

Resources

| Week | Topic | Description | links |
|------|--|---|---|
| 1 | Matlab | Getting started with matlab with Lynda "Matlab Master Classes" | Link , Link |
| 2 | Data manipulation | Python data science handbook has all essential tools for working with data: Numpy, Pandas, Matplotlib, Scikit-learn and other related tools, | Python data science handbook |
| | Data exploration | This paper addresses user goals for data manipulation process and data exploration. | Using Aggregation and Dynamic Queries for Exploring Large Data Sets |
| | Visualization techniques | Visualization with matlab | Visualization using Matlab |
| | Quandl API | | Quandl GitHub with Matlab |
| 3 | Probability, Statistical description, descriptive statistics | <ul style="list-style-type: none"> - How statistics are used in the fact based world, and how descriptive statistics are calculated - How to visualize data and different techniques used in data visualization | <ol style="list-style-type: none"> 1. https://fullfact.org/ 2. https://www.gapminder.org/videos/200-years-that-changed-the-world-bbc/ 3. http://pantheon.media.mit.edu/treemap/country_exports/DZ/all/-4000/2010/H15/pantheon 4. https://bigdata-madesimple.com/top-5-best-data-visualization-techniques-for-2018/ 5. Hogg, R.V., and Tanis, E.A. (2015). <i>Probability and Statistical Inference</i>, Prentice Hall, 9th edition. 6. Ott, R. L. and Longnecker, M. (2016). <i>An Introduction to Statistical Methods and Data Analysis</i>, 7th Edition, Cengage Learning. ISBN 13: 978-1-305-26947-7, ISBN 10: 1-305-26947-0 - Available via CMU Library |
| 4 | Statistical hypothesis testing, quantifying confidence | <ul style="list-style-type: none"> - Monitoring the quality of products, leveraging statics to make decisions, and different techniques used in statistical hypothesis testing. | <ol style="list-style-type: none"> 1. Hypothesis testing and examples 2. Power analysis 3. Confidence intervals 4. Techniques used in hypothesis testing pdf 5. Techniques for hypothesis testing 6. Akritas, M., (2015). <i>Probability & Statistics with R for Engineers and Scientists</i>, 1st edition, Pearson, ISBN-13: 978-0321852991. |
| 5 | Time series analysis, auto regression, moving averages | <ul style="list-style-type: none"> - Trends and their importance, and some statistical analysis on detecting trends - Real time decision making, difficulties of going live and the application of auto regression and moving averages doing the analysis | <ol style="list-style-type: none"> 1. Shumway R.H., Stoffer, D.S. (2012). <i>Time Series Analysis and Its Applications With R Examples</i>, 4th Edition, Springer. ISBN-978-3-319-52451-1 2. Time serie basics 3. Difference for trend and seasonality 4. |
| 6 | Linear regression, parameter estimation, model selection, evaluation | <ul style="list-style-type: none"> - Understanding the past and forecasting the future - Techniques for linear regression - Model evaluation | <ol style="list-style-type: none"> 1. Regression analysis and forecasting models 2. Build a stock predict algorithm 3. http://archive.ics.uci.edu/ml/index.php (interesting dataset) 4. StatLib (dataset from CMU) 5. Linear regression with python 6. Evaluating a linear regression |
| 7 | Machine learning essentials | Which machine learning algorithm should I use? | Machine Learning Algorithm |
| | Statistical learning:Pattern Recognition and Machine Learning | This book provides a comprehensive introduction to the fields of pattern recognition and machine learning | Pattern Recognition and Machine Learning |
| | The sitting rising Test | This paper present the research made to evaluate the association between the ability to sit and rise from the floor and all-cause mortality | The sitting and Rising Test |
| | Machine Learning Timeline | Gives a short history of Machine Learning | A Short History of Machine Learning |

| | | | |
|----|--|---|--|
| | Statistical learning:Data mining, Inference and Prediction | <p>This book gives an overview of Supervised Learning</p> <ul style="list-style-type: none"> -Explains the different methods for linear regression and discussion on their comparison -Explains the different methods for linear classification -Model assessment and selection -Model Inference and averaging -Unsupervised and Ensemble Learning -etc | The Elements of Statistical Learning |
| 8 | Linear models | About linear regression and parameter estimates | Multiple Linear Regression |
| | Linear regression | A case study on a linear regression approach to prediction of stock market trading volume | linear regression approach to prediction of stock market trading volume |
| | Confusion Matrix | Explanations and example on why to use and how to compute confusion matrix | Simple guide to confusion matrix Understanding confusion matrix |
| 9 | Mutual Information | Kinney & Atwal (2014).Equitability,mutual information and the maximal information coefficient, PNAS 111(9):3354-3359. | Link to paper |
| | Surrogate Data — A Qualitative and Quantitative Analysis | Maiwald, Thomas & Mammen, Enno & Nandi, Swagata & Timmer, Jens. (2007). Surrogate Data — A Qualitative and Quantitative Analysis. Understanding Complex Systems. 2008. 41-74. 10.1007/978-3-540-75632-3_2. | Link to paper |
| | Max Relevance, Minimum Redundancy (mRmR) | H. Peng, F. Long, C. Ding Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy IEEE Trans. Pattern Anal. Mach. Intell., 27 (8) (2005), pp. 1226-1238 | Link to paper https://ieeexplore.ieee.org/document/1453511 |
| 10 | Naïve Bayes | Basics theory of naive bayes algorithm for classification | link |
| | | Basics theory and implementation of naive bayes algorithm for classification in Python | link |
| | Decision Trees | Basics theory of classification and regression using decision trees | link |
| | | Implement a classification and regression tree algorithm from scratch using python | link |
| | | A complete tutorial on tree based modeling i.e. decision trees, random forest and gradient boosting from scratch (in R and Python) | link |
| | Neural network | A gentle introduction to theories and architectures of neural network | link |
| | | Master the core concepts of neural networks, including modern techniques for deep learning | link |
| | Linear discriminant analysis (LDA) | Basics theory of linear discriminant analysis algorithm for classification predictive modeling problems | link |
| | | Basics theory and implementation bit by bit of linear discriminant analysis algorithm for classification predictive modeling problems in python | link |
| | LDA and QDA | Basics theory and implementation of linear discriminant analysis and quadratic discriminant analysis | link |

| | | | |
|----|---|---|---|
| | | Matlab tutorial : Create and Visualize Discriminant Analysis Classifier | link |
| | Discriminant Analysis | A lesson on discriminant analysis offered by Eberly College of science | link |
| | | Discover Which Variables Discriminate Between Groups, Discriminant Function Analysis | link |
| | k-nearest neighbors algorithm (k-NN) | Basics theory on kNN algorithm for classification and regression | link |
| | | Basics theory and implementation kNN algorithm for classification in Python and R | link |
| | | KNN Model-Based Approach in Classification | link |
| | Parametric and Non-Parametric methods | Understanding the basics of Parametric and Nonparametric Machine Learning Algorithms | link |
| | Support Vector Machines (SVM) | Basics theory on SVM algorithm for machine learning | Link1 Link2 |
| | | Support Vector Machines for Binary Classification, Train SVM with different kernels (theory + implementation in Matlab) | link |
| | Kernel functions | A gentle introduction on kernel functions and how are they applied in machine learning? | Quora.com towardsdatascience.com |
| 11 | Principal Component Analysis (PCA) | Detailed introduction to Principal Component Analysis with the necessary mathematical proofs. | medium.com An online course at Penn State A tutorial on Principal Components Analysis |
| | Non-Negative Matrix Factorization (NMF) | Cover a gentle introduction to NMF, how it works and data preparation for NMF | Link1 , Link2 |
| | | Non-Negative Matrix Factorization (NMF) is described in the paper "Learning the Parts of Objects by Non-Negative Matrix Factorization" by D. D. Lee and H. S. Seung | Link to paper |
| | | Application of NMF properties in image processing, text mining and hyperspectral imaging in the paper "The Why and How of Nonnegative Matrix Factorization" by Nicolas Gillis | Link to paper |
| | Cluster Analysis | A detailed theoretical and practical coverage on cluster analysis algorithms; k-means, Self-Organizing Map and Hierarchical Clustering, and | Introduction to Data Mining (Second Edition) by pang, Anuj, vipin, michael. chap 7 and 8 |
| | K means | Overview of K-means clustering and a step by step practical implementation in python | geeksforgeeks.org |
| | Gaussian Mixture Models (GMM) | Basics theory on GMM and a step by step implementation in python | geeksforgeeks.org |
| | | | |