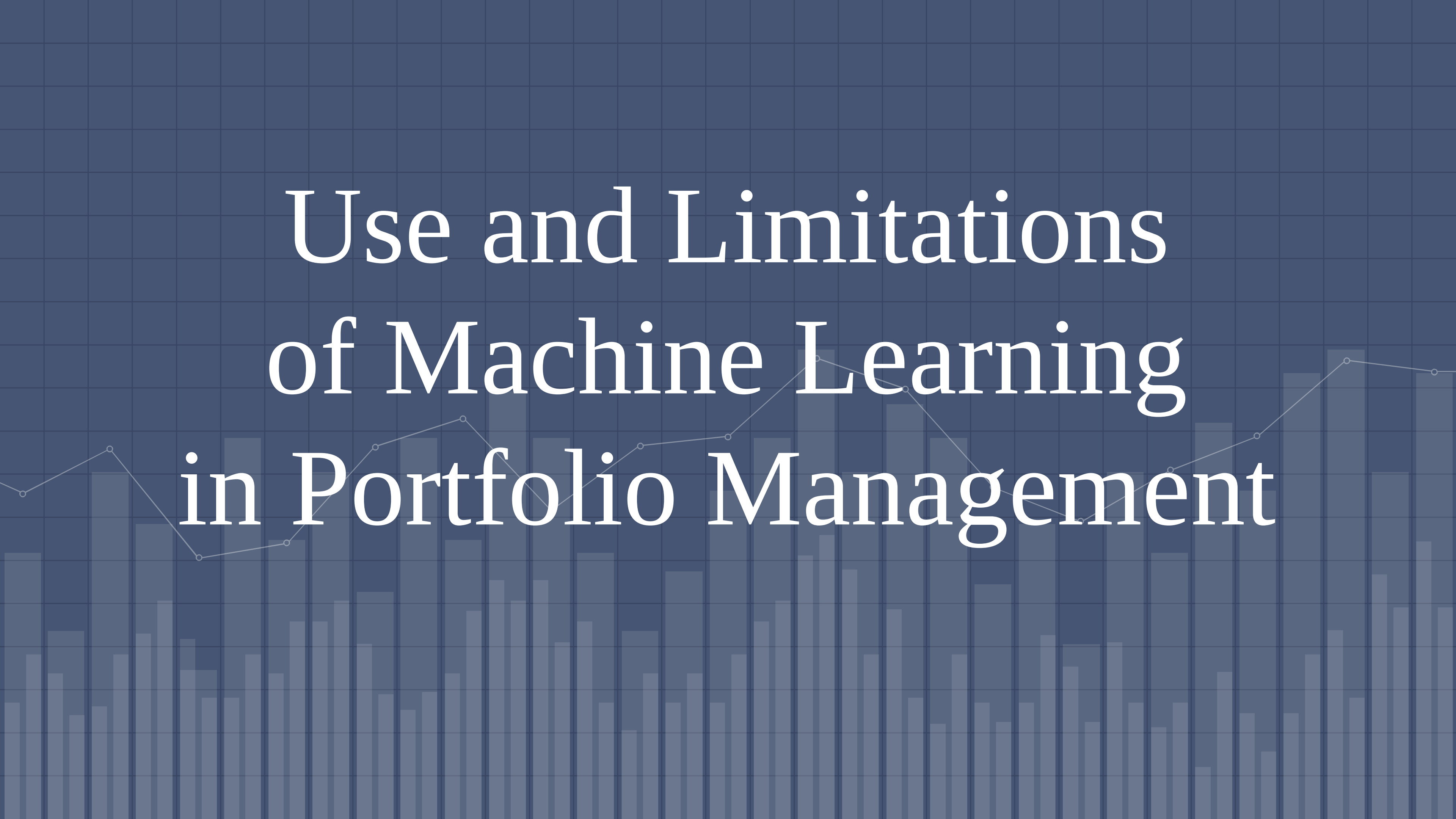
Use and Limitations of Machine Learning in Portfolio Management



Overview

1. Brief Introduction to Learning

2. Prediction
    - “Futurecasting”
    - “Nowcasting”
    - factor analysis

3. Similarity Measures
    - recommendation system

4. Generating Synthetic Datasets

A Brief Introduction to Learning

Learning: $Y|X$

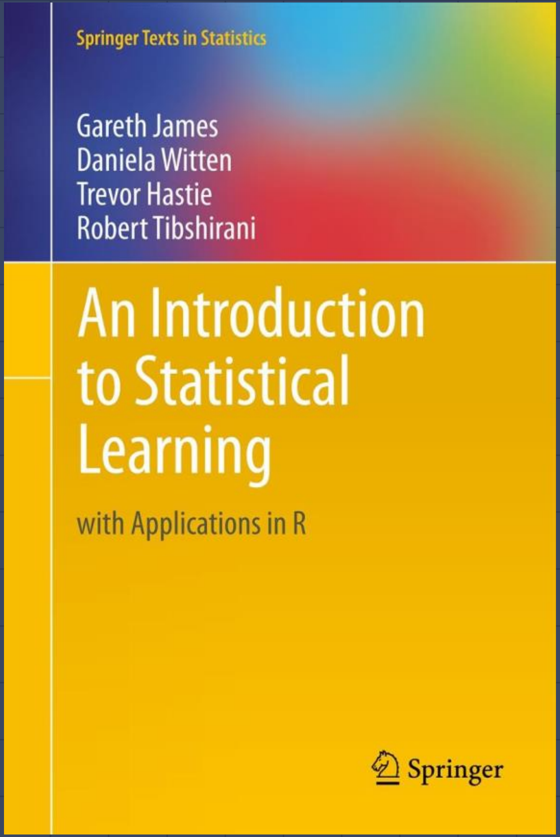
- Regression: $E[Y|X=x]$
- Classification: $P(Y=y|X=x)$
- Synthetic data generation: $Y|X=x$

To each problem its solution

- What we want to know from Y
- Dimensionality of the data (X and Y)
- Signal to noise of the data
- Risk function
- Stationarity
- Etc.

An Introduction to Statistical Learning

Great overview of classic machine learning techniques with examples of code in R



Springer Texts in Statistics

Gareth James
Daniela Witten
Trevor Hastie
Robert Tibshirani

An Introduction to Statistical Learning

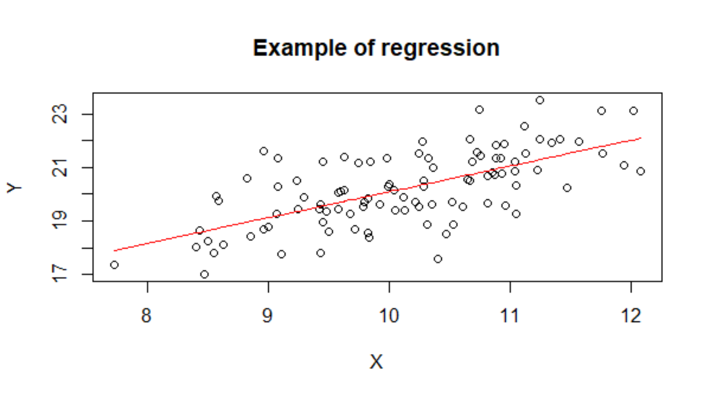
with Applications in R

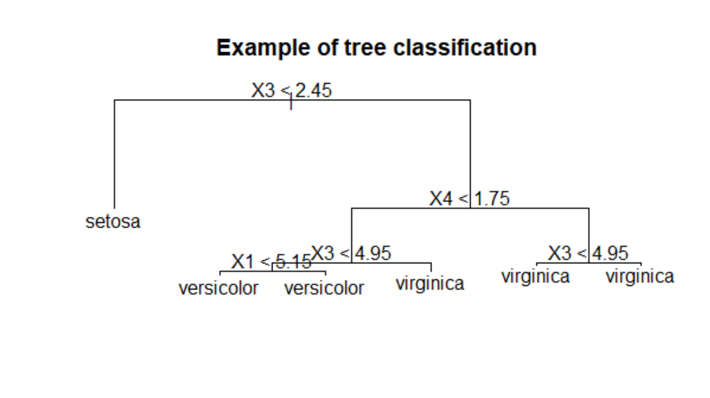
Springer

Prediction

Methods Used

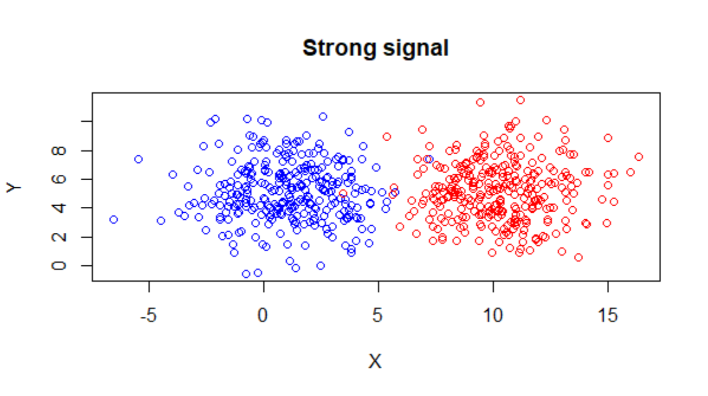
- OLS Regression
- Lasso, Ridge, Elastic Net
- Kernel Regression
- Trees
- Neural Nets
- Random Forests
- SVMs
- Etc.

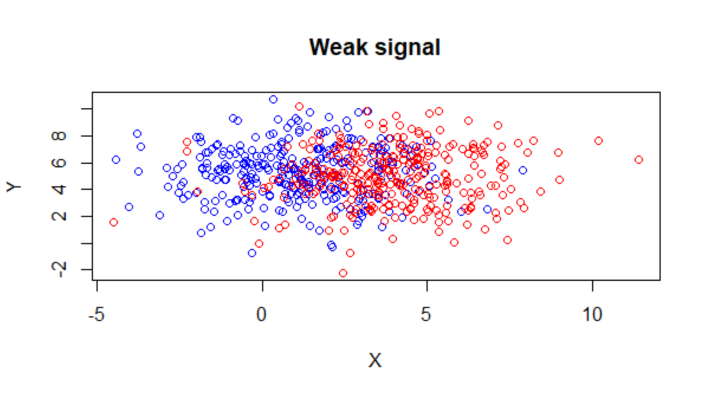




Prediction - Things to Consider

- Linear versus non-linear
- Dimensionality of the data
- Density of the data
- Signal to noise
- Risk function
- Interpretability
- Over-fitting





Prediction - “Futurecasting”

- No access to contemporaneous data
- Very difficult to do
- Markets tend to be efficient
- Signal to noise ratio is poor
- It is difficult to beat naïve predictors
- Boosted Trees is the leader at the moment

Big Data and AI Strategies

Good overview of the current use of machine learning in alpha generation and more

J.P.Morgan

May 2017

Big Data and AI Strategies

Machine Learning and Alternative Data Approach to Investing

Quantitative and Derivatives Strategy

Marko Kolanovic, PhD[AC]
marko.kolanovic@jpmorgan.com

Rajesh T. Krishnamachari, PhD
rajesh.tk@jpmorgan.com

See page 278 for analyst certification and important disclosures, including non-US analyst disclosures.

Completed 18 May 2017 04:15 PM EDT
Disseminated 18 May 2017 04:15 PM EDT

Prediction - “Nowcasting”

- Access to contemporaneous data
- Important data that is published with a lag or a low frequency
- Generating replicating portfolios (Stat Arb)
- Live estimates of
  - ERP
  - GDP
  - Macroeconomic indicators
  - Etc.

Prediction - Factor Analysis

- p: number of predictors
- n: number of observation
- It used to be n>>p
  - OLS was useful
- It is now p>n (zoo of factors)
  - curse of dimension
    - dimensionality reduction, PCA, clustering, etc.
    - best subset, Lasso, Ridge, etc.
    - K-fold cross validation
- Also useful for hedging

Similarity Measures

Useful For

- Manager selection
- Stock selection
- Style drift detection

Similarity Measures

Methods Used

- PCA
- Hierarchical Clustering
- K-means
- Supervised classifiers
- Etc.

Used For

- Alternative data
- Big data
- Improving analyst’s productivity

Similarity Measures - Things to Consider

- Supervised
  - labeling the target variable and letting the learner infer useful predictors
- Unsupervised
  - choosing predictors where “closeness” is of interest and letting the algorithm do the clustering
- Non stationarity of data
- Renormalization
- Availability of data for back testing

Generating Synthetic Data

Useful For

- Scenario analysis
- Stress testing
- Risk budgeting
- Option pricing
- OOS testing

Could be Useful For

- Training data for data intensive learners (deep learning, reinforcement learning, etc.)
- Testing systematic strategies

Generating Synthetic Data

Methods Used

- Fitting of parametric models
  - distributions (poisson, normal, cauchy, etc.)
  - DGP (EWMA, GARCH, variance gamma process, etc.)
- Kernel density estimation
- Eigen vector decomposition
- Factor analysis
- Auto Encoders
- LSTM NN

Generating Synthetic Data - Things to Consider

- Single versus multivariate inputs
- Single versus multivariate outputs
- Conditional versus unconditional outputs
- Linear versus non-linear relationships
- Bulk versus tails of the distribution
- Interpretability