Bibliography

1. G. James, D. Witten, T. Hastie, and R. Tibshirani, Introduction to statistical learning: with applications in R. Springer, 2017.
2. A.C. Muller, and S. Guido, Introduction to ML with Python. O'Reilly, 2016.
3. A.Geron, Hands on Machine Learning with Scikit-learn. O'Reilly, 2017.
4. C. Richert, ML with Python. Second Edition. PACKT, 2015.
5. G. Hackeling, Mastering ML with Scikit-learn. Packt, 2014.
6. J. VanderPlas, Python Data Science Handbook. O'Reilly, 2016
7. J. Avila, Scikit-learn Cookbook - Over_80_Recipes. Second Edition. Packt 2017.
8. M. Bowles, ML in Python. Wiley, 2015
9. P. Harrington, ML in Action. Second Edition. Packt 2015.
10. R. Garreta, G. Moncecchi, Learning Scikit-learn ML in Python. Packt, 2013.
11. W. Richert, L. P. Coelho, Building ML Systems with Python. Packt, 2013.

Topics

Introduction
- What is Data Science, AI, Machine Learning and Deep Learning?
- Historical overview
Main Concepts
- Supervised vs unsupervised learning
- Regression vs classification
- Parametric vs non-parametric models
- No-free-lunch theorem
- Ockham's razor principle
Regression: Model Accuracy
- Error decomposition into reducible and irreducible
- Bias-variance decomposition
Linear Regression
- Simple linear regression
- Multiple linear regression
- p-value, F-statistic, R squared
- Potential problems in regression: non-linearity, correlation of error terms, non-constant variance of the error terms, outliers, high-leverage points, collinearity
- Interaction terms
- Model selection: AIC, BIC, C_p, adjusted R-squared
Resampling Methods
- The validation set approach
- Leave-one-out cross validation
- K-fold cross validation
- Bootstrap

Regularization
- Ridge regression
- The Lasso
- Elastic net

Classification: Model Accuracy
- Accuracy and error rate of the classifier
- Classification measures – TPR, FPR, Precision, Sensitivity, ROC curve, AUC
- Baseline models

K-Nearest Neighbors
- Regression
- Classification

Classifiers
- Bayes classifier
- Naïve Bayes classifier
- Logistic regression

Classifiers
- Linear discriminant analysis
- Quadratic discriminant analysis

Decision Trees
- Regression
- Classification

Tree based methods
- Bagging
- Random forests

Tree based methods
- Boosting
- C5.0 Algorithm

Unsupervised Learning: Dimensionality Reduction
- Principal components analysis

Unsupervised Learning: Clustering
- K-means clustering
- Hierarchical clustering
- Dendrograms