Business Analytics & Machine Learning

Course Overview

Prof. Dr. Martin Bichler

Department of Computer Science

School of Computation, Information, and Technology

Technical University of Munich

Chair for Decision Sciences & Systems

Prof. Bichler (head), Prof. Brandt, Prof. Etesami

Focus in research:

Game theory, mathematical optimization, data analytics, market design, and social choice.

Core courses:

- *Winter term*
  - Auction Theory and Market Design
  - Business Analytics and Machine Learning
  - Computational Social Choice
  - Seminar on Markets, Algorithms, Incentives, and Networks
  - Seminar on Causal Reasoning
- *Summer term*
  - Operations Research
  - Algorithmic Game Theory
  - Seminar on Economics and Computation
  - Seminar on Learning in Games
  - Causal Inference in Time Series

Agenda for Today

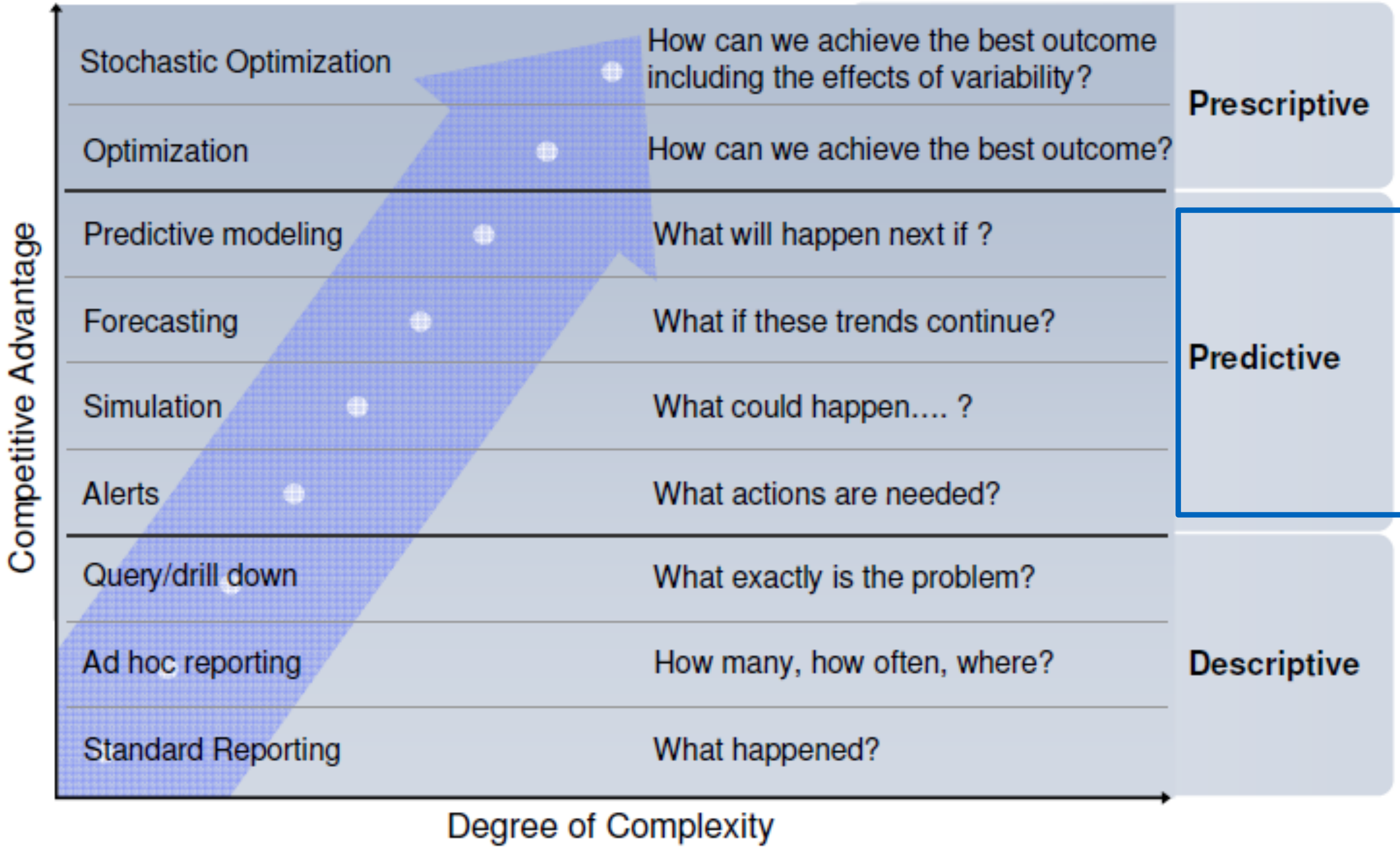
- Understand what this course is all about
- Learn about organization, grading, and tutor groups
- Refresh basic statistical concepts

Business Analytics:
Quantitative Methods for Management Problems

Business Analytics makes extensive use of statistical analysis, including explanatory and predictive modeling, and fact-based management to drive decision making. It is therefore closely related to management science. Analytics may be used as input for human decisions or may drive fully automated decisions.

Source: http://en.wikipedia.org/

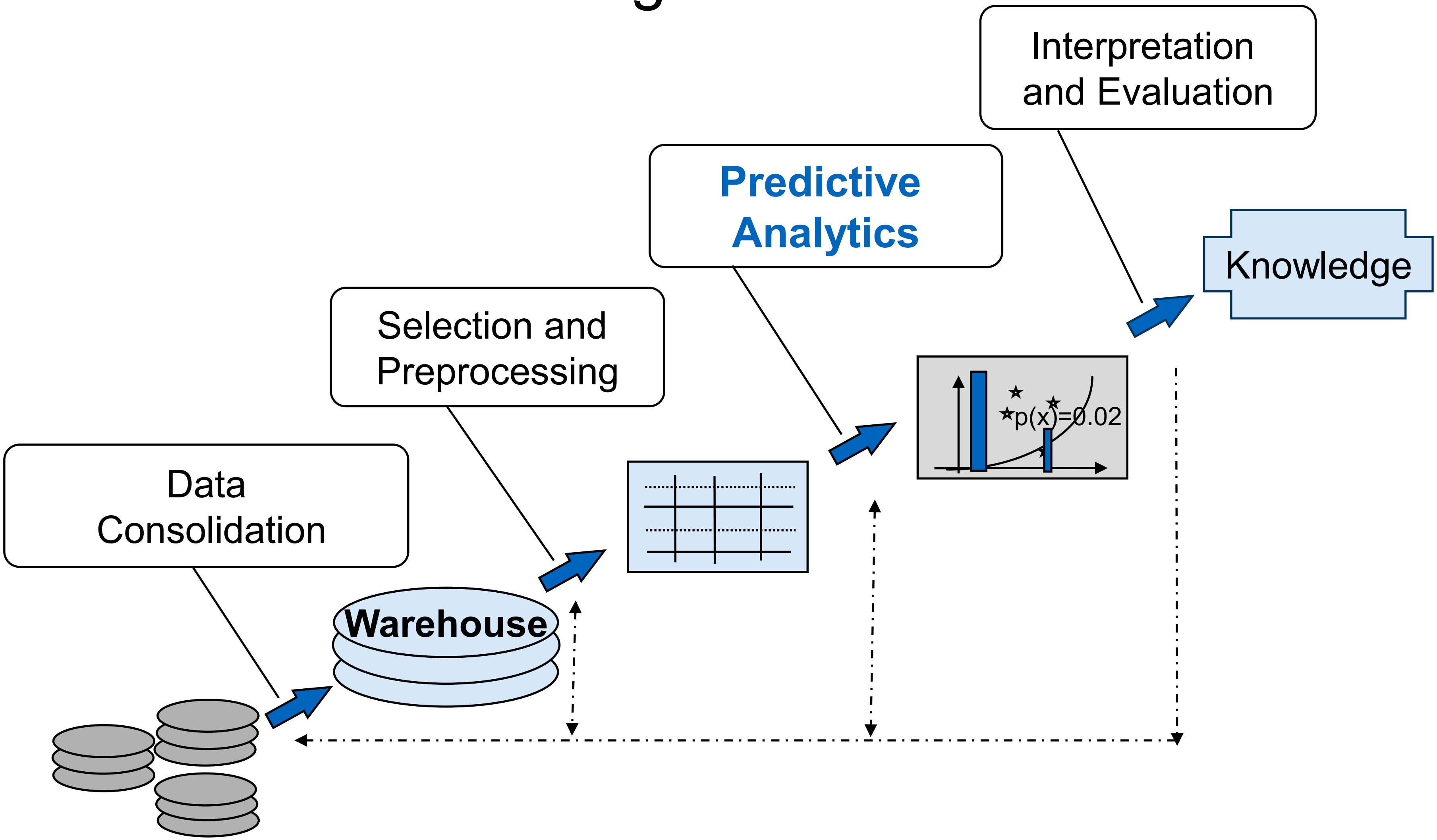
Business Analytics



			Academic Fields
Stochastic Optimization	How can we achieve the best outcome including the effects of variability?	Prescriptive	Operations Research
Optimization	How can we achieve the best outcome?		
Predictive modeling	What will happen next if ?	Predictive	Econometrics & Machine Learning
Forecasting	What if these trends continue?		
Simulation	What could happen.... ?		
Alerts	What actions are needed?		
Query/drill down	What exactly is the problem?	Descriptive	Data Management
Ad hoc reporting	How many, how often, where?		
Standard Reporting	What happened?		

Based on: Competing on Analytics, Davenport and Harris, 2007

From Data to Knowledge



Interpretation and Evaluation

Predictive Analytics

Selection and Preprocessing

Data Consolidation

Predictive Analytics

Predictive Analytics draw on methods from different fields, in particular from econometrics and machine learning.

Econometrics is the application of statistical methods to economic data in order to give empirical content to economic relationships. More precisely, it is "the quantitative analysis of actual economic phenomena based on the concurrent development of theory and observation, related by appropriate methods of inference".

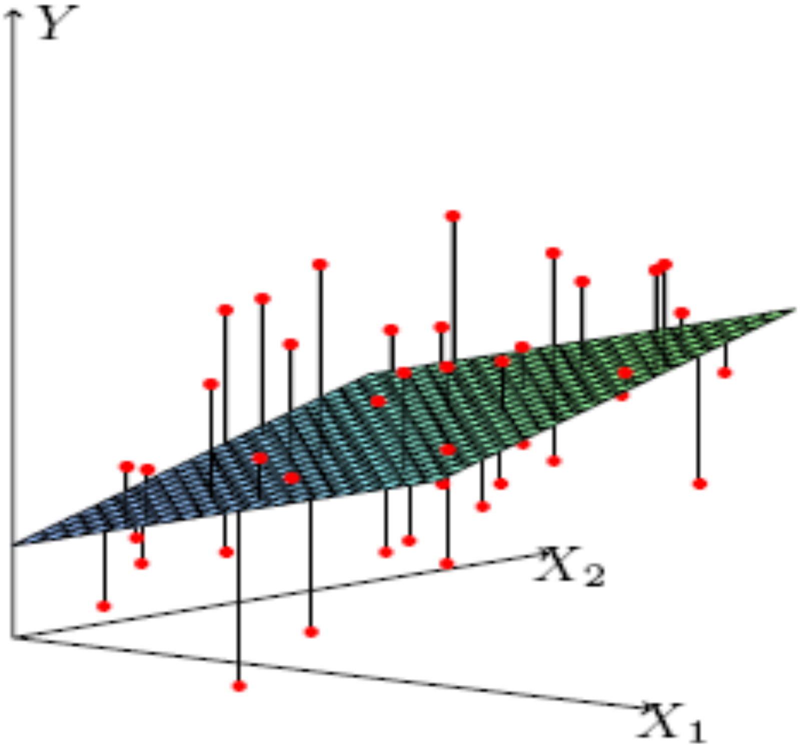
Machine learning (ML) is a field of inquiry devoted to understanding and building methods that 'learn', that is, methods that leverage data to improve performance on some set of tasks..

Source: en.wikipedia.org

Central **tasks** on both fields are numerical prediction, classification, and clustering.

Numerical Prediction

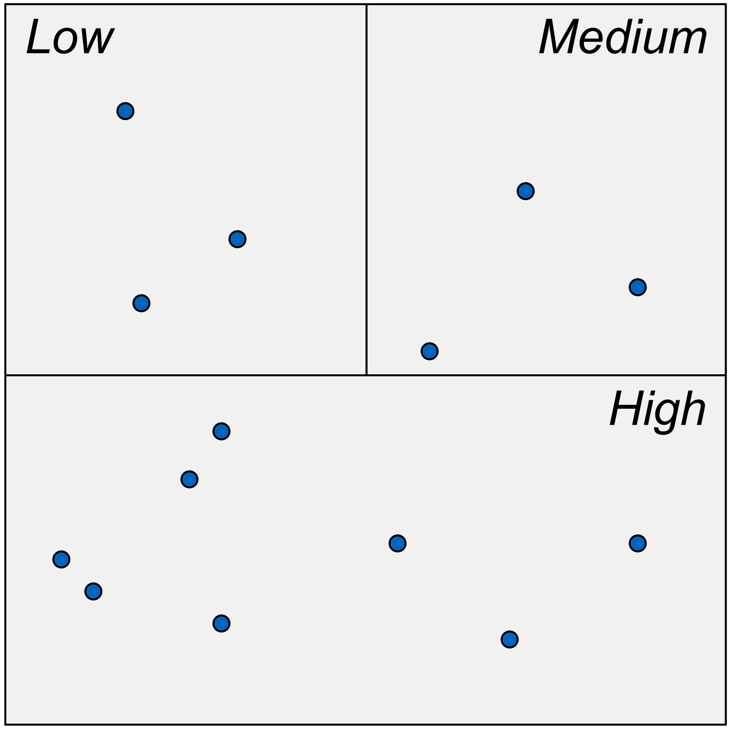
- given a collection of data with known numeric outputs, create a function that outputs a predicted value from a new set of inputs
- for example, given age and income of a person, predict monthly expenses dining



Source: Hastie et al. 2016

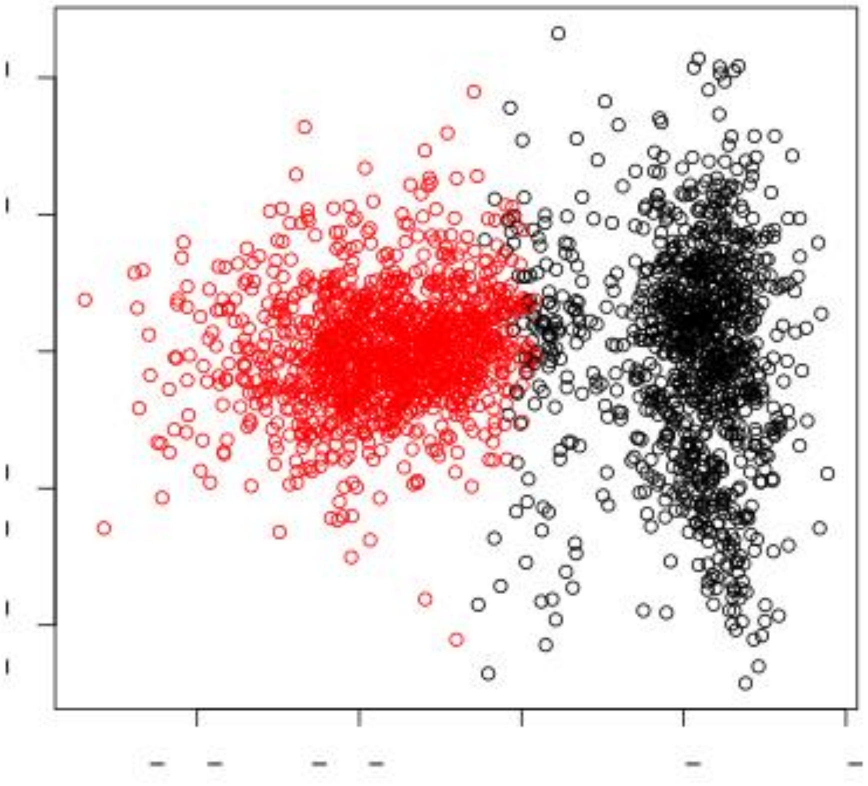
Classification

- from data with known labels, create a **classifier** that determines which label to apply to a new observation
- e.g. identify new loan applicants as low, medium, or high risk based on existing applicant behavior

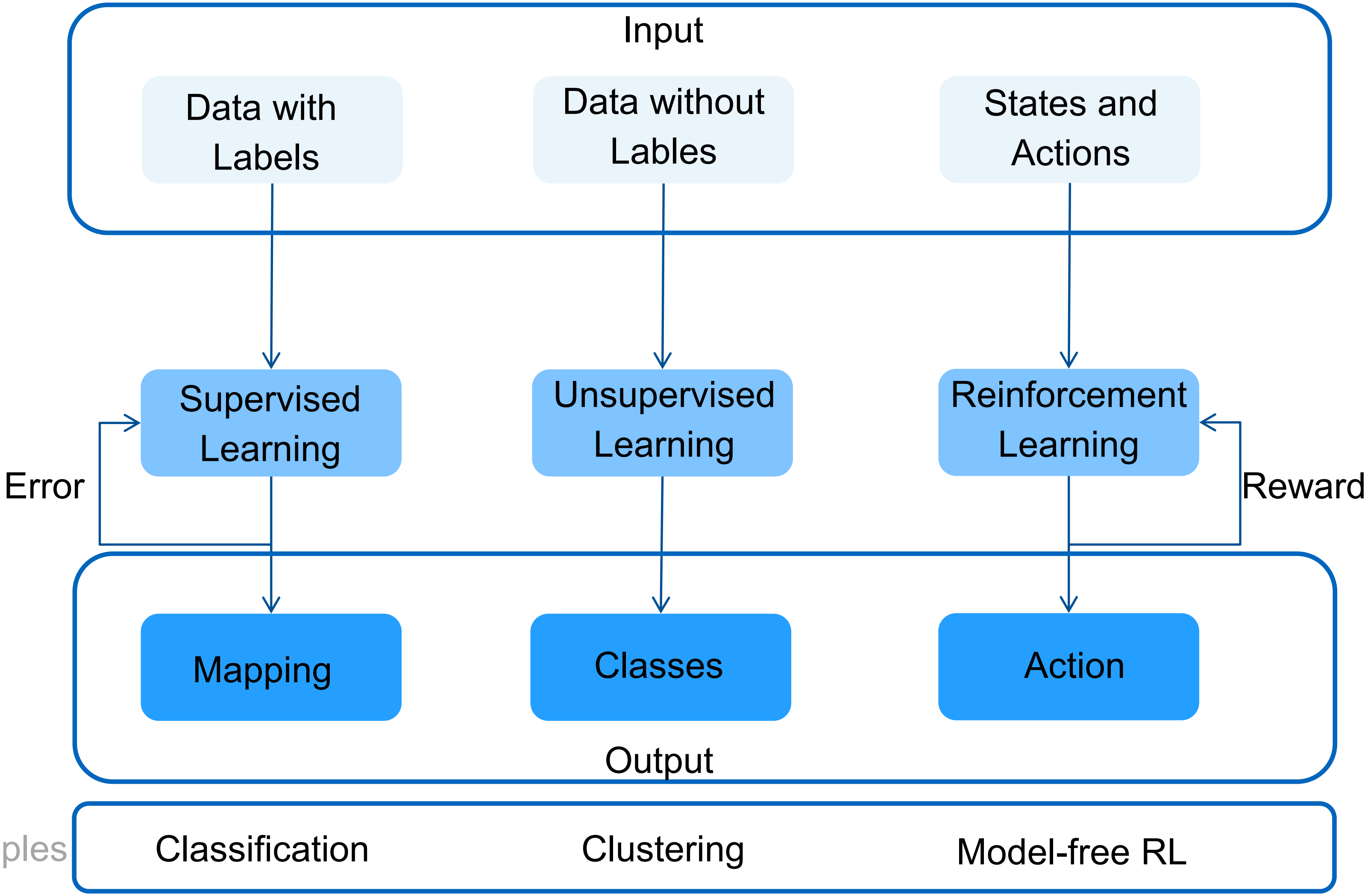


Clustering

- identify “natural” groupings in data
- unsupervised learning, no predefined groups
- e.g. identify clusters of “similar” customers



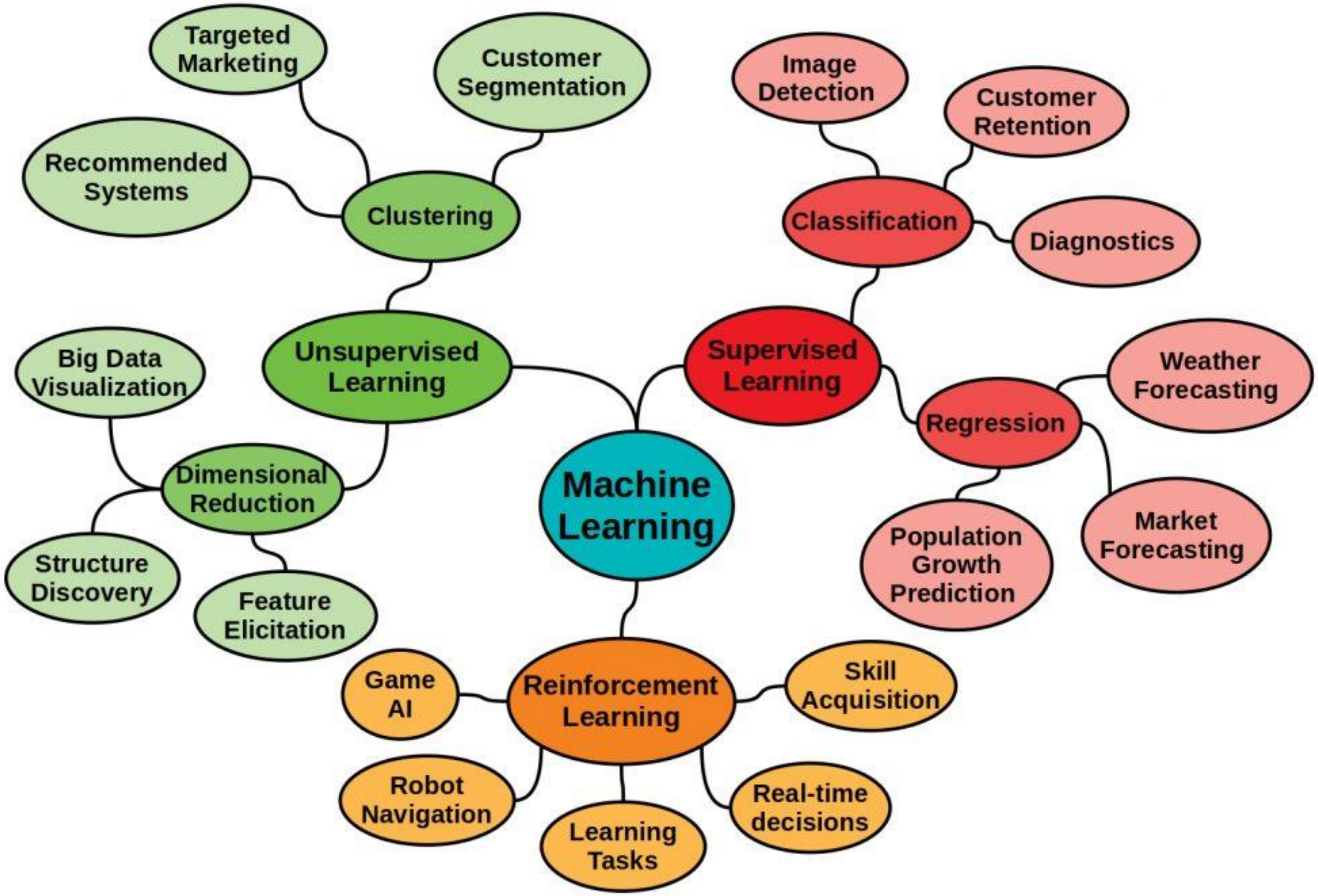
Machine Learning Terminology



Examples

Classification	Clustering	Model-free RL

Tasks in Machine Learning and Applications



Source: https://starship-knowledge.com/

Example Application: Churn Prediction

Churn: The proportion of contractual customers or subscribers who leave a supplier or service provider during a given time period.

Example:

- Churn rates of a telecom at 2.1% per month
- Causes: increased competition, lack of differentiation, market saturation
- Cost: $300 to $700 cost of replacement of a lost customer in terms of sales support, marketing, advertising, etc.
- Response: Targeted retention strategies

Churn Prediction as a Classification Task

Churn as a Classification problem:

Classify a customer i characterized by p variables (name, age, gender, ...)

$x_i = (x_{i1}, x_{i2}, \ldots, x_{ip})$ as

- Churner $y_i = +1$
- Non-churner $y_i = -1$

} perceptron

Churn is the response binary variable to predict: $y_i = f(x_i)$

Choice of the binary choice model $f(\,.\,)$?

What is a Model?

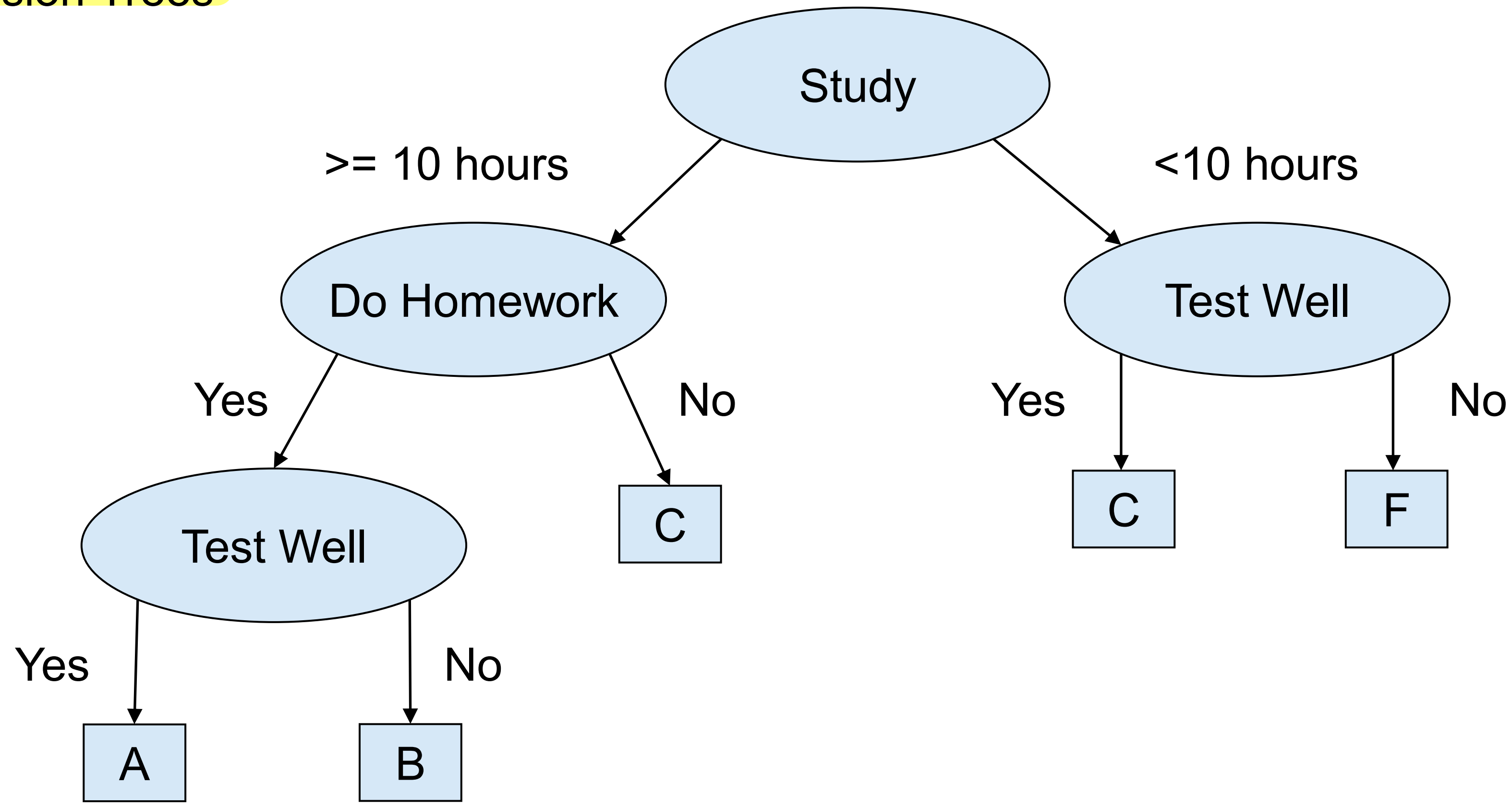
A representation of a system that allows for investigation of the properties of the system and, in some cases, prediction of future outcomes.

Linear functions as a well-known example

- mathematical combination of attribute values
- expenses based on age and income
- could be a non-linear or linear model such as $y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + e_i$

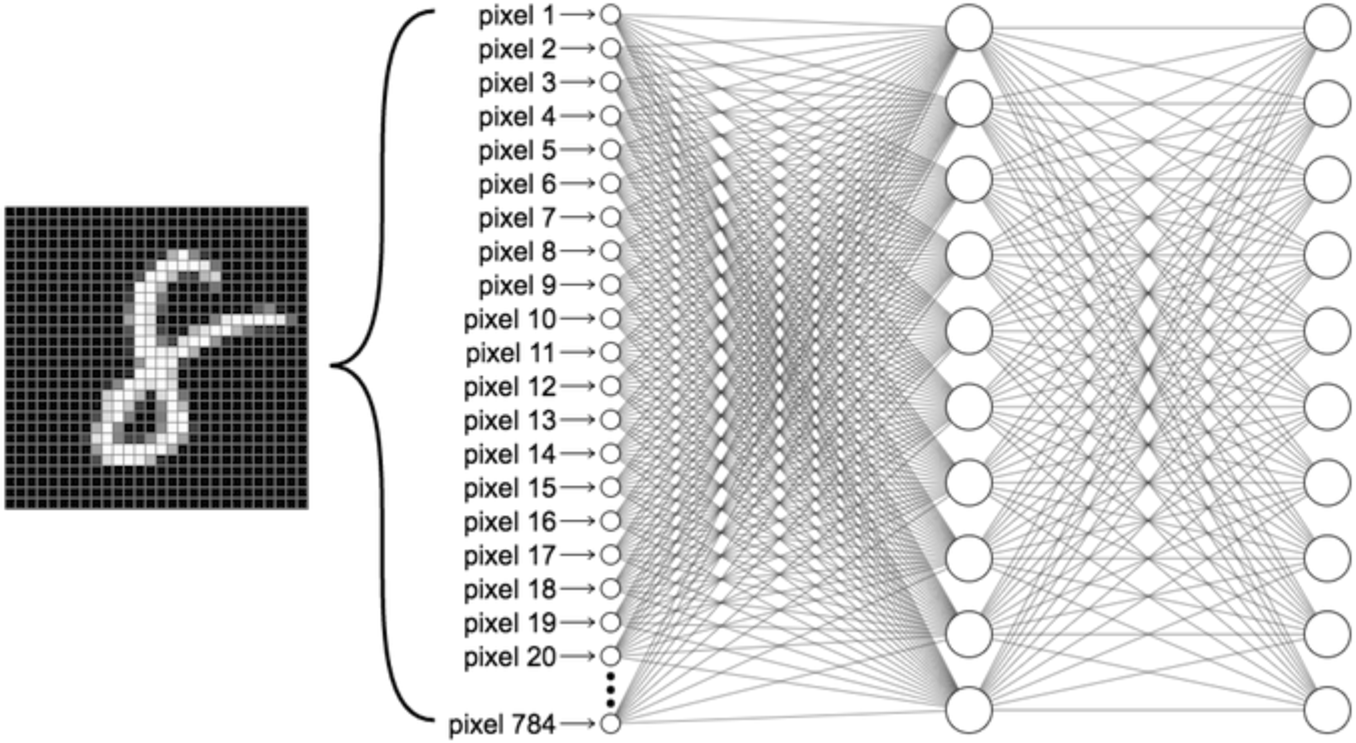
Models

Decision Trees



Models

Neural Networks



0 : $(0-0.412)^2$

1 : $(0-0.002)^2$

2 : $(0-0.112)^2$

3 : $(0-0.702)^2$

4 : $(0-0.012)^2$

5 : $(0-0.102)^2$

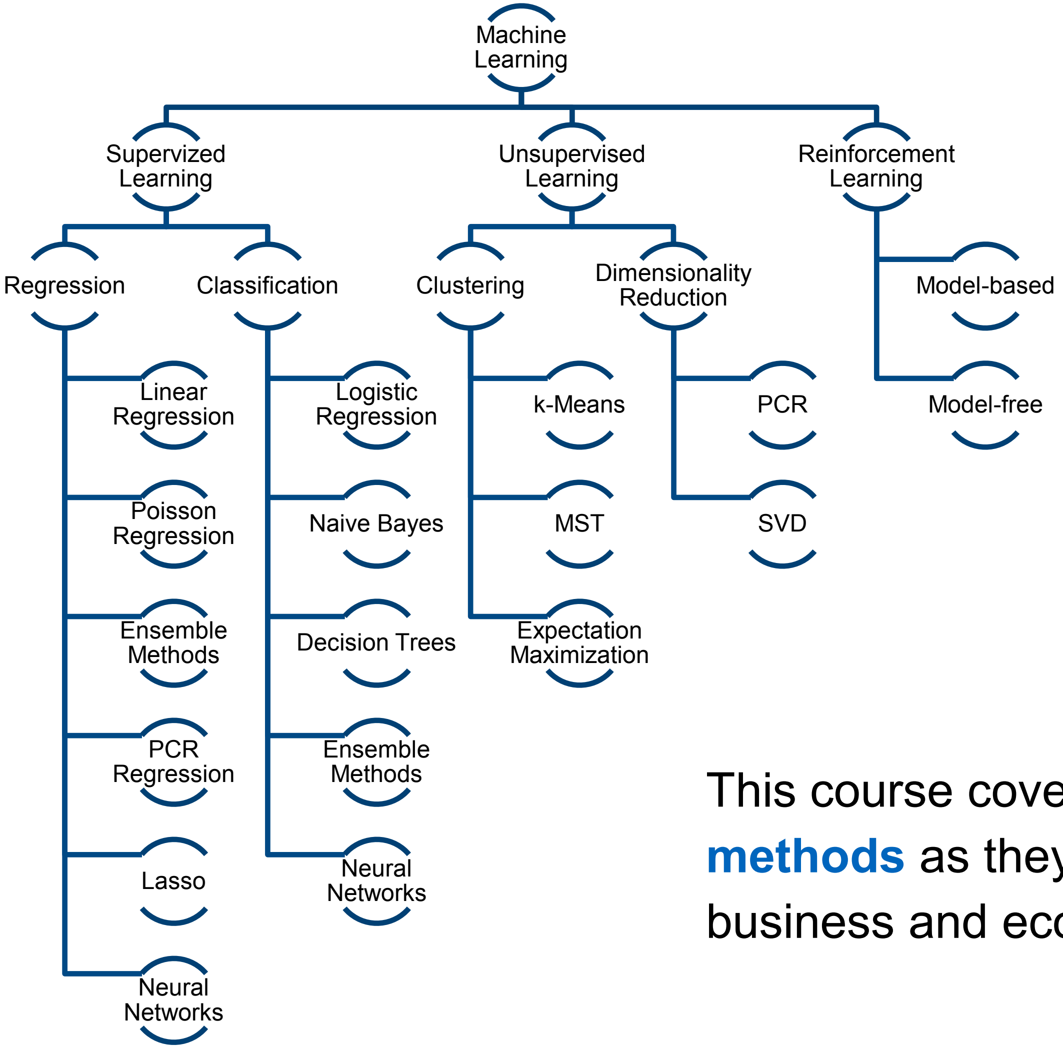
6 : $(0-0.112)^2$

7 : $(0-0.212)^2$

8 : $(1-0.902)^2$

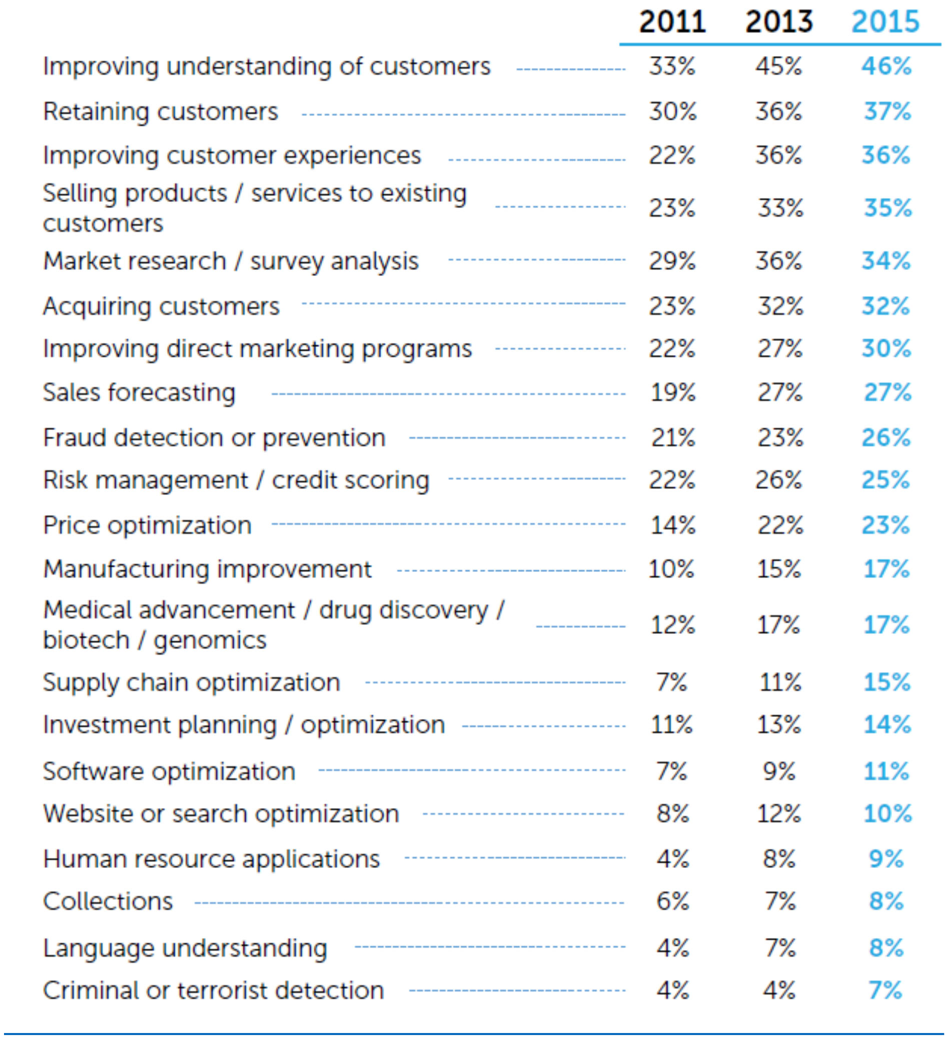
9 : $(0-0.812)^2$

Overview of Methods Discussed in BA&ML



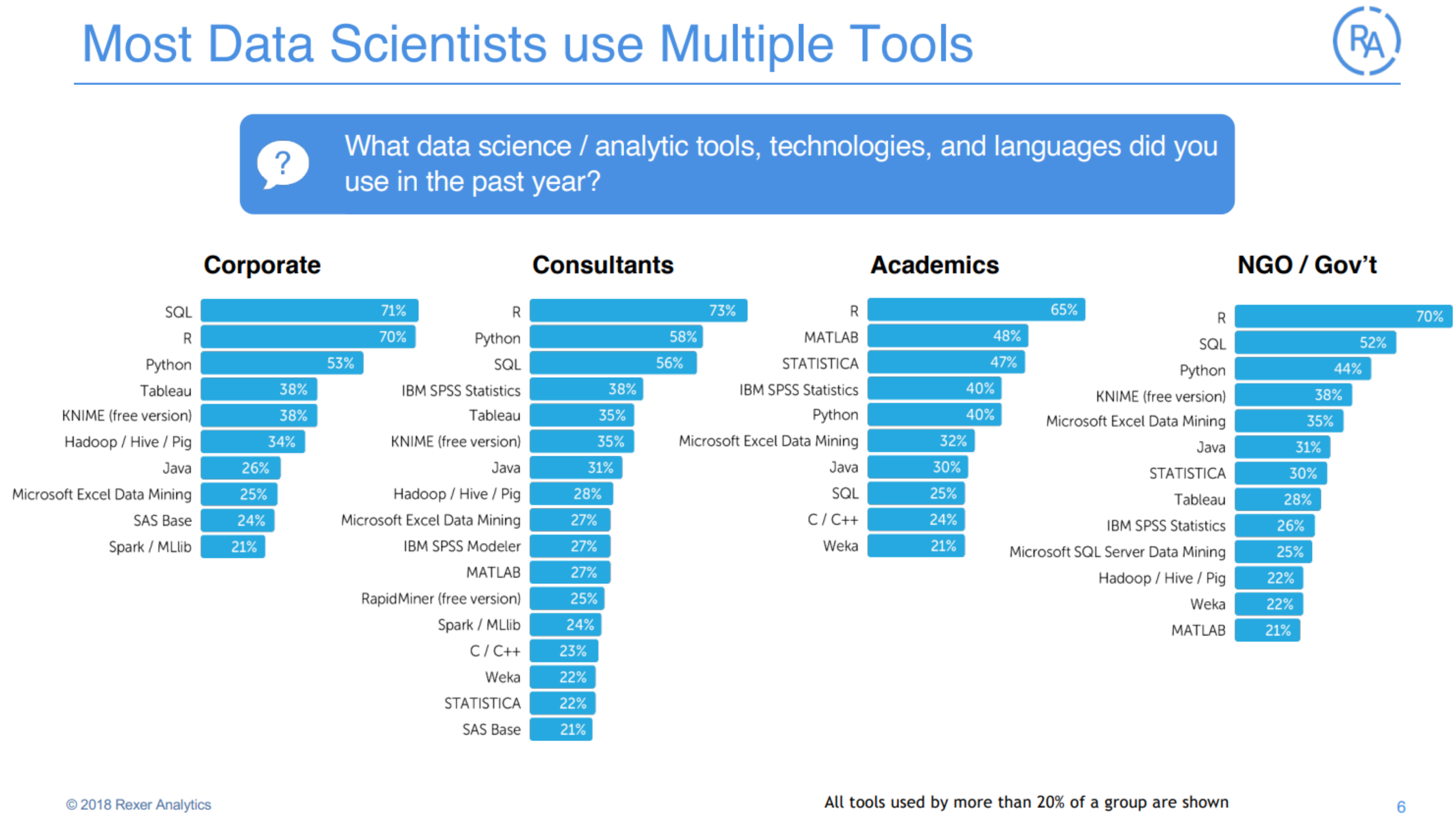
This course covers wide-spread **methods** as they are used in business and economics.

What do Data Scientists Work on?



	2011	2013	2015
Improving understanding of customers	33%	45%	46%
Retaining customers	30%	36%	37%
Improving customer experiences	22%	36%	36%
Selling products / services to existing customers	23%	33%	35%
Market research / survey analysis	29%	36%	34%
Acquiring customers	23%	32%	32%
Improving direct marketing programs	22%	27%	30%
Sales forecasting	19%	27%	27%
Fraud detection or prevention	21%	23%	26%
Risk management / credit scoring	22%	26%	25%
Price optimization	14%	22%	23%
Manufacturing improvement	10%	15%	17%
Medical advancement / drug discovery / biotech / genomics	12%	17%	17%
Supply chain optimization	7%	11%	15%
Investment planning / optimization	11%	13%	14%
Software optimization	7%	9%	11%
Website or search optimization	8%	12%	10%
Human resource applications	4%	8%	9%
Collections	6%	7%	8%
Language understanding	4%	7%	8%
Criminal or terrorist detection	4%	4%	7%

Source: Rexter Analytics Data Science Survey, 2016



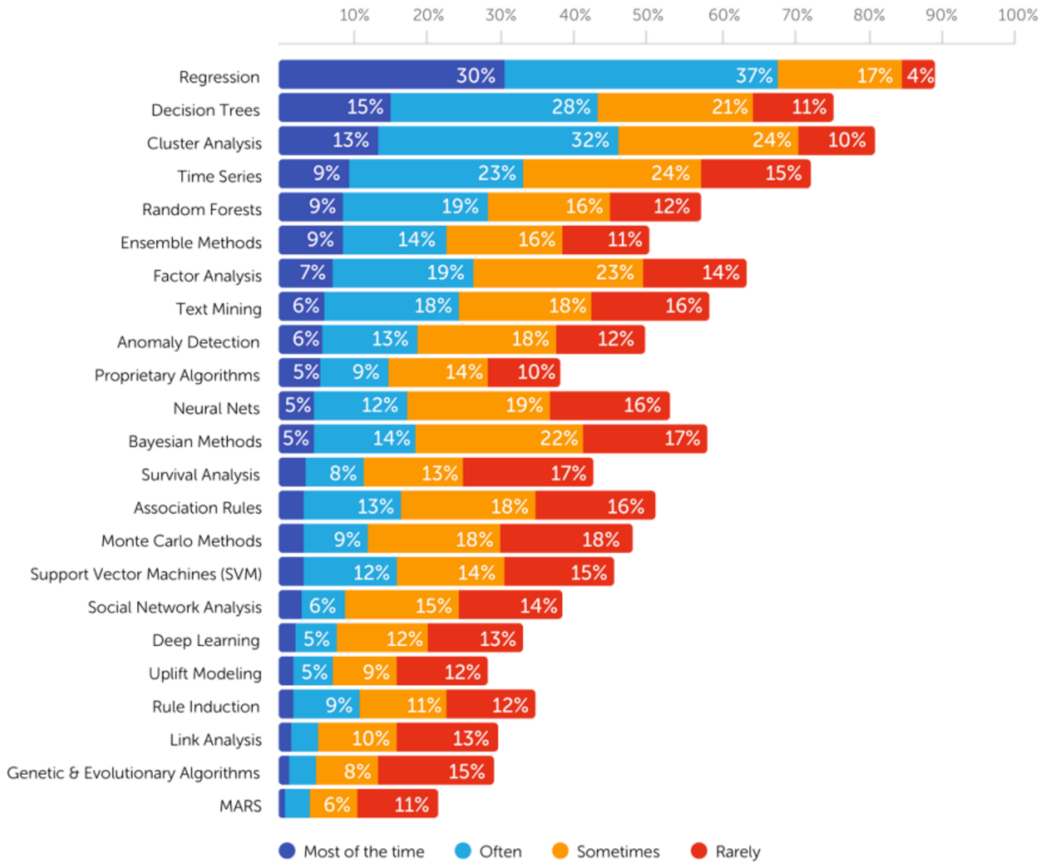
Most Data Scientists use Multiple Tools

What data science / analytic tools, technologies, and languages did you use in the past year?

All tools used by more than 20% of a group are shown

Source: Rexter Analytics, Data Science Survey, 2018

What algorithms/analytic methods do you TYPICALLY use?

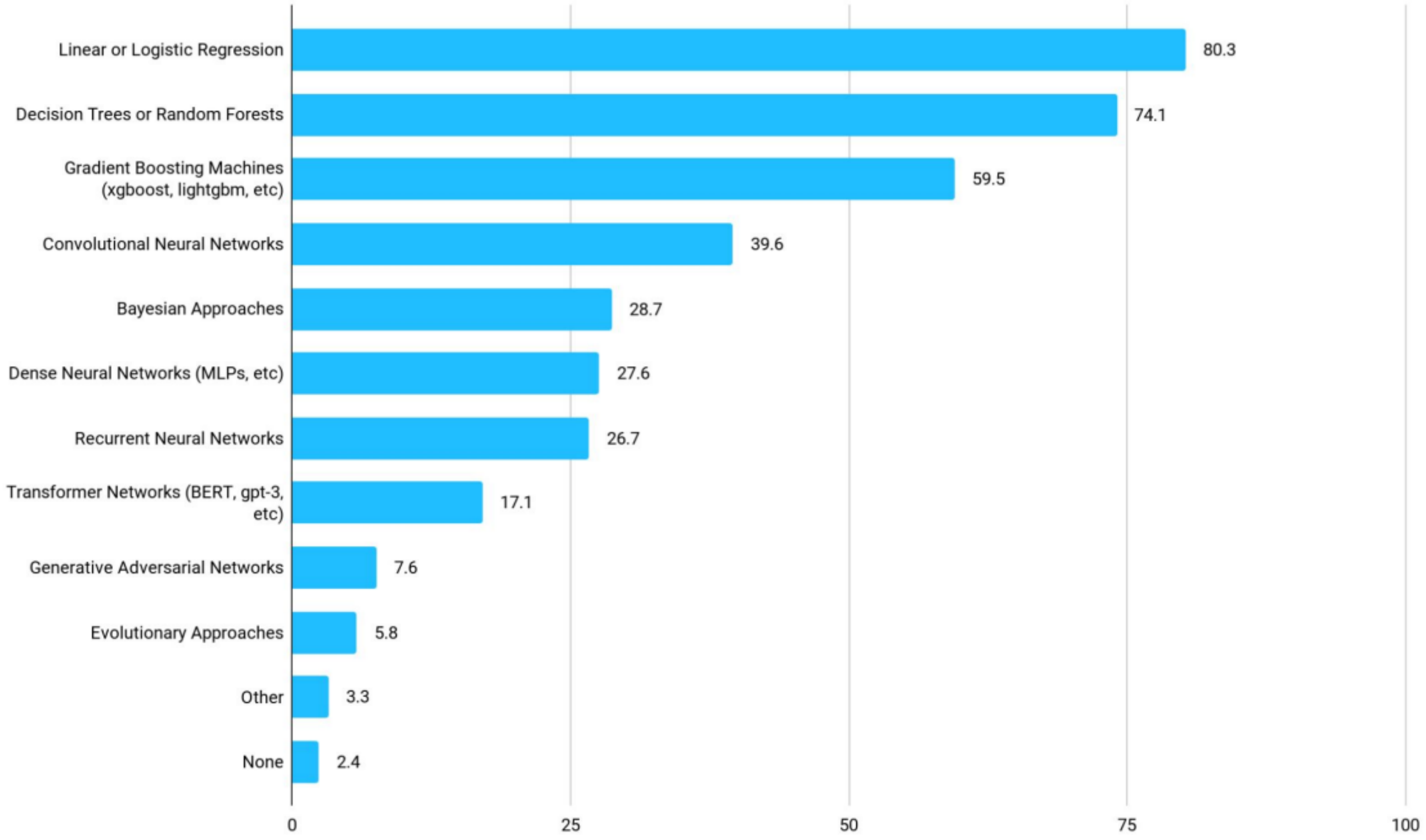


Rexer Analytics' surveys since 2007 have consistently shown that data scientists primarily work with the algorithm triad of regression, decision trees, and cluster analysis. In every survey since 2007, over half of respondents reported using each of these methods in the prior year. Among these three, regression is clearly dominant, with more than two thirds (67%) of respondents indicating that they use regression "often" or "most of the time".

On average 2017 respondents use 11 different algorithms in the course of their work (slightly down from 12 in 2015). Despite extensive media hype about AI, Cognitive Computing, Deep Learning, and the rise of machine learning and its related algorithms, no algorithms showed substantial increased usage since the 2015 survey.

Source: Rexter Analytics, 2018

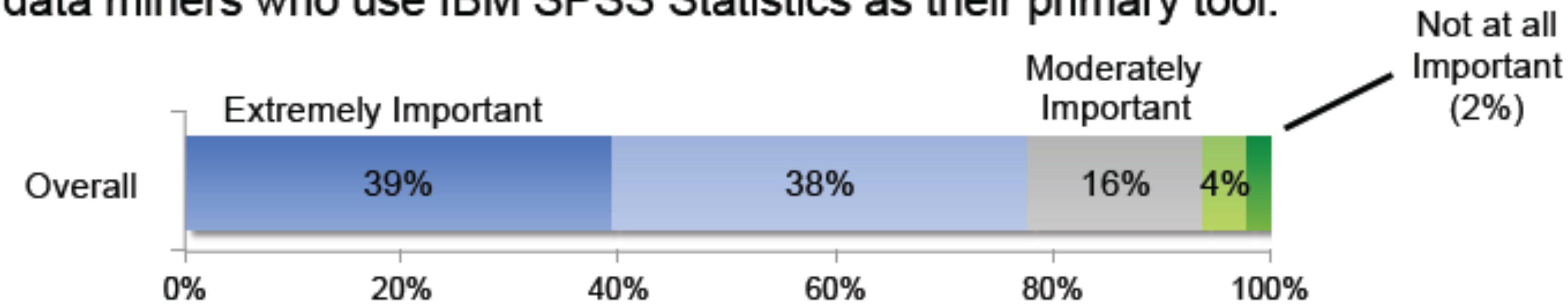
Methods and Algorithms Usage



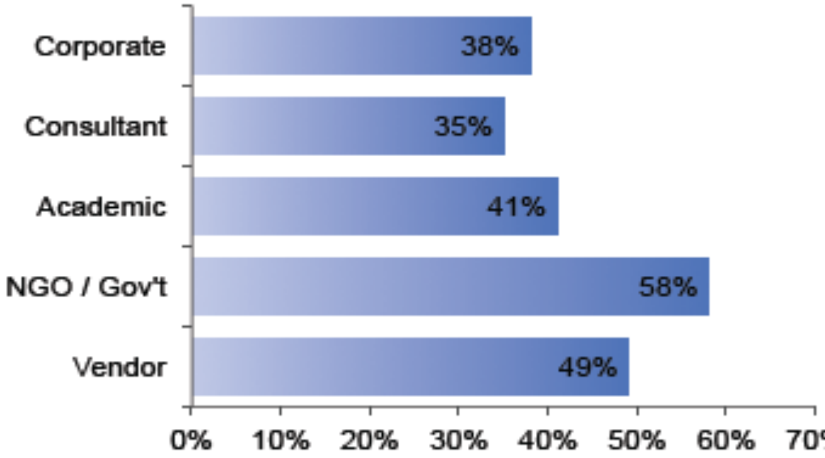
Kaggle | State of ML & Data Science 2021

Importance of Model Explainability

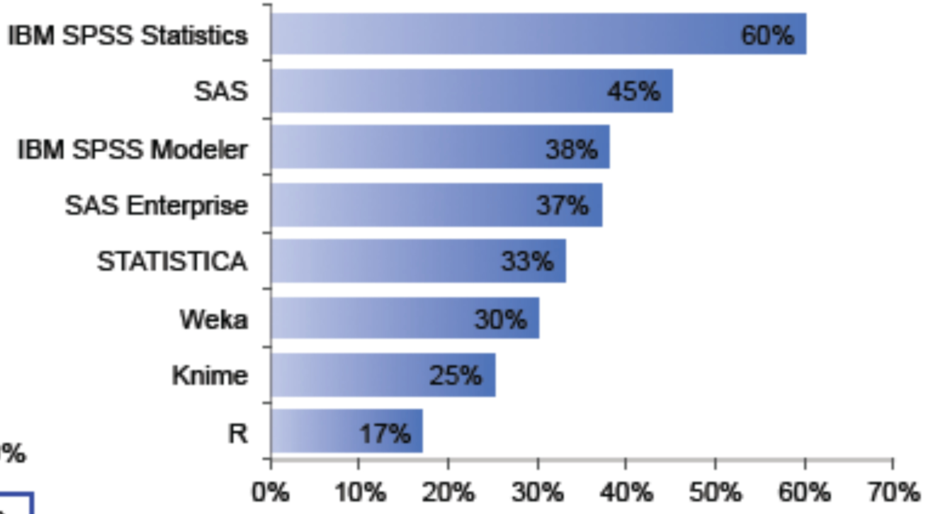
- Model explainability / transparency is important to most data miners.
- It is particularly important to data miners working in NGO / Gov't settings and to data miners who use IBM SPSS Statistics as their primary tool.



Percent indicating Model Explainability is "Extremely Important" by Employment Type



Percent indicating Model Explainability is "Extremely Important" by Primary Tool Used



Question: How important is model explainability / transparency to you?

Goals of this Course

Learn data analysis methods with a focus on

- problems in **management and economics**, and
- **causal inference**, which is particularly challenging when analyzing human decision behavior

Understand and interpret techniques for

- **numerical prediction**
- **classification**
- **clustering and dimensionality reduction**
- **reinforcement learning**

Learn to analyze data with the *Python* programming language

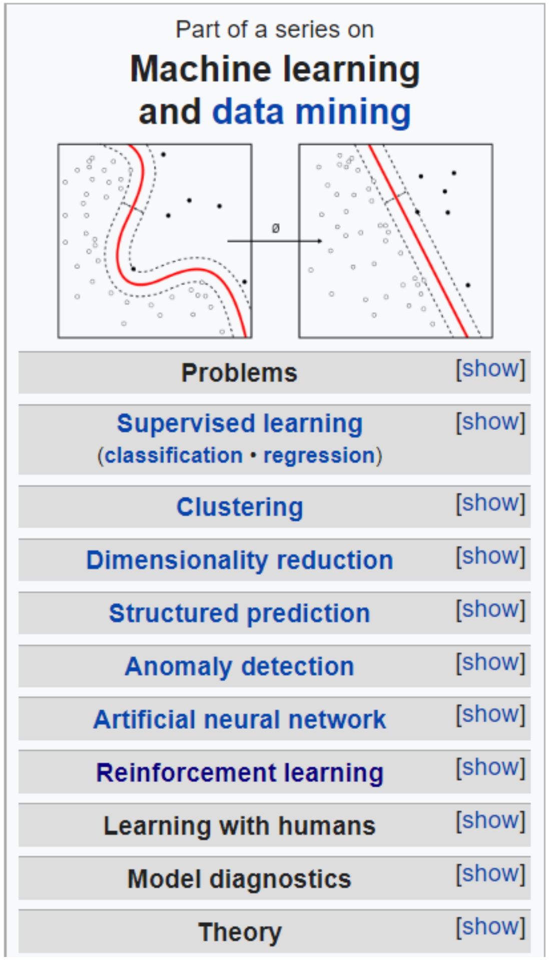
- during the Analytics Cup you analyze data sets as part of the tutorials in small groups

Please note:

- Students in this class have very different backgrounds.
- We expect that you had an introductory course in statistics, algebra, and analysis.
- What is easy for some is difficult and new for others, who lack the prerequisites.

Course Content

- Introduction
- Regression Analysis
- Regression Diagnostics
- Logistic and Poisson Regression
- Naive Bayes and Bayesian Networks
- Decision Tree Classifiers
- Data Preparation and Causal Inference
- Model Selection and Learning Theory
- Ensemble Methods and Clustering
- Dimensionality Reduction
- Convex Optimization
- Neural Networks
- Reinforcement Learning



Part of a series on

Machine learning and data mining

Problems [show]

Supervised learning (classification • regression) [show]

Clustering [show]

Dimensionality reduction [show]

Structured prediction [show]

Anomaly detection [show]

Artificial neural network [show]

Reinforcement learning [show]

Learning with humans [show]

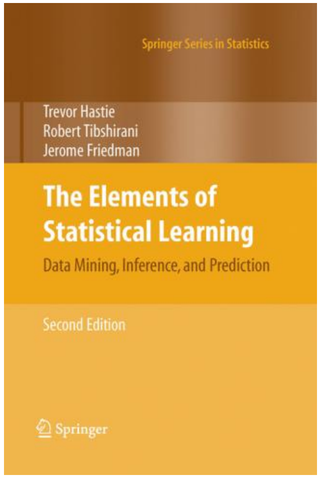
Model diagnostics [show]

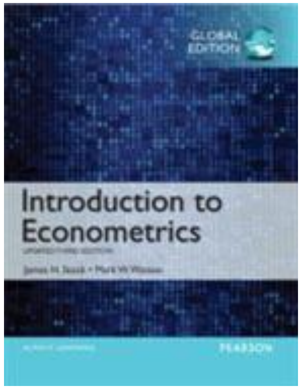
Theory [show]

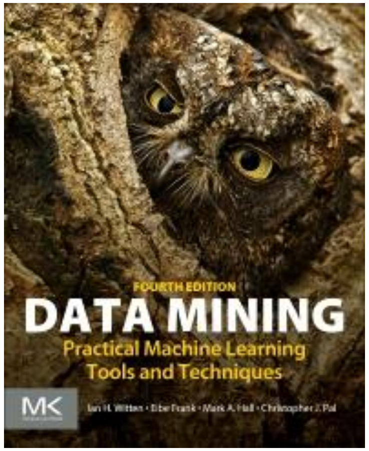
Check out related entries on wikipedia.org!

Primary Literature

- **Introduction to Econometrics**
  - James H. Stock and Mark W. Watson
- **The Elements of Statistical Learning**
  - Trevor Hastie, Robert Tibshirani, Jerome Friedman
  - https://web.stanford.edu/~hastie/ElemStatLearn/
- **Data Mining: Practical Machine Learning Tools and Techniques**
  - Ian H. Witten, Eibe Frank, Mark A. Hall, Christopher Pal
  - http://www.cs.waikato.ac.nz/ml/weka/book.html
- **An Introduction to Statistical Learning: With Applications in R**
  - Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani
  - http://www-bcf.usc.edu/~gareth/ISL/


Springer Series in Statistics
Trevor Hastie
Robert Tibshirani
Jerome Friedman
The Elements of Statistical Learning
Data Mining, Inference, and Prediction
Second Edition
Springer




FOURTH EDITION
DATA MINING
Practical Machine Learning Tools and Techniques
MK
Ian H. Witten • Eibe Frank • Mark A. Hall • Christopher J. Pal

Parts of the slides have kindly been provided by
Prof. Dr. Gregory Piatetsky-Shapiro and Prof. Gary Parker (Univ. of Connecticut).

This Course is Available to Students from ...

- **BSc** Information Systems
- MSc Informatics, MSc Games Engineering, MSc Data Engineering & Analytics
- MSc Management and Technology
- MSc Mathematics, MSc Mathematics in Operations Research

Students from IN, GE, and DE&A can choose one class in Analytics and one class in Machine Learning:

Analytics

•Data Mining, IN2030, 2V, WS, Prof. Runkler
•Business Analytics & Machine Learning, IN2028, **2V+2Ü**, WS, Prof. Bichler
•Data Analysis and Visualization in R, IN2339, 2V+4Ü, WS, Prof. Gagneur

Machine Learning

•Statistical Modeling & Machine Learning, IN2332, 4V+4Ü, SS, Prof. Gagneur
•Machine Learning, IN2064, 4V+2Ü, WS, Prof. Günnemann

Remarks

This course provides an introduction to data analysis with a **focus on methods** used to solve prediction problems in economics and management.

After this course you should be able to analyze mid-sized data sets, perform regression and classification tasks with wide-spread methods, and be able to interpret the results.

This course has **prerequisites** *[see next slide]*. If you don‘t feel you have the necessary background, please choose another class. **A prior statistics course covering estimation, hypothesis testing, and linear regression is essential.**

Slides and textbooks:

– Some students want the slides to be self-contained and demand that everything discussed in class can be found on the slides. Some students want lean slides with not too much text.

– We try to cover the main topics relevant for the final exam on the slides, but urge you to use a **textbook** that suits your background. We provide recommendations that are widely used in related classes.

– Each class can take up to **two hours**!

Prerequisites

In the initial classes of this course you need:

MA9712: Statistics for Business Administration OR
IN0018: Discrete Probability Theory

- probability calculus, random variables, statistical inference (hypothesis tests, confidence intervals, estimation, **simple linear regression**)

In later classes:

MA0901: Linear Algebra for Informatics

- vector and matrix calculus, vector spaces, determinants, eigenvalues

MA0902: Analysis for Informatics

- differentiation, integration, differential calculus of functions of several variables (gradients, Hesse matrix)

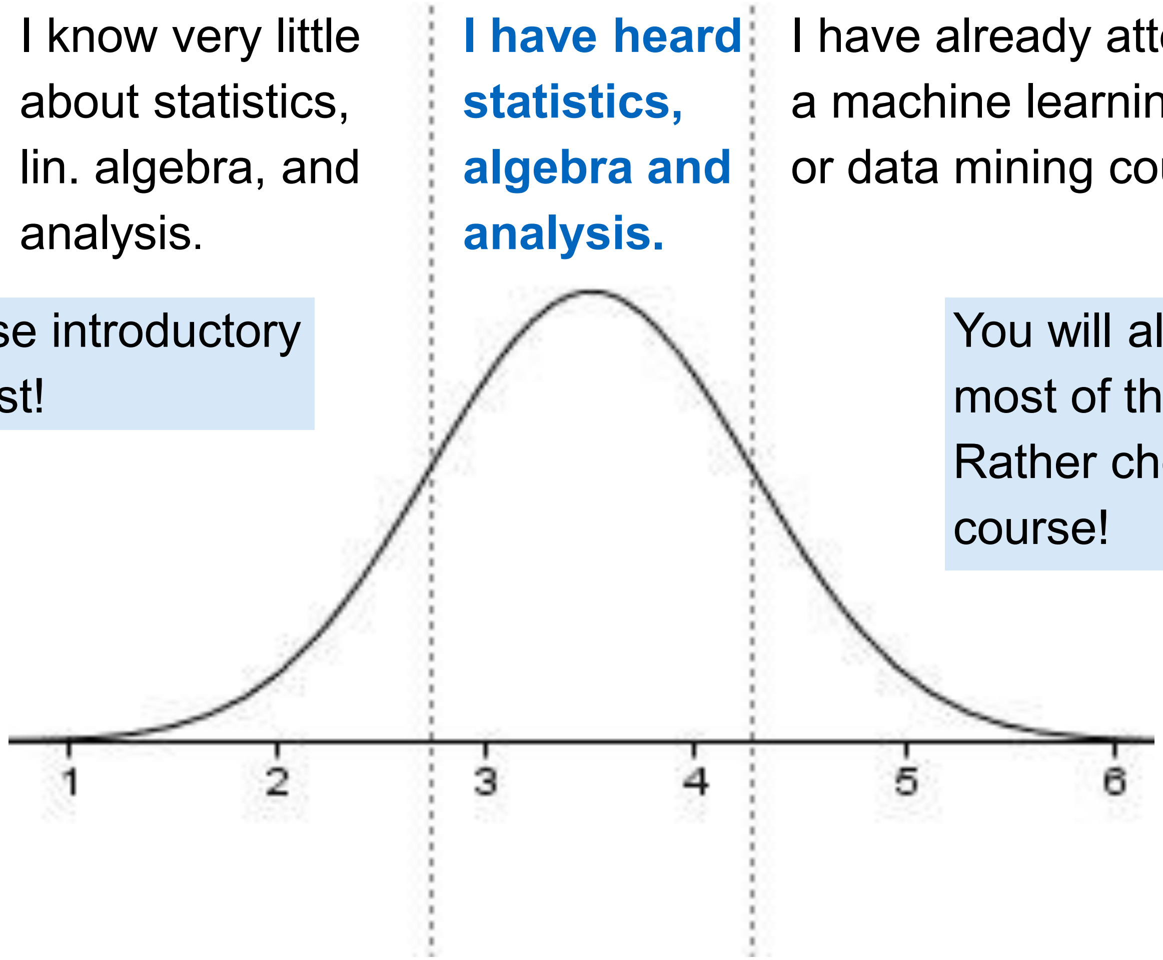
This Course is for …

I know very little about statistics, lin. algebra, and analysis.

I have heard statistics, algebra and analysis.

I have already attended a machine learning or data mining course.

Thake these introductory courses first!



You will already know most of the material. Rather choose another course!

Agenda for Today

1. Understand what this course is all about
2. **Learn about organization, grading, and tutor groups**
3. Refresh basic statistical concepts

Agenda for Today

1. Understand what this course is all about
2. Learn about organization, grading, and tutor groups
3. Homework: **Refresh basic statistical concepts!**

In the first tutorials, we will recap important concepts from inferential statistics and introduce the *Python* programming language, required for the rest of the course.

Check out http://onlinestatbook.com as an online source.

For this week please revisit the concepts on the following slides. Slides are only meant as a refresher.

Statistics

Descriptive statistics can be used to summarize the data, either numerically or graphically, to describe the sample (e.g., mean and standard deviation).

Inferential statistics is used to model patterns in the data, accounting for randomness and drawing inferences about the larger population. These inferences may take the form of

- estimates of numerical characteristics (estimation),
- answers to yes/no questions (hypothesis testing),
- forecasting of future observations (forecasting),
- descriptions of association (correlation), or
- modeling of relationships (regression).

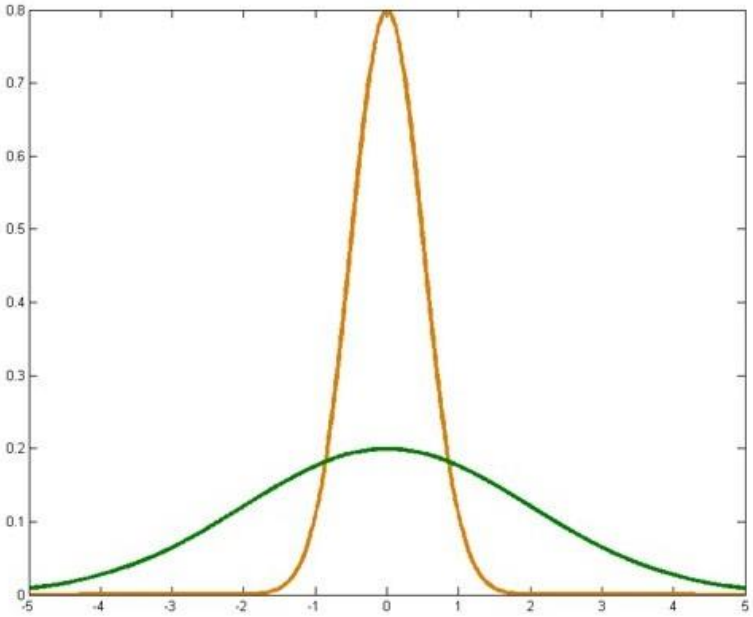
Random Variables

X is a random variable if it represents a random draw from some population and is associated with a probability distribution.

- a **discrete** random variable can take on only selected values (e.g., Binomial or Poisson distributed)
- a **continuous** random variable can take on any value in a real interval (e. g., uniform, Normal or Chi-Square distributions)

For example, a Normal distribution, with mean μ and variance σ^2 is written as $N(\mu, \sigma^2)$ has a probability density function (pdf) of:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



The Standard Normal

Any random variable can be "standardized" by subtracting the mean, μ, and dividing by the standard deviation, σ , so $E(Z) = 0$, $Var(Z) = 1$.

Thus, the standard normal, $N(0,1)$, has the probability density function (pdf):

$$\boxed{\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}}$$

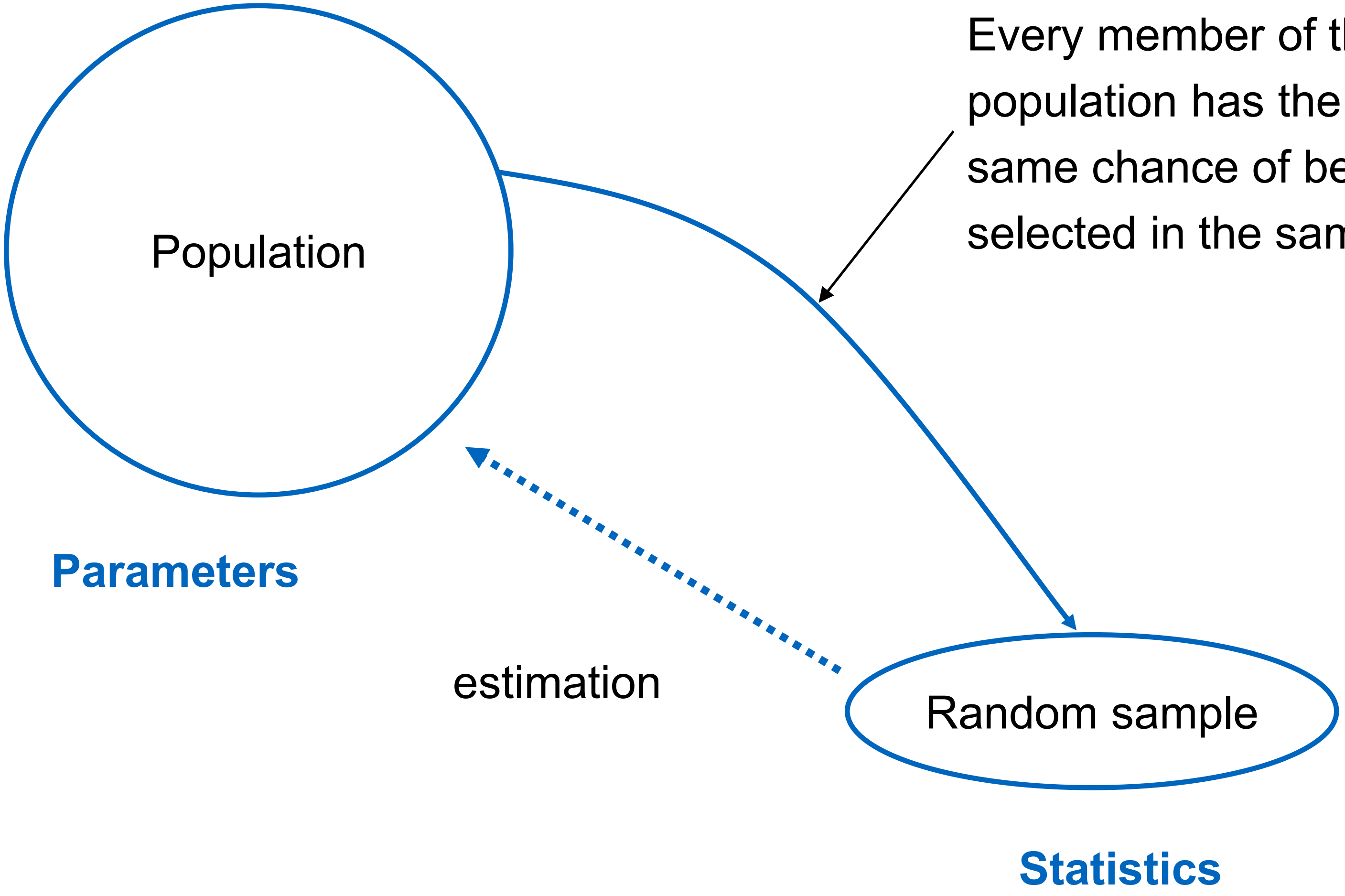
For a pdf, $f(x)$, where $f(x)$ is $P(X = x)$, the cumulative distribution function (cdf), $F(x)$, is $P(X \le x)$; $P(X > x) = 1 - F(x) = P(X < -x)$

For the standard normal, $\varphi(z)$, the cdf is:

$$\Phi(z) = P(Z \le z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{\frac{-t^2}{2}} dt$$

Note: We'll use the term "Gaussian" as synonym for a Normal distribution.

Statistical Estimation



Every member of the population has the same chance of being selected in the sample.

Expected Value of X: Population Mean $E(X)$

The expected value of a probability weighted average of $X,\ E(X),$ is the mean or expected value of the distribution of X, denoted by $\mu_{X.}$

Let $f(x_i)$ be the (discrete or continuous) probability that $X = x_i$, then

$$\mu_X = E(X) = \overset{\text{discrete}}{\sum_{i=1}^{n}} x_i f(x_i)\ or \overset{\text{continuous}}{\int_{-\infty}^{\infty}} xf(x)dx$$

Example: Expected Value

Students were surveyed and told to pick the number of hours that they play online games each day. The probability distribution is given below.

# of Hours *x*	Probability *P*(*x*)
0	.3
1	.4
2	.2
3	.1

Compute a “weighted average” by multiplying each possible time value by its probability and then adding the products.

$$E(X) = 0(.3) + 1(.4) + 2(.2) + 3(.1) = 1.1$$

Random Samples and Sampling

For a random variable X, repeated draws from the same population can be labeled as $X_1, X_2, \ldots, X_n$.

If every combination of n sample points has an equal chance of being selected, this is a random sample.

A random sample is a set of independent, identically distributed (i.i.d) random variables.

Examples of Estimators

- suppose we want to estimate the **population mean**

- suppose we use the formula for $E(X)$, but substitute $1/n$ for $f(x_i)$ as the probability weight since each point has an equal chance of being included in the sample, then we can calculate the sample mean:

$$\bar{X} = \frac{1}{n}\sum_{1}^{n} X_i$$

- $\bar{X}$ describes the random variable for the **arithmetic mean of the sample**, while $\bar{x}$ is the mean of a particular realization of a sample

Estimators Should be Unbiased

An estimator (e.g., the arithmetic sample mean) is a statistic (a function of the observable sample data) that is used to estimate an unknown population parameter (e.g., the expected value).

We want the estimator to be right, on average, i.e. unbiased.

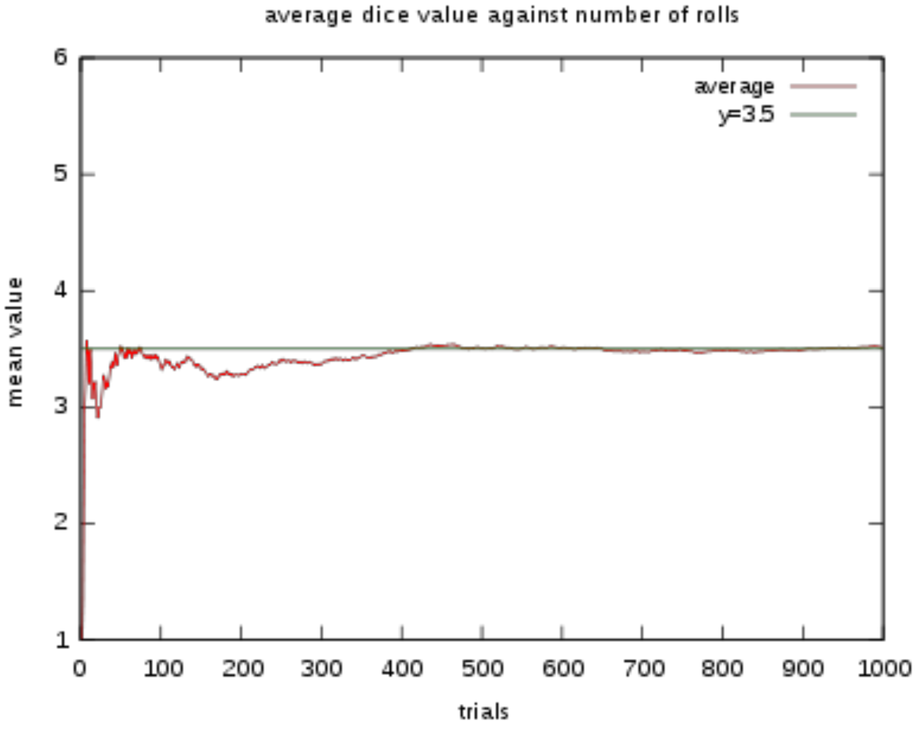
In our case, the sample mean $\bar{X}$ should be an unbiased estimator for the population mean μ_X:

$$E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i)$$

$$= \frac{1}{n}\sum_{i=1}^{n} \mu_X = \frac{1}{n} n\mu_X = \mu_X$$

Rolling a Dice

Expected value of 3.5 as the number of die rolls grows.



According to the **law of large numbers**, the sample mean converges to the expected value of the population distribution.

Standard Error of the Sample Mean

$$Var(\bar{X}) = Var\left(\frac{1}{n}\sum_{i=0}^{n}(X_i)\right) = \frac{1}{n^2}\sum_{i=1}^{n}Var(X_i)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}$$

$$\sigma_{\bar{X}} = SD(\bar{X}) = \sqrt{Var(\bar{X})} = \frac{\sigma}{\sqrt{n}}$$

+const in Var → not relevant

scalar in Var → squared

Rule: $Var[aX + \overset{\times}{b}] = a^2 Var[X]$

The **standard error of the sample mean** is an estimate of how far the sample mean is likely to differ from the population mean. This means, the standard error of the mean tells you how accurate your estimate of the mean is likely to be.

The **standard deviation** of the sample is the degree to which individuals within the sample differ from the sample mean.

The Central Limit Theorem

The **central limit theorem** states that the standardized average of any population of i.i.d. random variables X_i with mean μ_X and variance σ^2 is asymptotically $\sim N(0,1)$ as n goes to infinity.

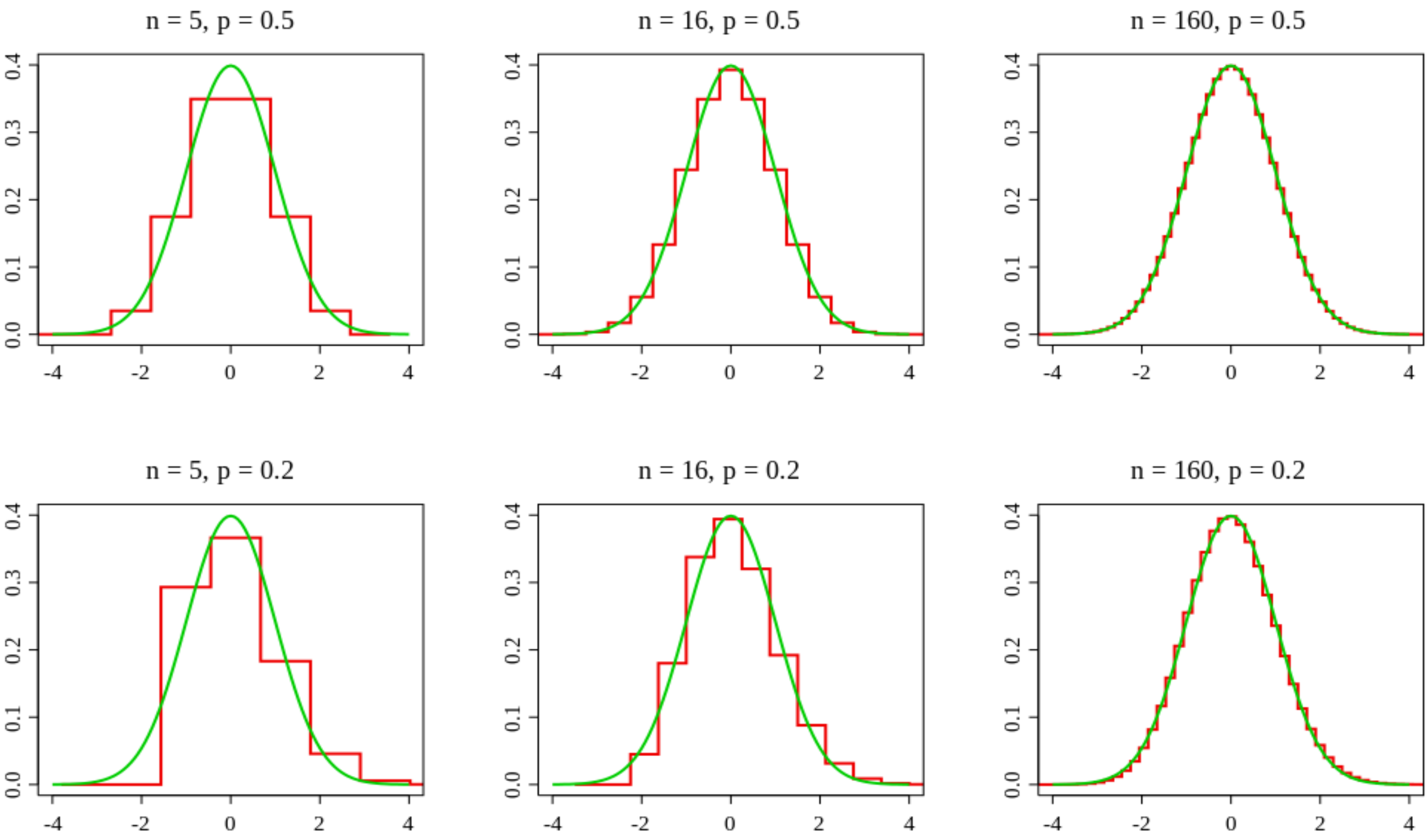
$n \gtrsim 30$, usually

$$\boxed{Z = \frac{\bar{X} - \mu_X}{\sigma/\sqrt{n}} \sim N(0,1)}$$

This means, when independent random variables are added, their normalized sum tends toward a normal distribution even if the original variables themselves are not normally distributed.

independent n samples, with probability p
and 1-p

Binomial Distributions and the Normal Distribution



Source: Wikipedia

Summary: Sampling Distribution of the Mean

We can say something about the distribution of sample statistics (such as the sample mean).

The sample mean is a random variable, and consequently it has its own distribution and variance.

The distribution of sample means for different samples of a population is centered on the population mean.

The mean of the sample means is equal to the population mean.

If the population is normally distributed or when the sample size is large, sample means are distributed normally (Central Limit Theorem).

Question

What is the probability that a sample of 100 **randomly selected** elements with a mean of 300 or more gets selected if the true population mean is 288 and the population standard deviation is 60?

$P(\bar{X} > 300)$ $\mu = 288$ $\sigma = 60$ $n = 100$

$$P(\bar{X} > 300) = P\left(\frac{\bar{X} - 288}{60/\sqrt{100}} > \frac{300-288}{60/\sqrt{100}} \right)$$

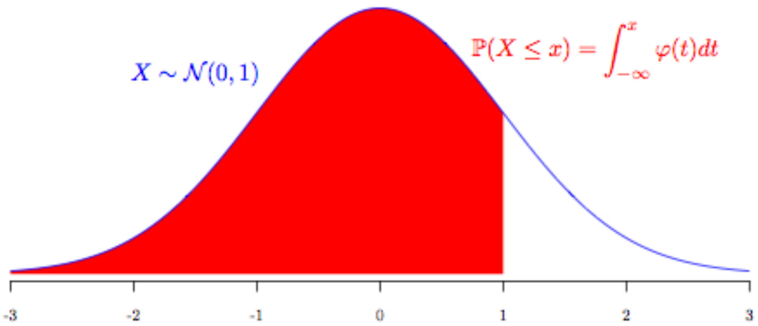
$Z \sim N(0,1)$

$$= P\left(\frac{\bar{X} - 288}{6} \right) > 2)$$

$$= 1 - P(Z \leq 2)$$

$$= 1 - 0.9772$$

$$= 0.0228$$



2.00 2.01 ...

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990