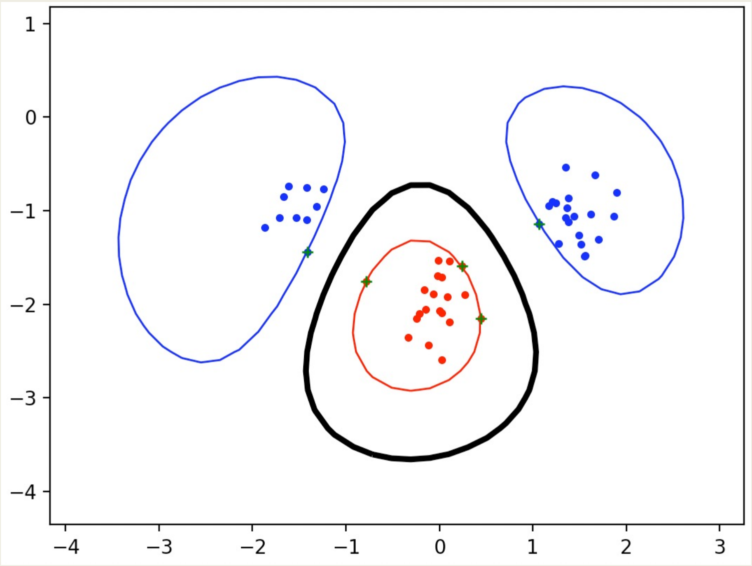
4

low variance, high bias ⟷ high variance, low bias

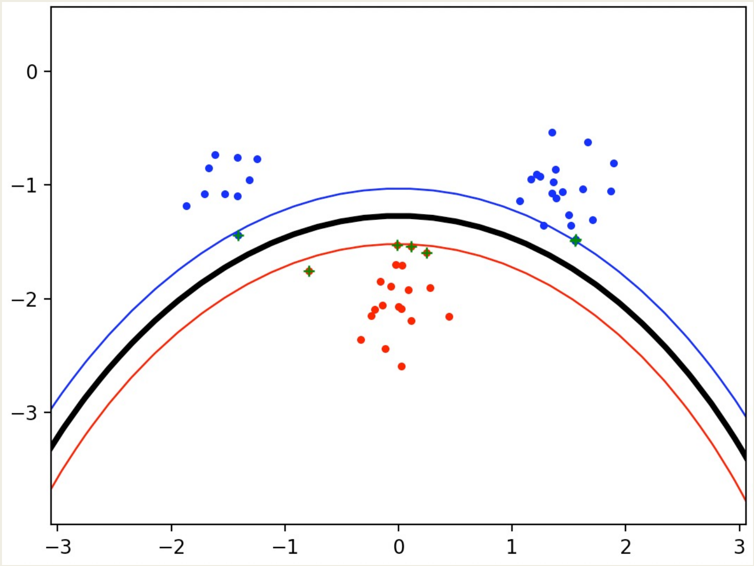
Bias-variance tradeoff
RBF kernel - sigma



.1



1



10

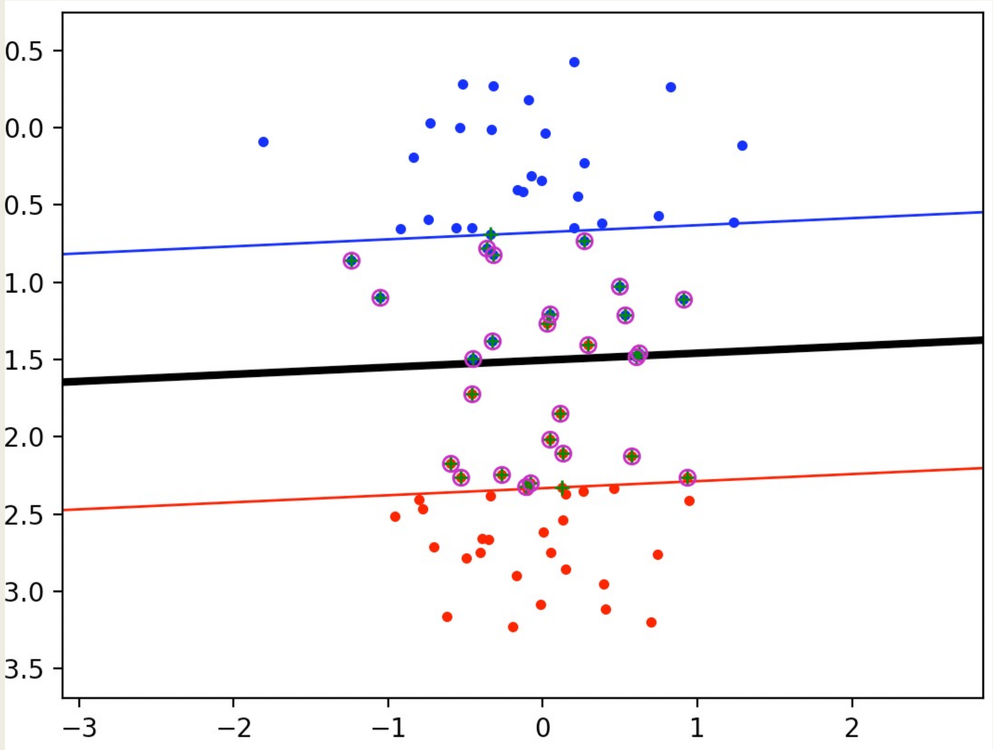
high variance, low bias — low variance, high bias

Slack variables

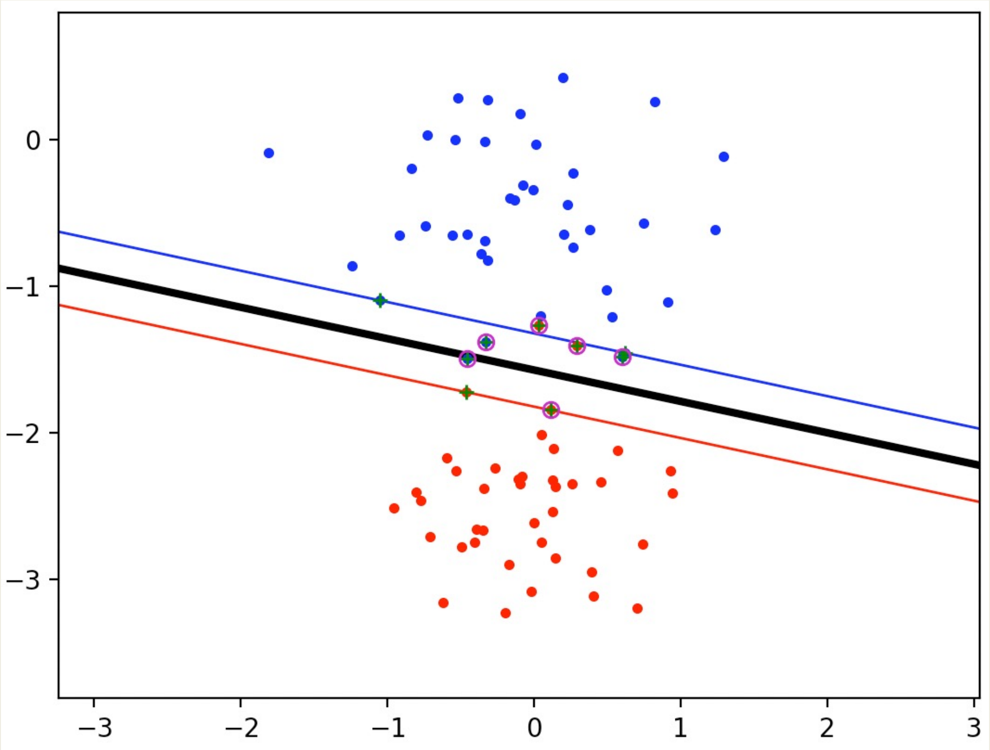
Weakening the constraints

$$t_i(\vec{w}^T \cdot \phi(\vec{x}_i) - b) \geq 1 - \xi_i$$

Influence of slack parameter C - noise



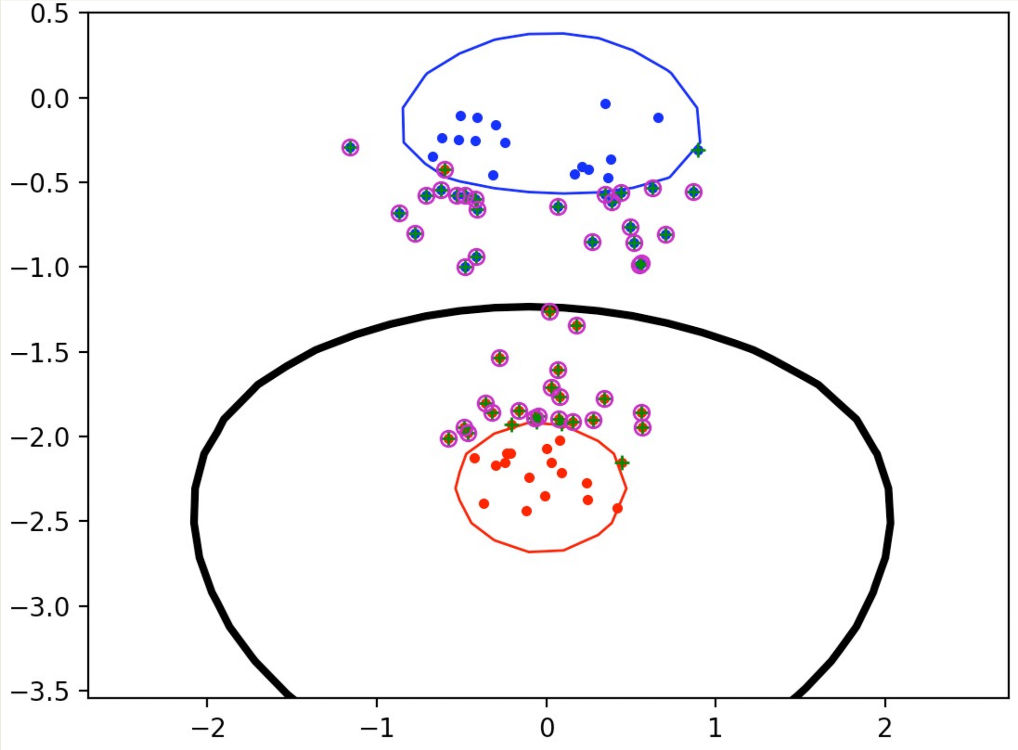
.1



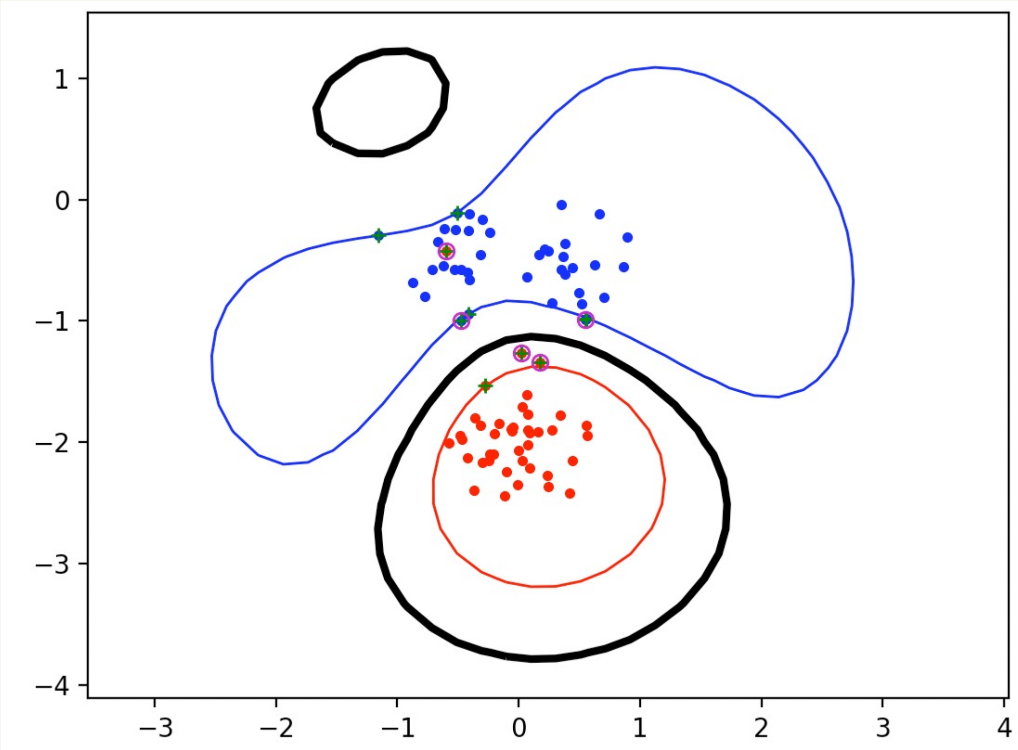
10

C controls how much we penalize points lying inside of the margin

Influence of slack parameter C – wrong label



.1



10

C controls how much we penalize points lying inside of the margin

What to do with difficult data?
More slack or complex model?

High noise? → More slack!

Otherwise overfitting

Complicated data? → Complex model!

Otherwise underfitting