**Obsah**

[1. Regressions – Polynomial Linear Regression 2](#_Toc4353464)

[1.1. SVM – Support-vector machine 2](#_Toc4353465)

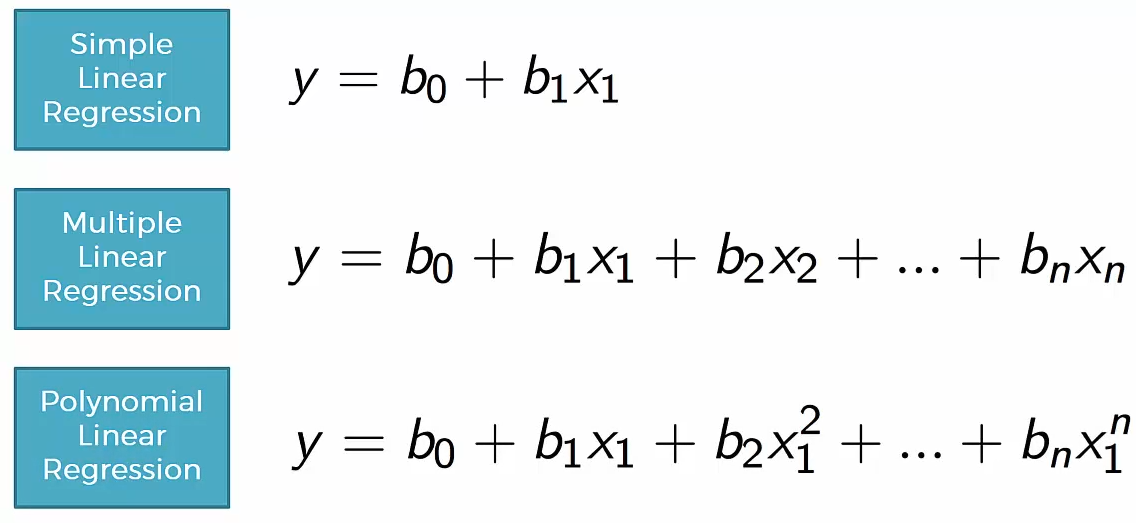
[1.2. Probabilistic classification 2](#_Toc4353466)

[1.3. Ensemble learning 2](#_Toc4353467)

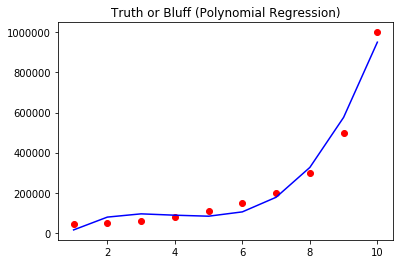
[1.4. Feature scaling 3](#_Toc4353468)

[2. Random Forest 4](#_Toc4353469)

1. Regressions – Polynomial Linear Regression



Linear refers to the coefficient. Non-linear would be e.g. y = (b0 + b1x1)/b2x2. A version of MLR.



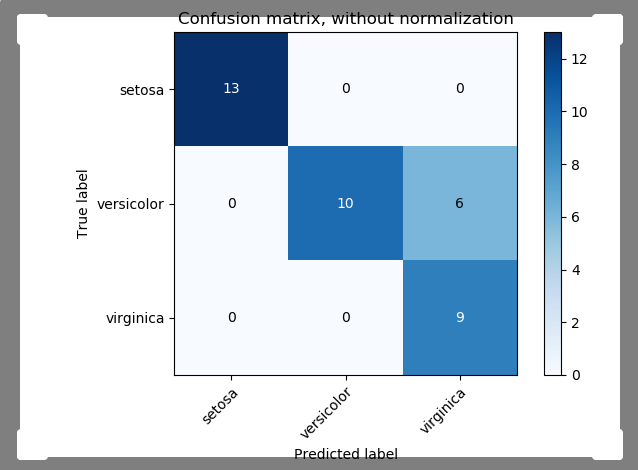
* 1. SVM – Support-vector machine

Types of kernels:

* Linear
* Poly – polynomial
* Rbf – Gaussian
* Sigmoid
* Precomputed

“All models are wrong, but some are useful”

* + 1. Confusion Matrix



Example of confusion matrix usage to evaluate the quality of the output of a classifier on the iris data set. The **diagonal elements** represent the number of points for which the predicted label is equal to the **true** label, while **off-diagonal elements** are those that are **mislabeled** by the classifier. The higher the diagonal values of the confusion matrix the better, indicating many correct predictions.

* 1. Probabilistic classification

In machine learning, a probabilistic classifier is a classifier that is able to predict, given an observation of an input, a probability distribution over a set of classes, rather than only outputting the most likely class that the observation should belong to. Probabilistic classifiers provide classification that can be useful in its own right or when combining classifiers into ensembles.

* 1. Ensemble learning

In statistics and machine learning, ensemble methods use multiple learning algorithms to obtain better predictive performance than could be obtained from any of the constituent learning algorithms alone. Unlike a statistical ensemble in statistical mechanics, which is usually infinite, a machine learning ensemble consists of only a concrete finite set of alternative models, but typically allows for much more flexible structure to exist among those alternatives.

* 1. Feature scaling

Feature scaling is a method used to standardize the range of independent variables or features of data. In data processing, it is also known as data normalization and is generally performed during the data preprocessing step.

1. Random Forest

* Use of multiple trees reduce the risk of overfitting (means we have fit the data so close what we have in our sample that we pick up on all the weird parts and instead of predicting the overall data you’re predicting the weird staff)
* Training time is less
* Runs efficiently on large database
* For large data it produces highly accurate predictions
* Random Forest **can maintain accuracy when a large proportion of data is missing**

Operates by constructing multiple Decision Trees during training phase. The Decision of majority of the trees is chosen by the random forest as the final decision.

1. Entropy – measure of randomness or unpredictability
2. Information gain – measure of decrease in entropy in entropy after the dataset is split
3. Leaf Node – carries the classification or the decision
4. Decision Node – has 2 or more branches
5. Root Node – the top most Decision Node

