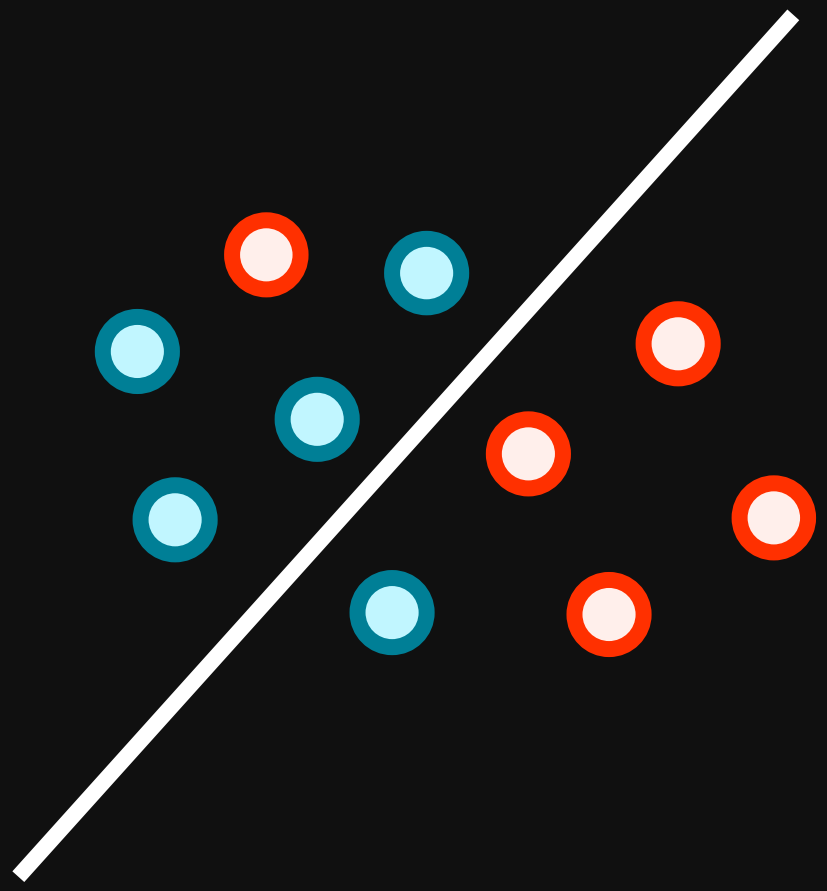
Basics of Machine Learning

Dmitry Ryabokon, github.com/dryabokon



Basics of Machine Learning

Summary

Summary

11. Non parametrical ML methods
12. Unsupervised Learning
13. Non-Bayesian approaches
14. Dimension reduction
15. Ensemble learning: RF
16. Ensemble learning: XGB, Adaboost
17. Benchmarking
18. Time Series
19. Introduction to Deep Learning
20. AE, VAE

Lesson 01: Intro

- Introduction
- Goals
- Overview of the course
- Prerequisites
- Data engineering vs Data Science vs Machine Leaning
- Overview of practical assignments
- Collaboration and feedback

https://www.mygreatlearning.com/blog/top-artificial-intelligence-interview-questions/?highlight=interview

Lesson 02: Environments

- Installing Python
- Virtual environments
- Using docker
- Jupyter notebook
- Google Collab
- IDEs
- Code repositories
- Datasets
- Example: vanilla classification problem

https://habr.com/ru/company/edison/blog/480408/

Overview

Lesson 03: Data Manipulation

Numpy

- Data inspection
- Combining data
- Insert and delete
- Reshape
- Slicing
- Sorting
- Aggregating

Pandas

- Creation
- Inspection
- Sorting
- Slicing
- Grouping

Overview

Lesson 04: Data Visualization

- Pairwise analysis
- Regression analysis
- Density chart
- Feature importance
- Confidence level
- Highcharts

https://www.machinelearningplus.com/plots/top-50-matplotlib-visualizations-the-master-plots-python/

Overview

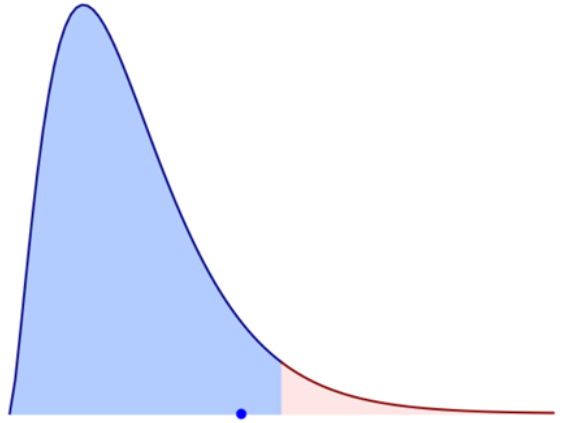
Lesson 05: Feature Engineering

- Handling Outliers
- Handling Missing Values
- Imputation
- Encoding
- Imbalanced dataset
- Sampling

Overview

Lesson 06: Statistical ML

- P Value and null hypothesis significance testing
- Chi-squared test statistical hypothesis test
- Consistency check: KS value

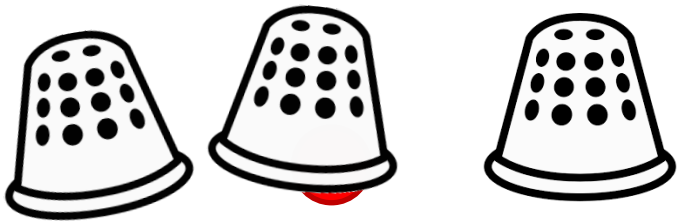


- Feature importance
- Numerical and categorical features
- Multi collinearity and Variance Inflation Factor

Overview

Lesson 07: Bayesian approach

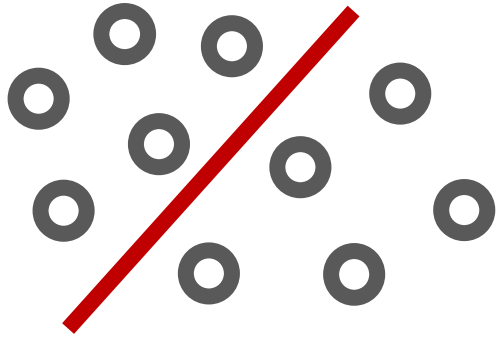
- Bayesian decision model
- Bayesian risk
- Decision strategy
- Example: TBD
- Homework: TBD



Overview

Lesson 08: Regression methods

- Linear discrimination
- Linear regression for data prediction
- Logistic regression for data classification

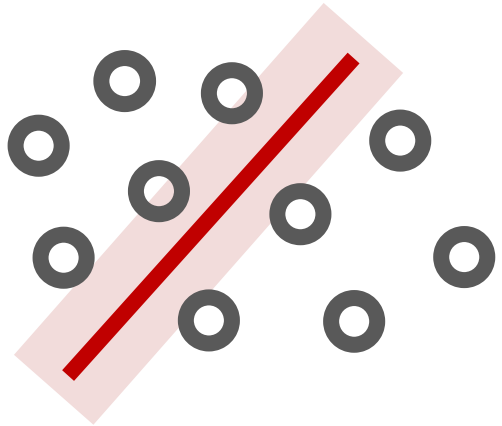


https://www.w3schools.com/python/python_ml_multiple_regression.asp

https://medium.com/codex/step-by-step-guide-to-simple-and-multiple-linear-regression-in-python-867ac9a30298

Lesson 09: Confidence intervals for regression methods

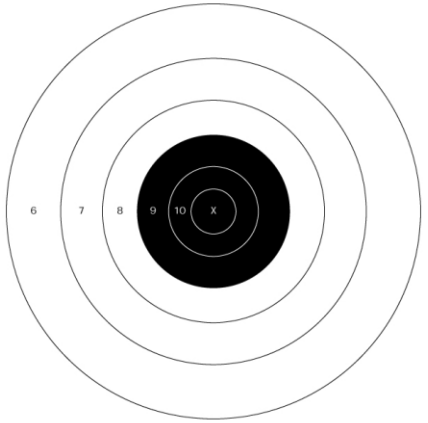
- Coding session
- Confidence interval evaluation
- Out-of-box solutions



Overview

Lesson 10: Supervised Learning parametrical methods

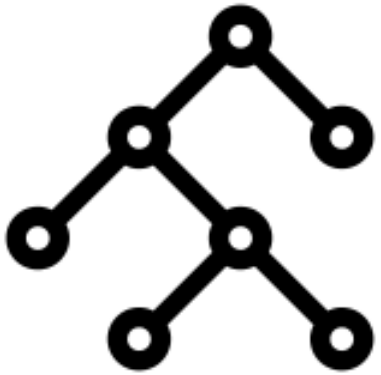
- Naive Bayes classifier
- Gaussian classifier
- SVM



Overview

Lesson 11: Supervised Learning non-parametrical methods

- KNN: K-nearest-neighbors
- Decision tree
- Bias vs Variance tradeoff



Overview

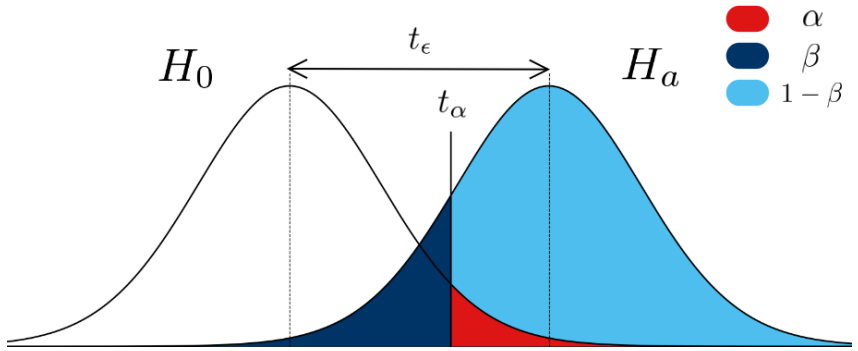
Lesson 12: Unsupervised Learning

- K-means
- EM algorithm
- dbscan

Overview

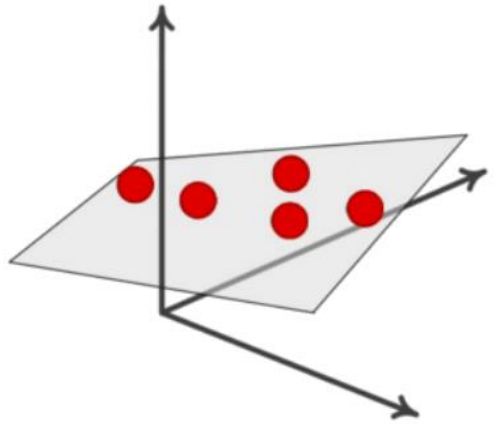
Lesson 13: Non-Bayesian Approach

- Error Type I
- Error Type II
- Neyman-Pearson approach
- Minimax approach
- Example: TBD
- Homework: TBD



Lesson 14: Dimension Reduction

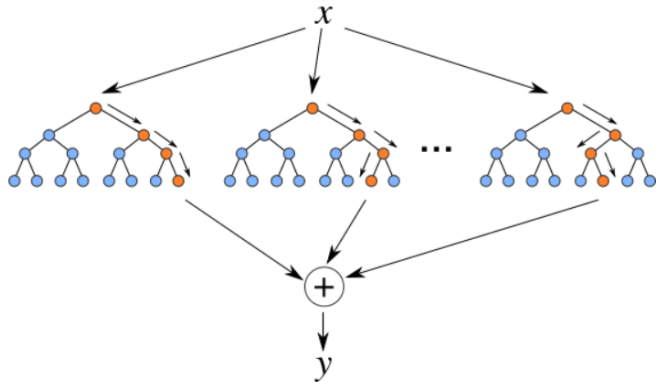
- PCA
- tSNE
- umap



Overview

Lesson 15: Ensemble learning - RF

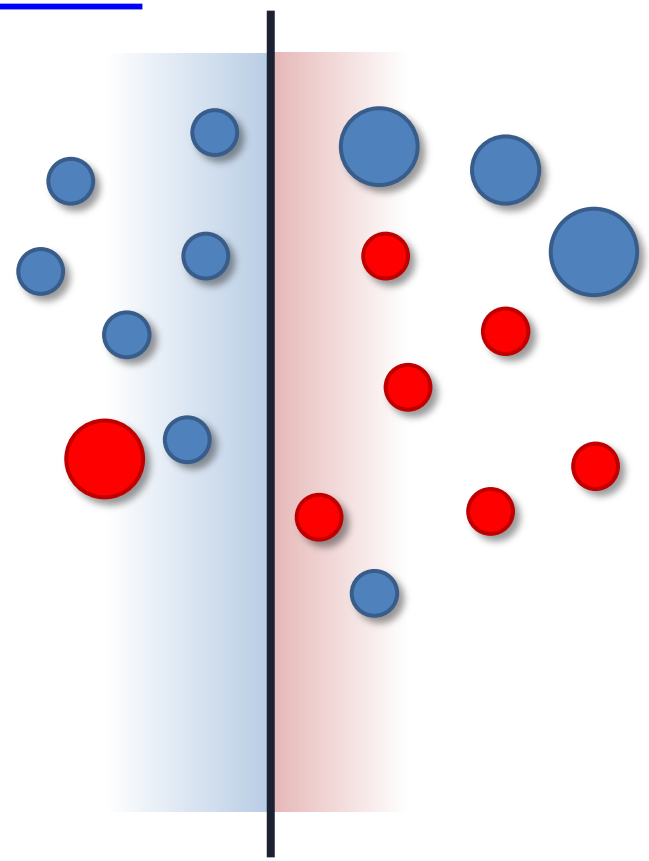
- Bagging
- Boosting
- Bootstrapping
- Random Forest



Overview

Lesson 16: Ensemble learning - Adaboost and XGB

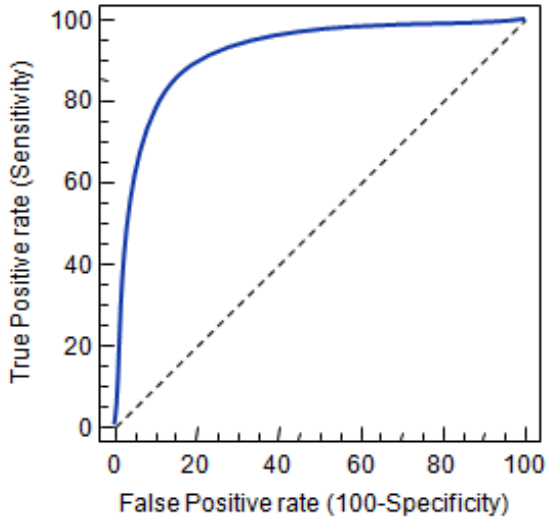
- Adaboost
- XGB



Overview

Lesson 17: Benchmarking

- Accuracy
- Precision
- Recall
- F1 score
- ROC curve
- PR curve



Overview

Lesson 18: Time series

- Stationarity, Trends, Seasonality
- Regression approaches
- Regularization with Ridge regression
- ARIMA
- Gated Recurrent Units (GRU)

Overview

Lesson 19: Introduction to Deep Learning

- Basic operations
- CNNs out of box
- Convolution in details
- Transfer learning
- Engineering CNN with tensorflow.keras
- Visualization of layers and kernels