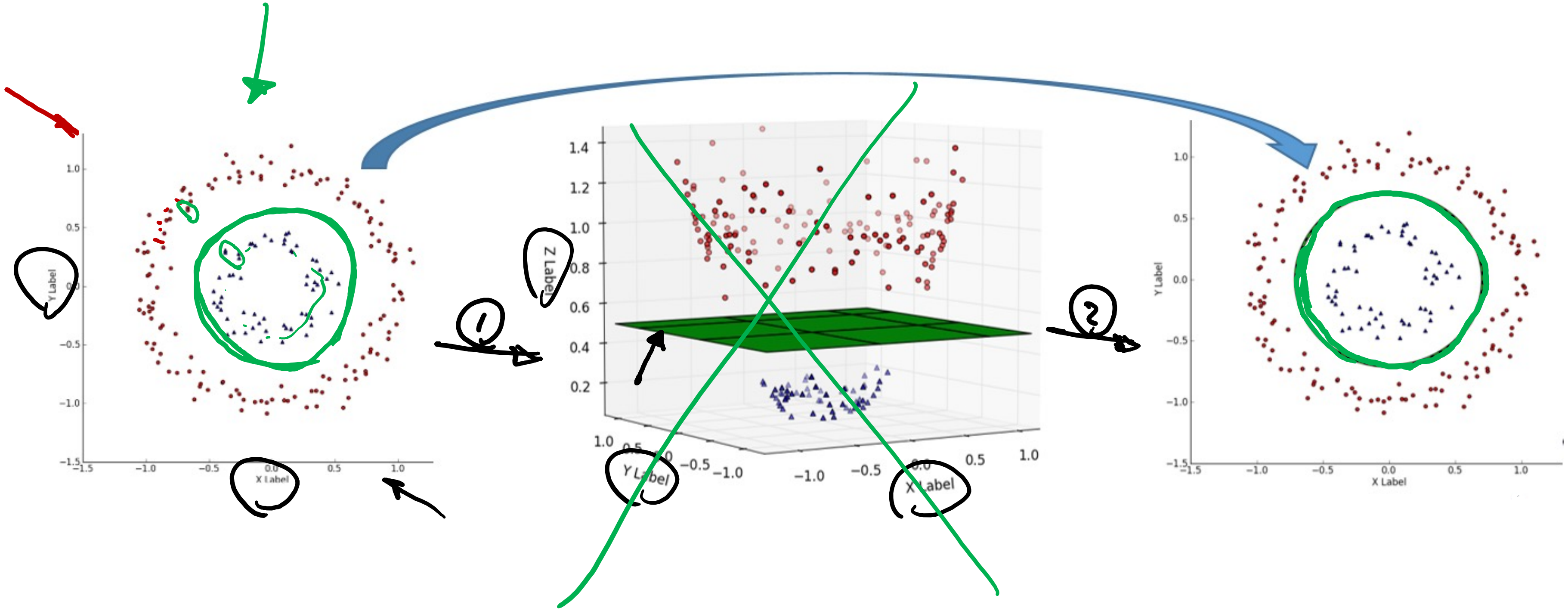
Part 25- Support Vector Machines
The Kernel Trick

Prof. Pedram Jahangiry



Topics

Part 23

- SVM Geometry
- SVM Motivation

Part 24

- Maximum Margin Classifier (MMC)
- Support Vector Classifiers (SVC)

Part 25

- Support Vector Machines (SVM)

Soft Mg + Kernel

Part 26

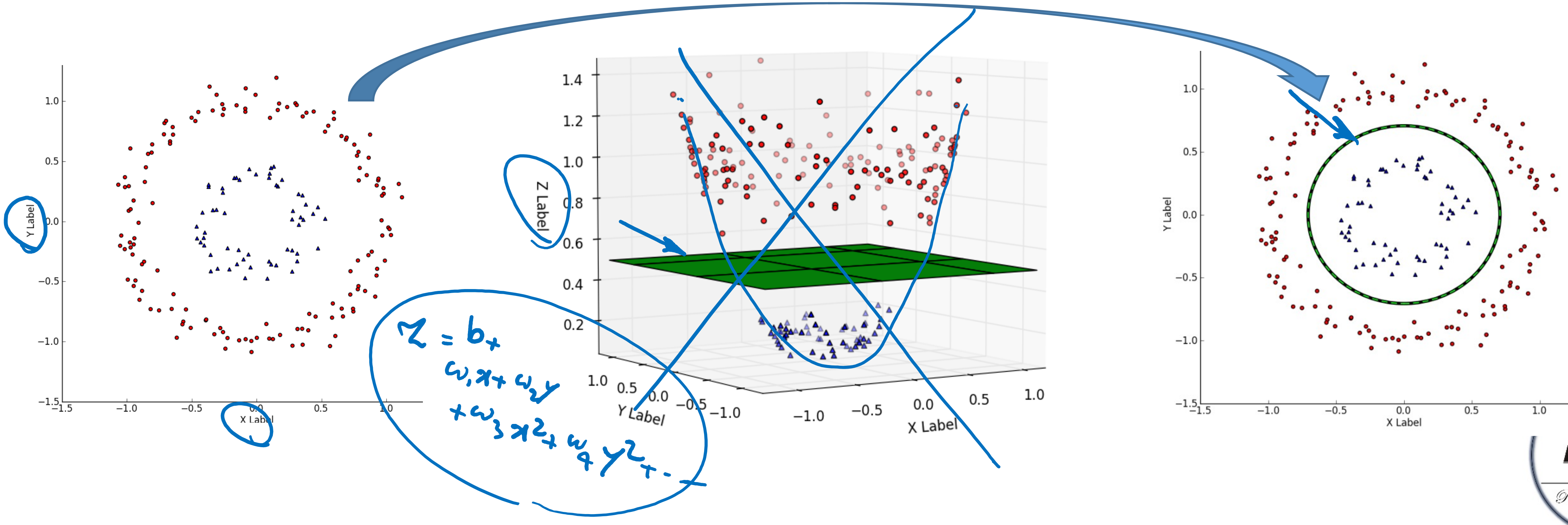
- Support Vector Regressors (SVR)

Part 27

- Multiple class classification
- SVM pros and cons
- SVM applications in Finance

Kernel Trick!

- Non-linearly separable data: sometime a linear boundary simply won't work, no matter what value of C.
- We need a non-linear decision boundary!
- Mapping to higher dimensional space, finding the hyper plane and projecting it back to low dimensional space can be computationally expensive.
- Solution: Kernel Trick!



Support Vector Machines (SVM)

- SVM generalizes the SVC to a nonlinear model, via the kernel ϕ which is applied to the input points $x_{i,k}$.
- The Kernel $\phi(x_{i,k})$ is a function that quantifies the similarities between observations by summarizes the relationship between every single pairs in the training set.

Hard Margin = MMC

SVC (Soft Margin) + linear Kernel = SVC

SVC + Non-linear Kernel = SVM

$$\underset{w,b}{\text{Min}} \ \frac{1}{2}||w||^2 + C\sum_{i=1}^{I}\xi_i$$

$$s.t. \ \ y_i\left(\sum_{k=1}^{K} w_k\phi(x_{i,k}) + b\right) \geq 1 - \xi_i, \qquad \xi_i \geq 0 \ \ \forall_i$$

Support Vector Machines (SVM)

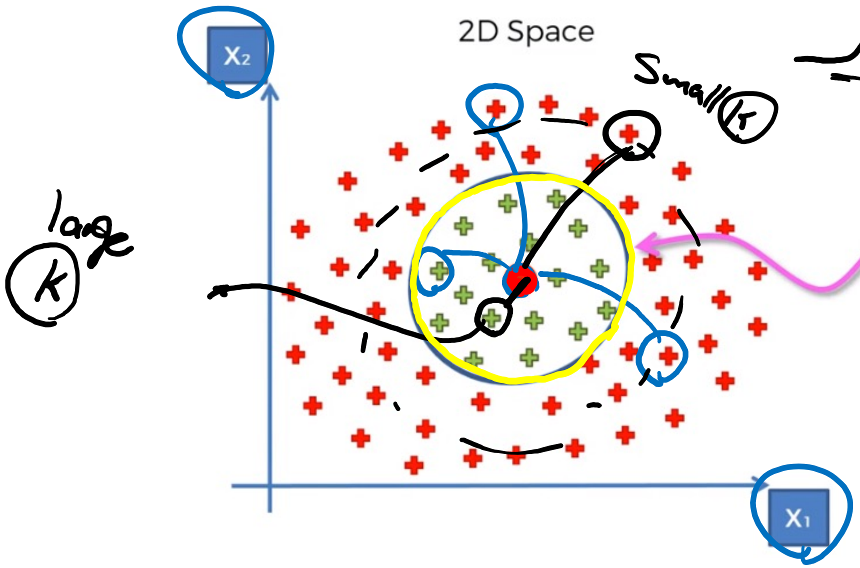
- SVM generalizes the SVC to a nonlinear model, via the kernel ϕ which is applied to the input points $x_{i,k}$.
- The Kernel $\phi\left(x_{i,k}\right)$ is a function that quantifies the similarities between observations by summarizes the relationship between every single pairs in the training set.

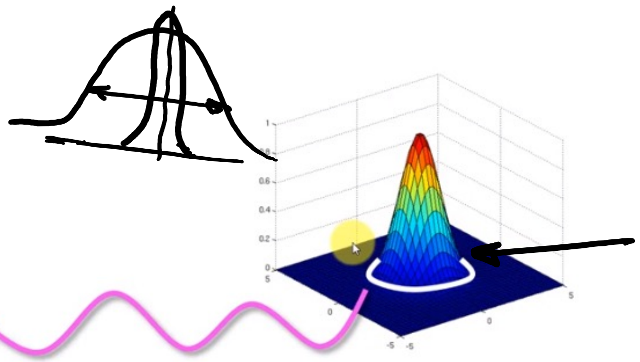
SVC + Non-linear Kernel = SVM

$$\underset{w,b}{\text{Min}} \ \frac{1}{2}\left|\left|w\right|\right|^2 + C\sum_{i=1}^{I}\xi_i$$

$$s.t.\ \ y_i\left(\sum_{k=1}^{K} w_k\phi(x_{i,k}) + b\right) \geq 1-\xi_i, \qquad \xi_i \geq 0\ \ \forall_i$$

The Gaussian RBF Kernel (Radial Basis Function)

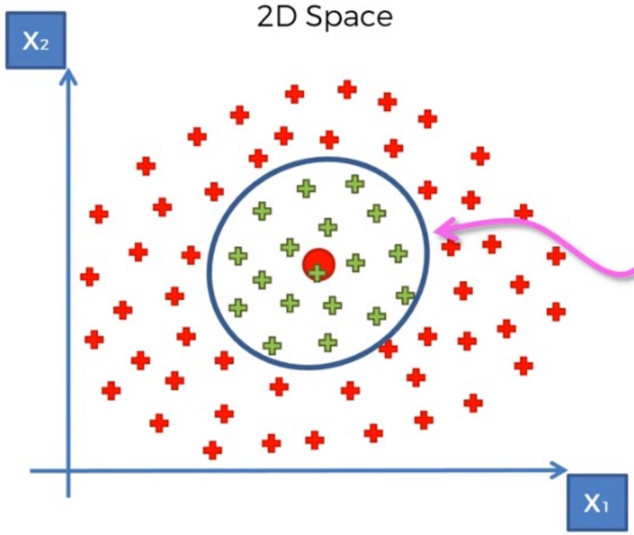


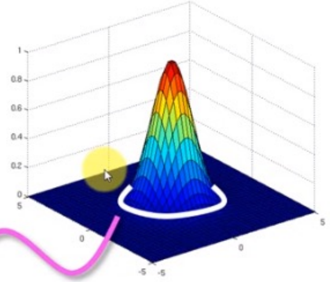


$\gamma \uparrow \quad \sigma \downarrow$

$$\boxed{K(\vec{x},\vec{l}^{\,i})} = e^{-\frac{\|\vec{x}-\vec{l}^{\,i}\|^2}{2\sigma^2}} \quad , \quad \gamma = \frac{1}{2\sigma^2}$$

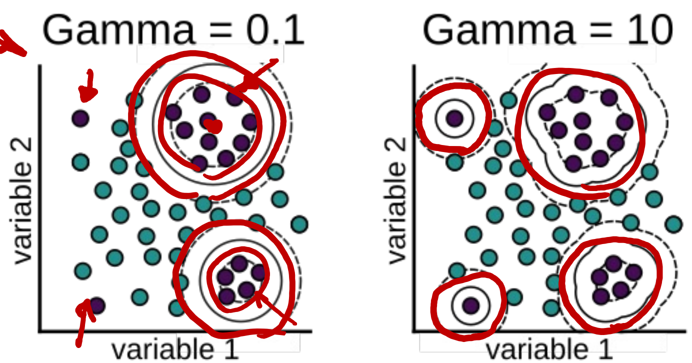
The Gaussian RBF Kernel (Radial Basis Function)





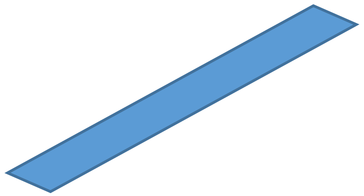
$$K(\vec{x}, \vec{l}^i) = e^{-\frac{\|\vec{x}-\vec{l}^i\|^2}{2\sigma^2}} \quad , \qquad \gamma = \frac{1}{2\sigma^2}$$

γ

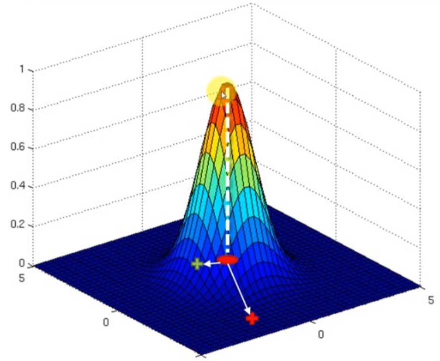


Most common types of Kernel

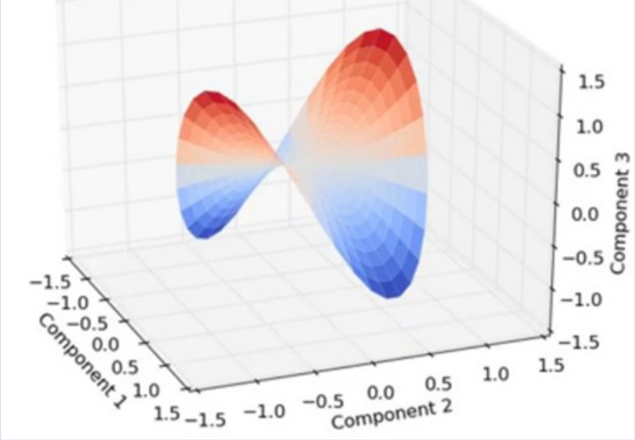
Linear Kernel



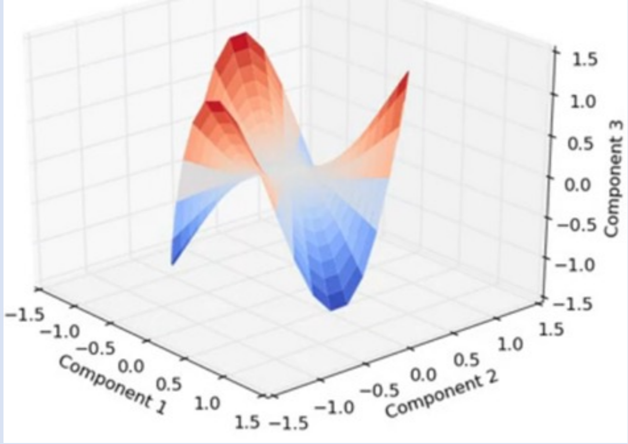
The Gaussian RBF Kernel



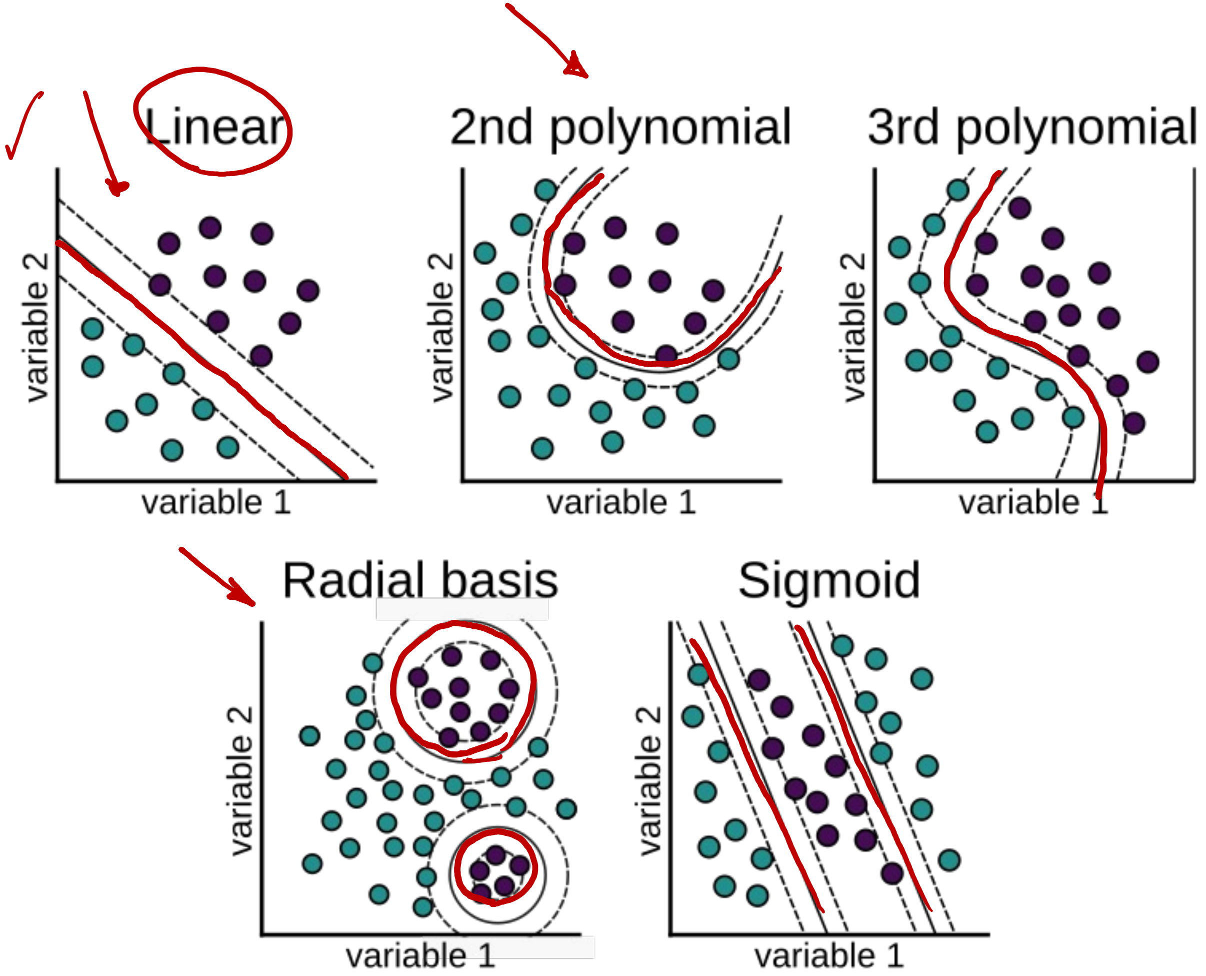
Polynomial Kernel



Sigmoid Kernel



Decision boundaries with different Kernels

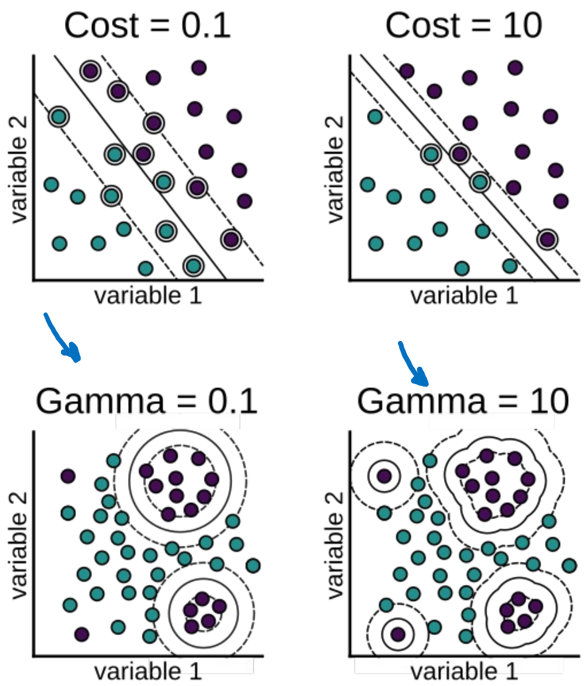


Source: [Machine learning with R, tidyverse, and mlr]

Tuning hyperparameters

SVM hyperparameters:

1) C, Cost of misclassification: controls bias variance trade off
2) Kernel
3) Gamma, controls how far the influence of a single training set reaches



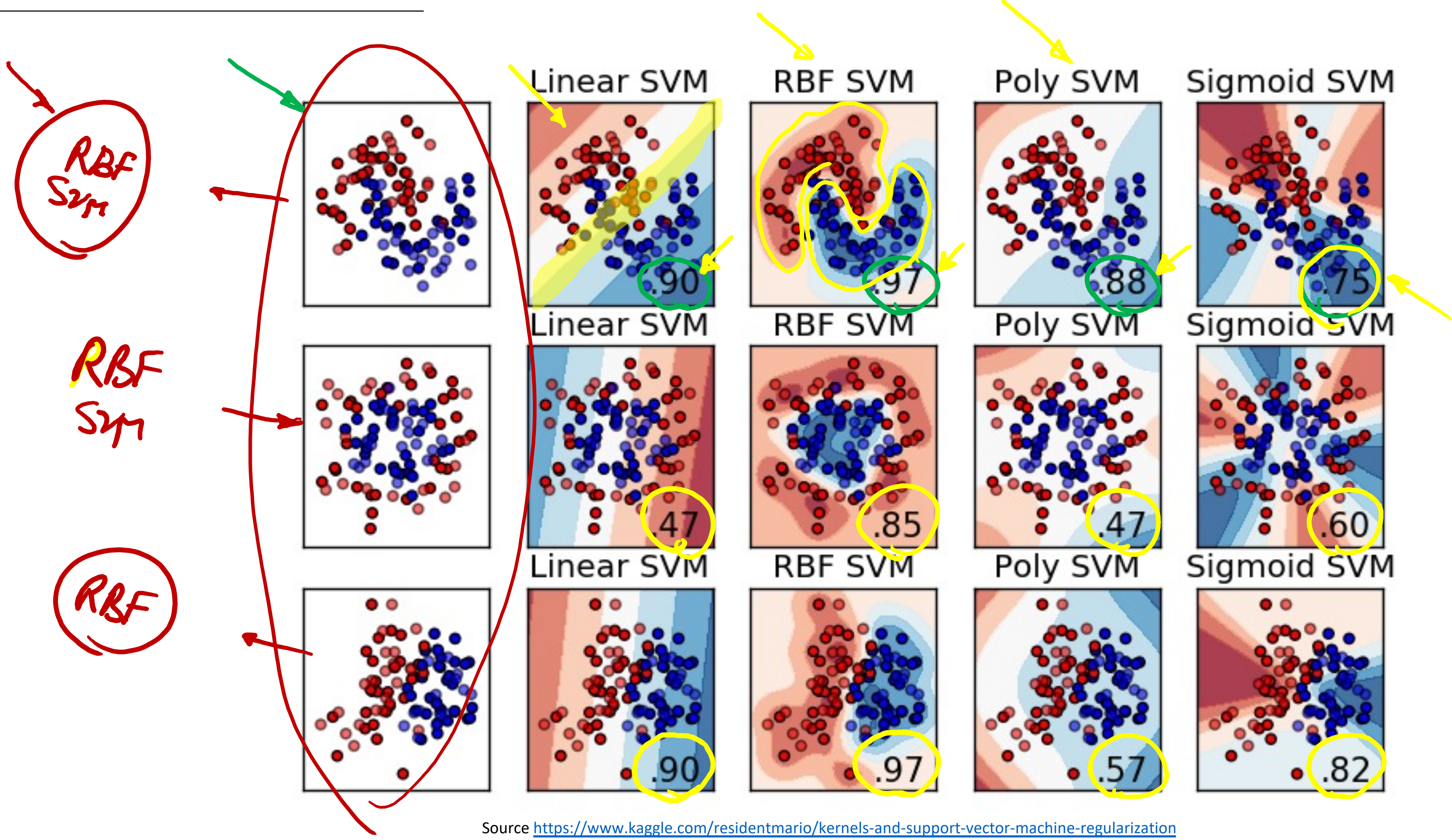
Source: [Machine learning with R, tidyverse, and mlr]

Grid search cross validation is used to tune the hyper parameters.

Kernel	C	Gamma	CV
Linear, rbf, poly, ….	0.1, 1, 10, 100, …	0.001, 0.01, 0.1, 1, …	5,10,…

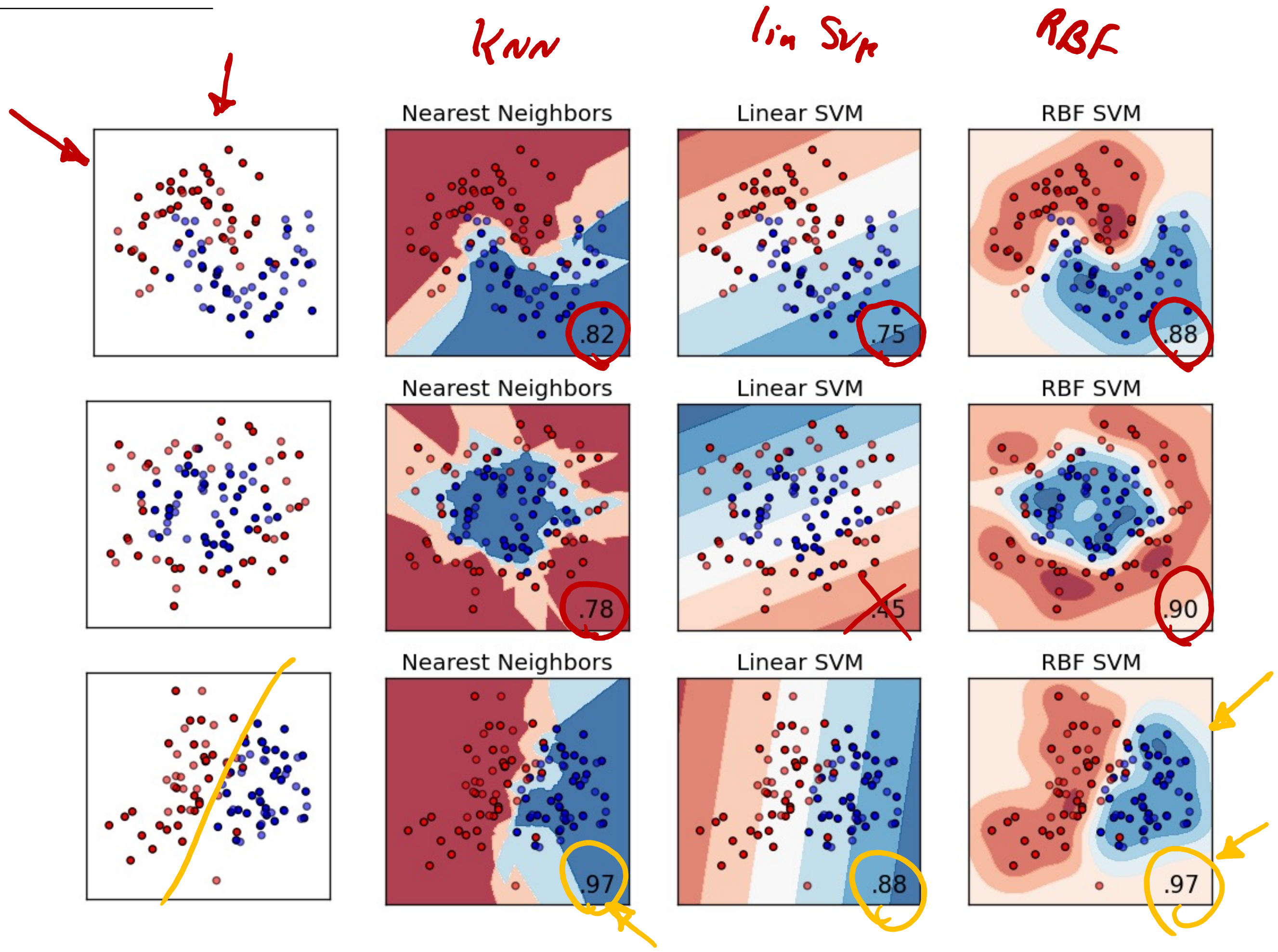
3 × 4 × 4 × 5 = 240

Decision boundaries with different Kernels



Source https://www.kaggle.com/residentmario/kernels-and-support-vector-machine-regularization

Comparing classifiers (so far)



Source: https://scikit-learn.org/stable/auto_examples/classification/plot_classifier_comparison.html