Machine Learning & Data Mining @ NuIEEE

Tcharammmmmm - Final session

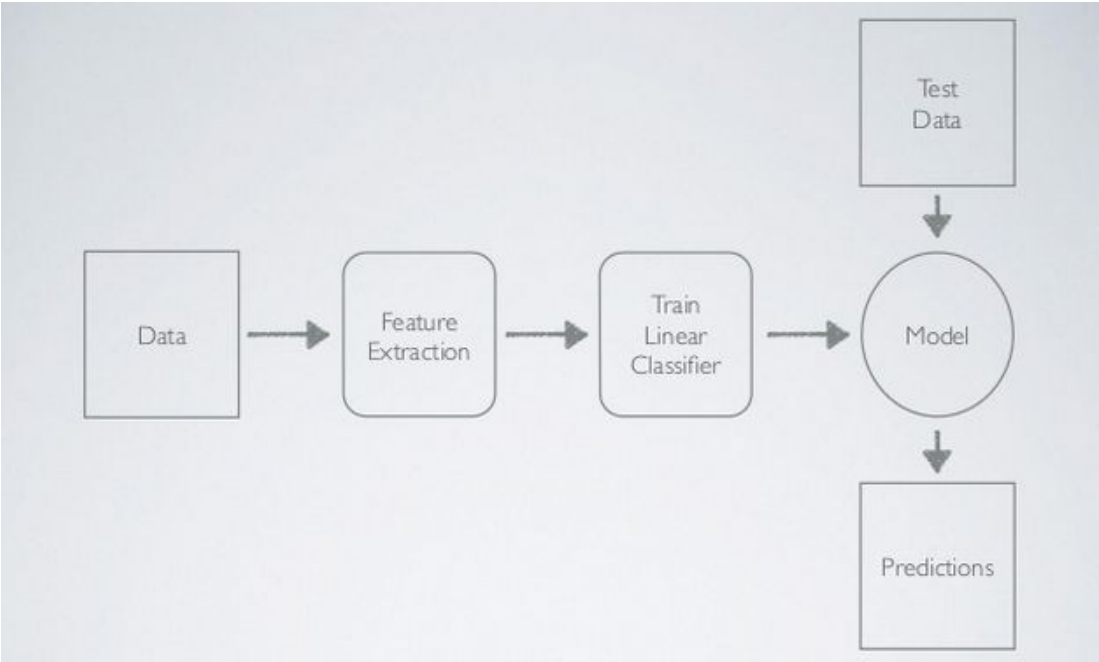
Miguel Sandim & Paula Fortuna

08-03-2016

Final session topics

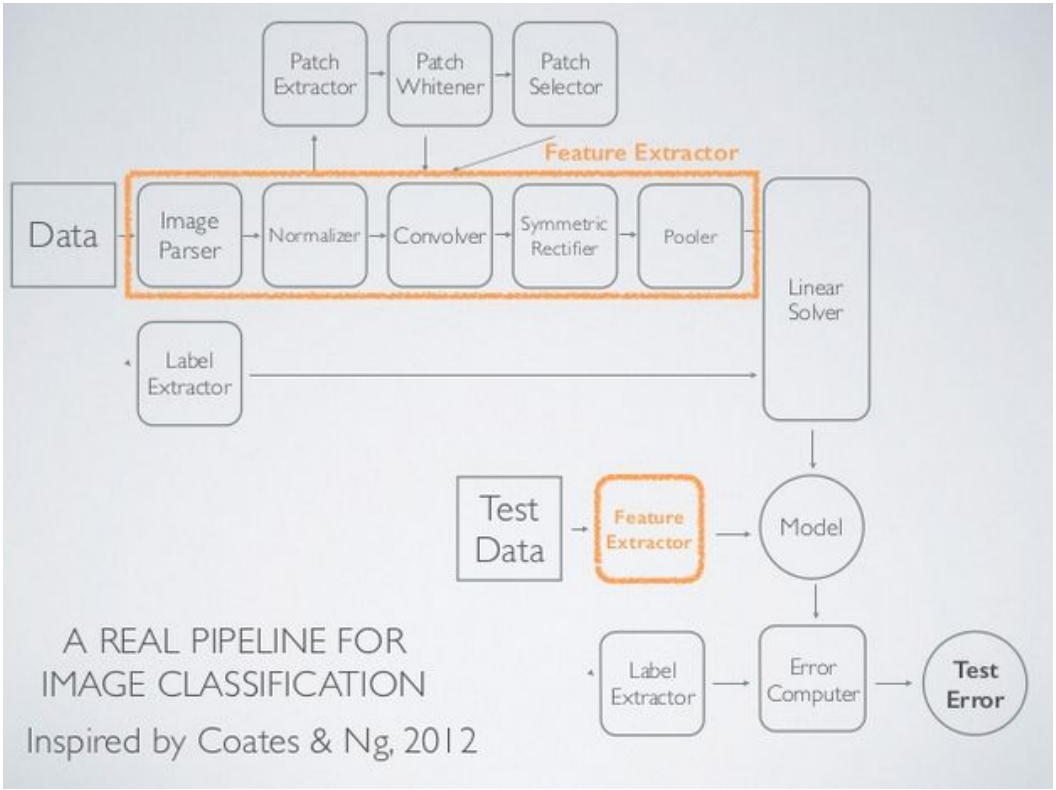
- ML pipeline
- How to choose a family of algorithms?
- How to evaluate if an algorithm is working
- Overfitting & underfitting
- Validation Methods - K-folds

Machine Learning Pipeline



[1]

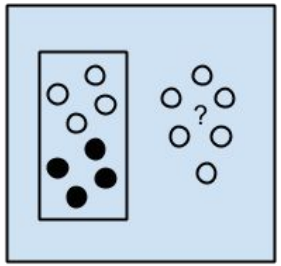
Machine Learning Pipeline



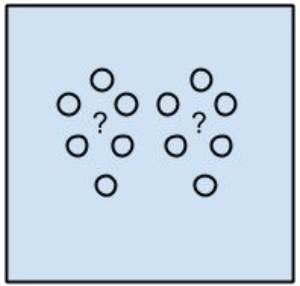
[1]

Machine Learning Model - Algorithms

Learning Style



Supervised Learning Algorithms

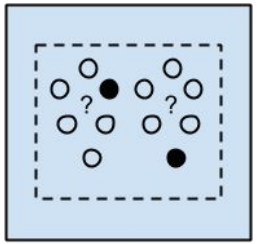


Unsupervised Learning Algorithms

[2]

Machine Learning Model - Algorithms

Learning Style

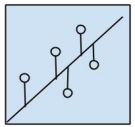


Semi-supervised
Learning Algorithms

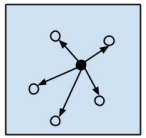
[2]

Algorithms Grouped By Similarity and examples

Regression Algorithms
- Linear Regression

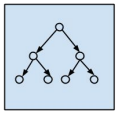

Regression Algorithms

Instance-based Algorithms
- k-Nearest Neighbour (kNN)


Instance-based Algorithms

Regularization Algorithms
- Ridge Regression

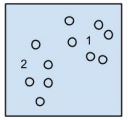
Decision Trees
- C4.5


Decision Tree Algorithms

Bayesian Algorithms
- Naive Bayes

Bayesian Algorithms

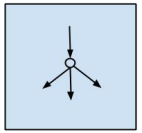
Clustering Algorithms
- K-Means



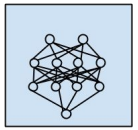
Clustering Algorithms

Association Rule Algorithms
- Apriori Algorithm

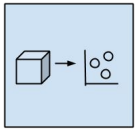
Neural Networks
- Back-propagation


Artificial Neural Network Algorithms

Deep Learning
- Deep Boltzmann Machine


Deep Learning Algorithms

Dimensionality Reduction
- PCA


Dimensional Reduction Algorithms

Emsemble Algorithms
- Random Forests

Feature Selection
- Wrapper Models

Other
- SVM

[2]


HELP ME STACKOVERFLOW...
...YOU ARE MY ONLY HOPE
Bayesian
Naive Bayes
Averaged One-Dependence Estimators (AODE)
Bayesian Belief Network (BBN)
Gaussian Naive Bayes
Multinomial Naive Bayes
Bayesian Network (BN)
Decision Tree
Classification and Regression Tree (CART)
Iterative Dichotomiser 3 (ID3)
C4.5
C5.0
Chi-squared Automatic Interaction Detection (CHAID)
Decision Stump
Conditional Decision Trees
M5
Dimensionality Reduction
Principal Component Analysis (PCA)
Partial Least Squares Regression (PLSR
Sammon Mapping
Multidimensional Scaling (MDS)
Projection Pursuit
Principal Component Regression (PCR)
Partial Least Squares Discriminant Analysis
Mixture Discriminant Analysis (MDA)
Quadratic Discriminant Analysis (QDA)
Regularized Discriminant Analysis (RDA)
Flexible Discriminant Analysis (FDA)
Linear Discriminant Analysis (LDA)
Instance Based
k-Nearest Neighbour (kNN)
Learning Vector Quantization (LVQ)
Self-Organizing Map (SOM)
Locally Weighted Learning (LWL)
Clustering
k-Means
k-Medians
Expectation Maximization
Hierarchical Clustering
Deep Learning
Deep Boltzmann Machine (DBM)
Deep Belief Networks (DBN)
Convolutional Neural Network (CNN)
Stacked Auto-Encoders
Ensemble
Random Forest
Gradient Boosting Machines (GBM)
Boosting
Bootstrapped Aggregation (Bagging)
AdaBoost
Stacked Generalization (Blending)
Gradient Boosted Regression Trees (GBRT)
Neural Networks
Radial Basis Function Network (RBFN)
Perceptron
Back-Propagation
Hopfield Network
Regularization
Ridge Regression
Least Absolute Shrinkage and Selection Operator (LASSO)
Elastic Net
Least Angle Regression (LARS)
Rule System
Cubist
One Rule (OneR)
Zero Rule (ZeroR)
Repeated Incremental Pruning to Produce Error Reduction (RIPPER)
Regression
Linear Regression
Ordinary Least Squares Regression (OLSR)
Stepwise Regression
Multivariate Adaptive Regression Splines (MARS)
Locally Estimated Scatterplot Smoothing (LOESS)
Logistic Regression

[2]

How to choose the right algorithm?

- Use Past Experience To Choose An Algorithm
- Start with the simplest approach
- Visualize data
- Use Trial And Error To Choose An Algorithm
    - use diverse algorithms
    - tuning the algorithm parameters
    - using ensemble methods
- Spot Check Algorithms in R - process to automatically apply a set of algorithms [3]

Variance and Bias

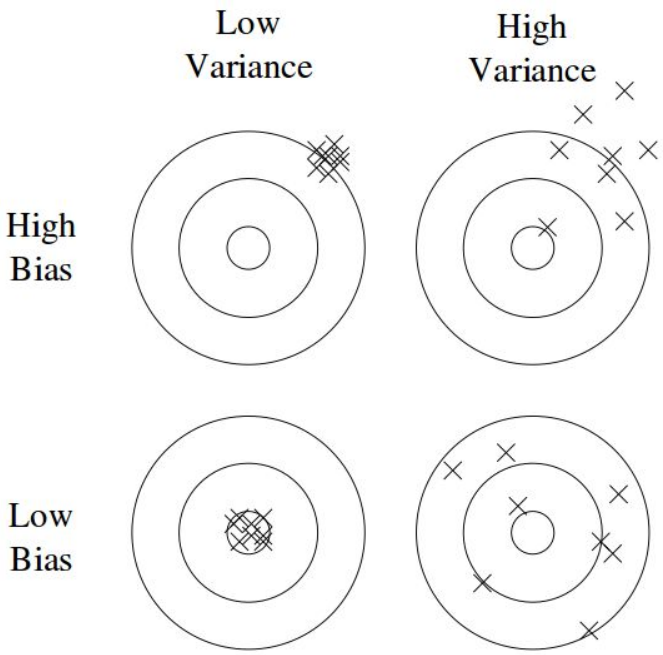
- ***Bias*** is a learner's tendency to consistently learn the wrong thing (difficulty in finding the "pattern" in the data);

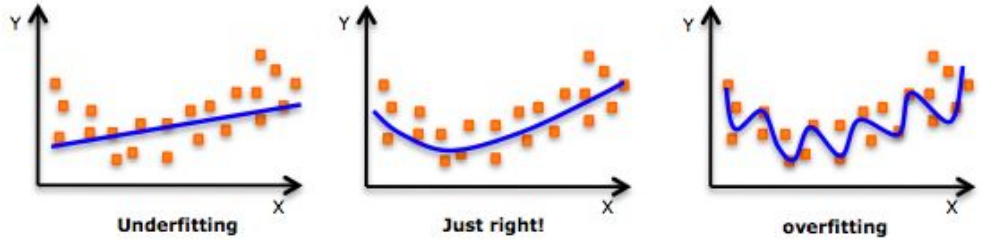
    - **Example**: approximating a real-world problem by a simple linear regression model may lead to high bias.

- ***Variance*** is a learner's tendency to learn random things irrespective of the "real signal" (learning everything is a "pattern", including the "noise" in the data). Another interpretation is the sensitivity of the model to changes in the train set.

    - **Example**: decision trees learned on different training sets generated by the same phenomenon are often very different, when in fact they should be the same.

Overfitting and Underfitting



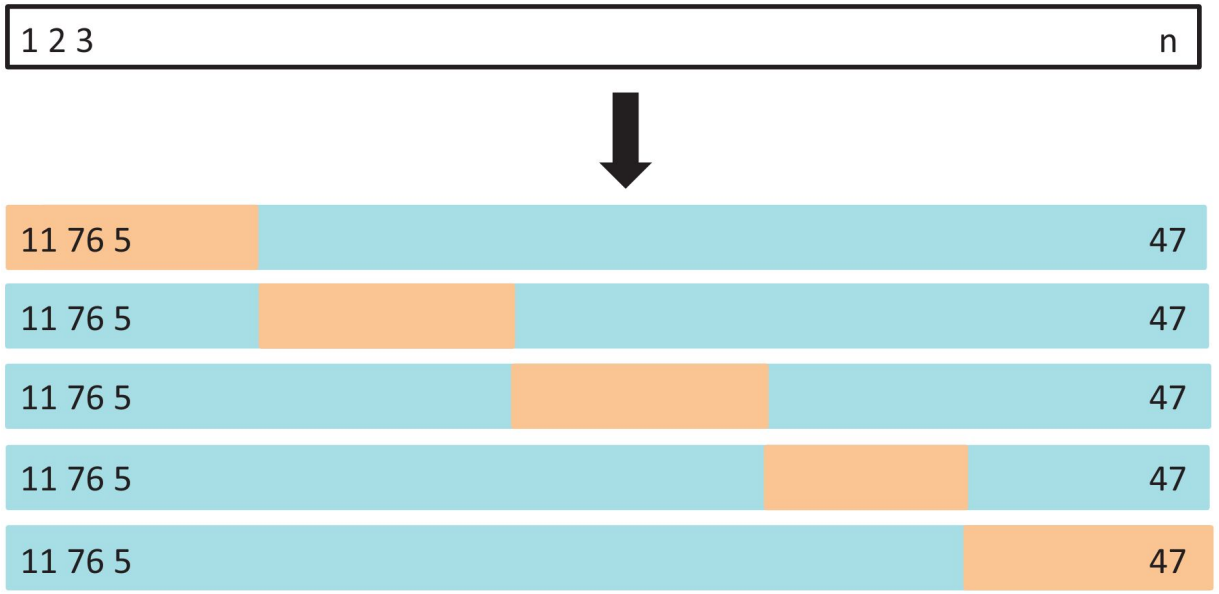


Overfitting is the major problem!

- How to solve?
  - **Regularization** - include an additional term that penalizes highly complex models;

  - **Cross-Validation** - choose the best model's parameters based on the error rate on a validation set. This set should be different from the training set (used to train the model) and the test set (used to assess the final model's accuracy).
    - Leave-One-Out Cross-Validation
    - k-Fold Cross-Validation

k-Fold Cross Validation



$$\mathrm{CV}_{(k)} = \frac{1}{k}\sum_{i=1}^{k}\mathrm{MSE}_i$$

References

[1] http://www.slideshare.net/jeykottalam/pipelines-ampcamp

[2] http://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/

[3]http://machinelearningmastery.com/evaluate-machine-learning-algorithms-with-r/

[4] Domingos, Pedro. "A few useful things to know about machine learning." *Communications of the ACM* 55.10 (2012): 78-87.

[5] James, Gareth, et al. An introduction to statistical learning. Vol. 112. New York: springer, 2013.