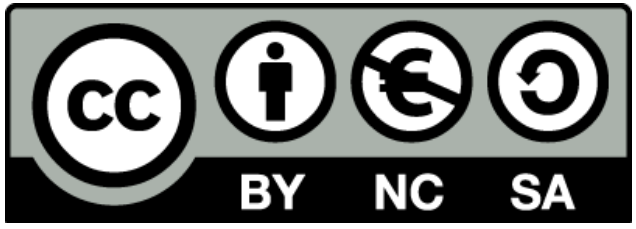
Rail-Data

Prof. Dr. Raphael Pfaff

Fachhochschule Aachen



FH AACHEN
UNIVERSITY OF APPLIED SCIENCES

Abschnitt 1

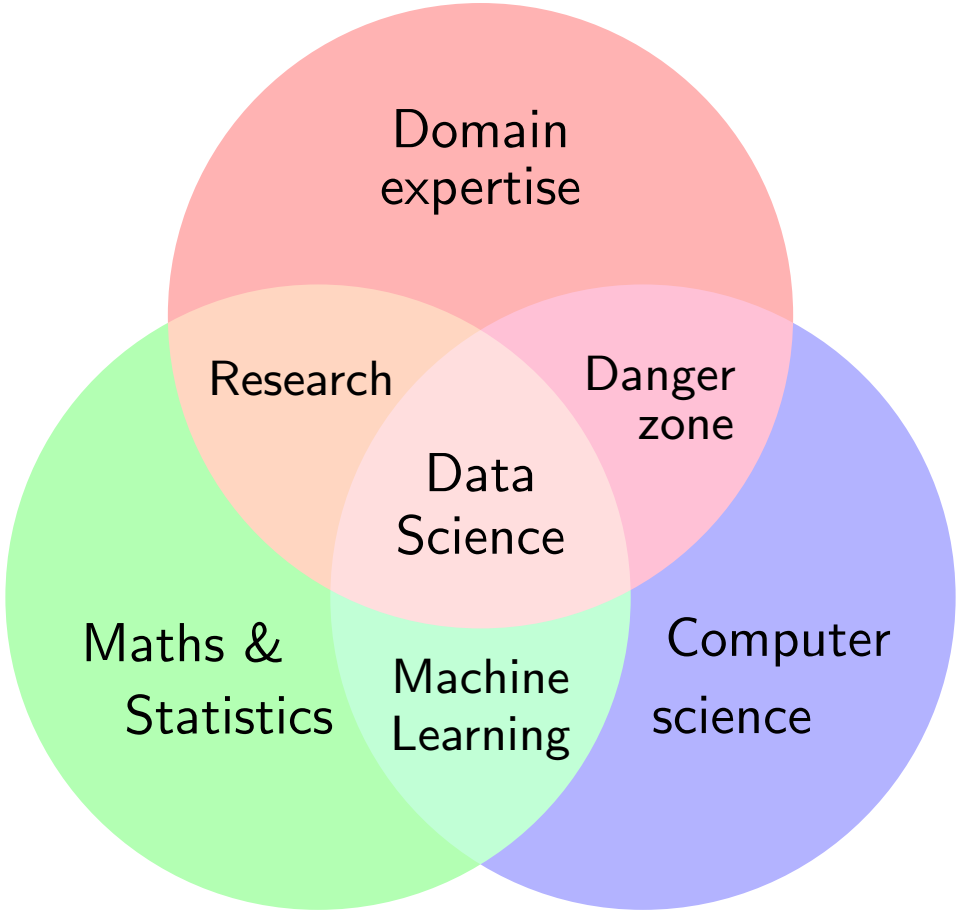
Brief Introduction to Data Science

What is Data Science?

Data science is our means of taming unstructured information and gathering insight. - Matthew Mayo, KDnuggets

- Interdisciplinary field of
  - Systems,
  - Methods and
  - Processs to extract insight or knowledge from data.
- Term coined in 2001, gained popularity in 2010
- Integrates:
  - Data Engineering
  - Scientific Method
  - Mathematics
  - Statistics
  - Advanced Computational Methods
  - Visualisation
  - Hacker Mindset
  - Domain Expertise

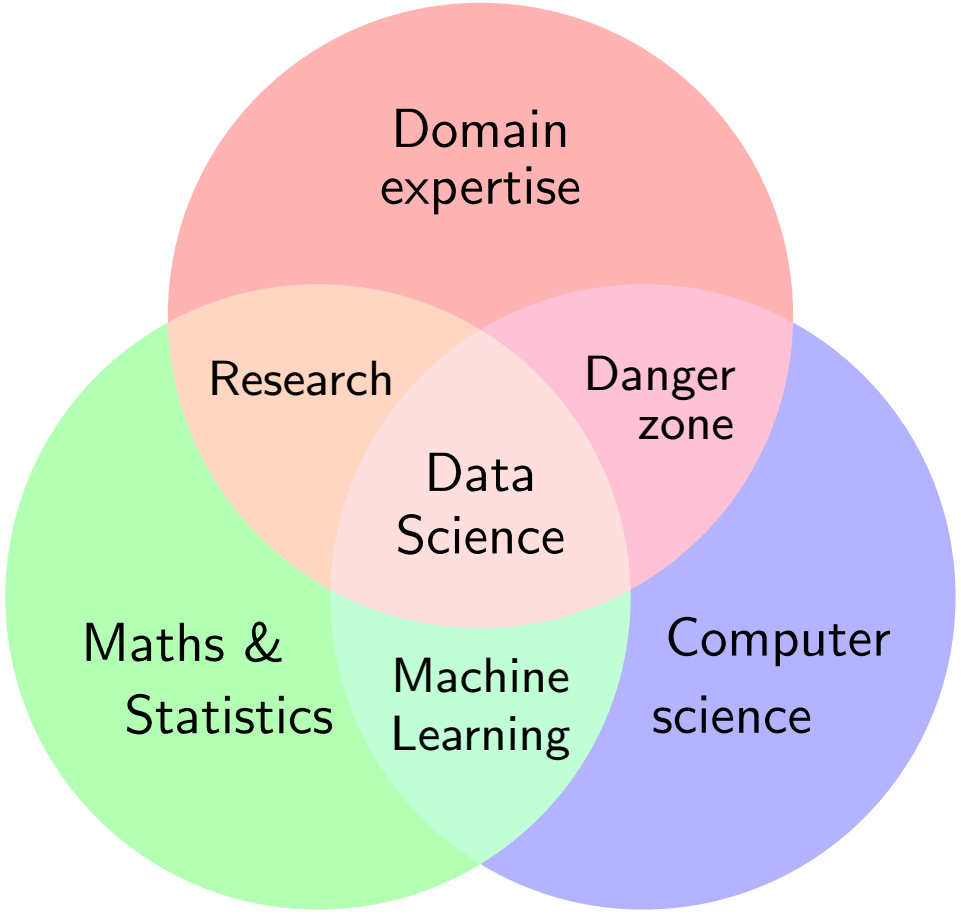
How does Data Science integrate to Mechanical Engineering?



Why Data Sciene?

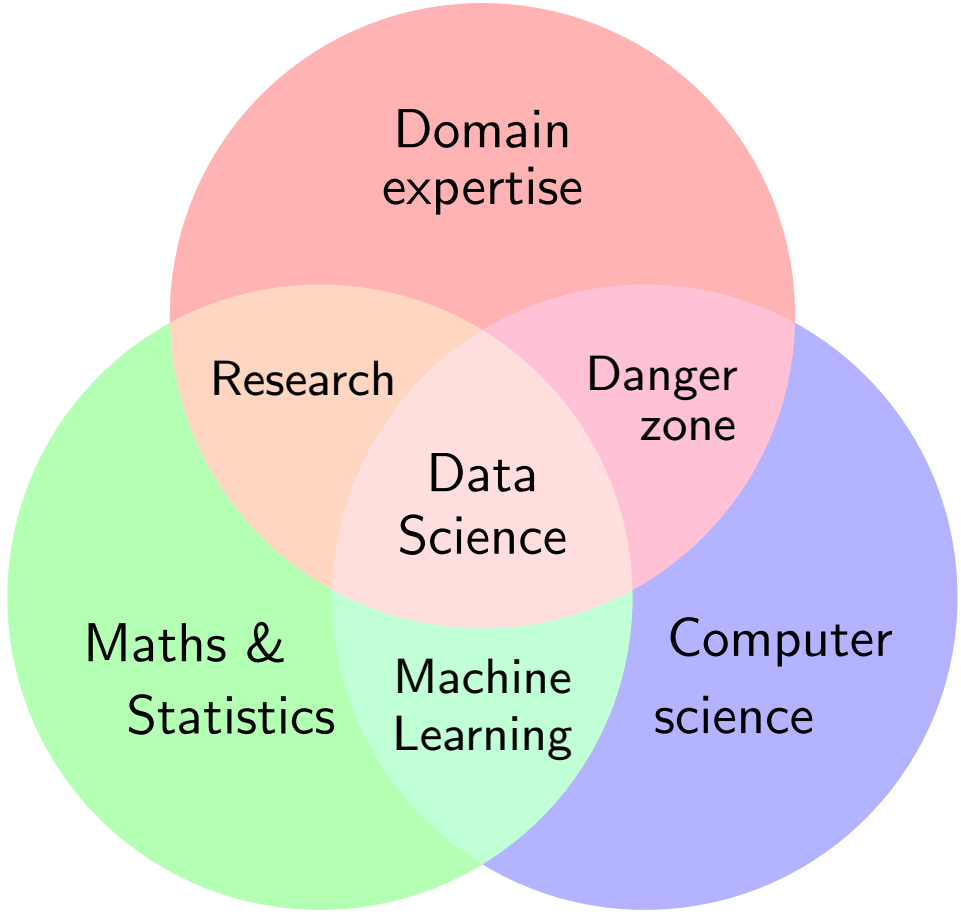
Applying the right tools and techniques, you bring more value!

- Turn data to information
  - Inform decisions
  - Increase insight
- Companies:
  - Collect large amounts of data
  - Do rarely integrate them
  - Frequently decide based on the "gut"

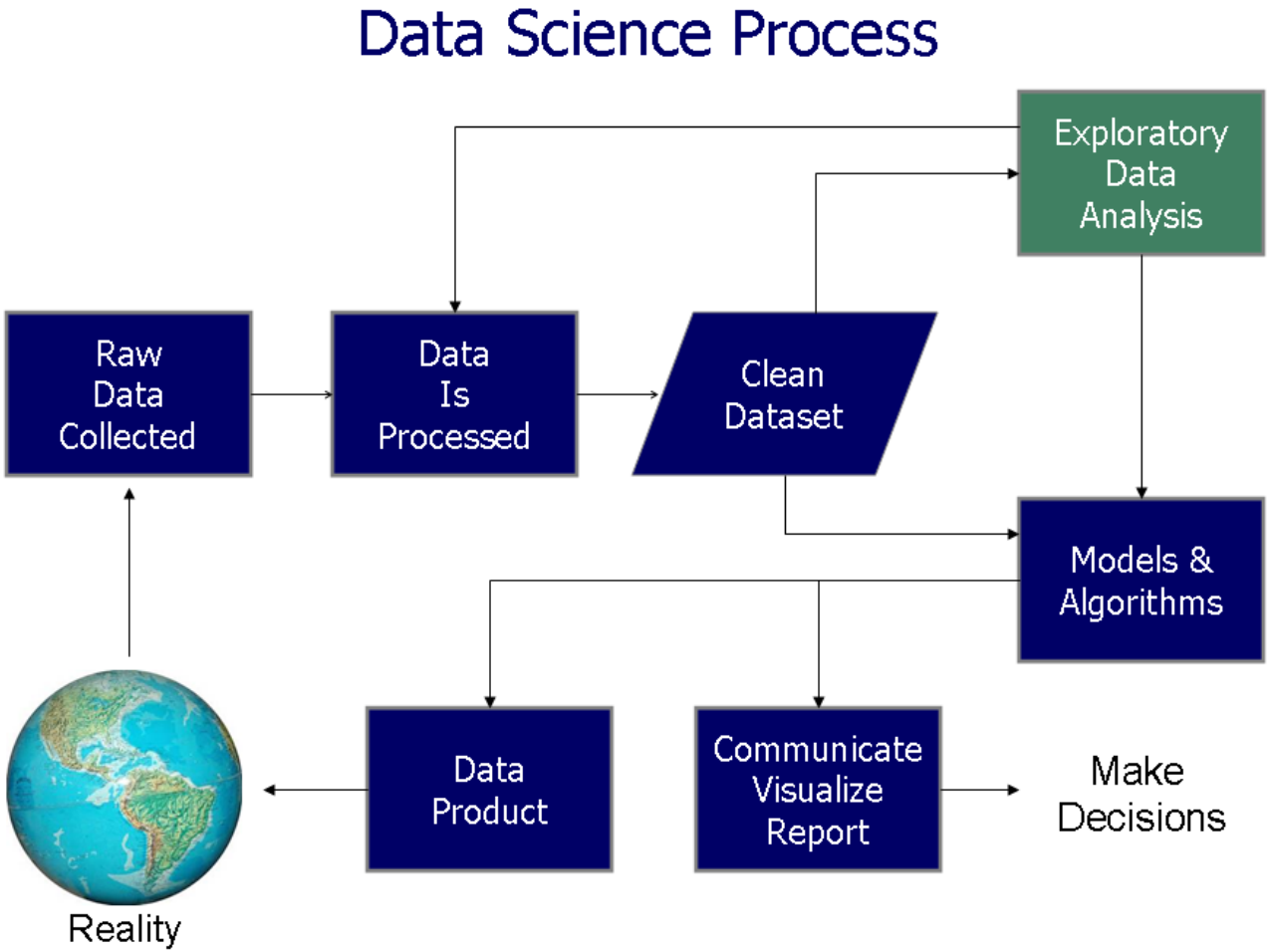


How do you incrase Value with Data Science?

- Improve decision making
  - Empower management
  - Supply data driven evidence
- Identify trends and bring to action
- Challenge your colleagues
- Find opportunities for improvement
- Test decisions
- Understand customers



The data science process



Source: Farcaster/English Wikipedia

How to get started

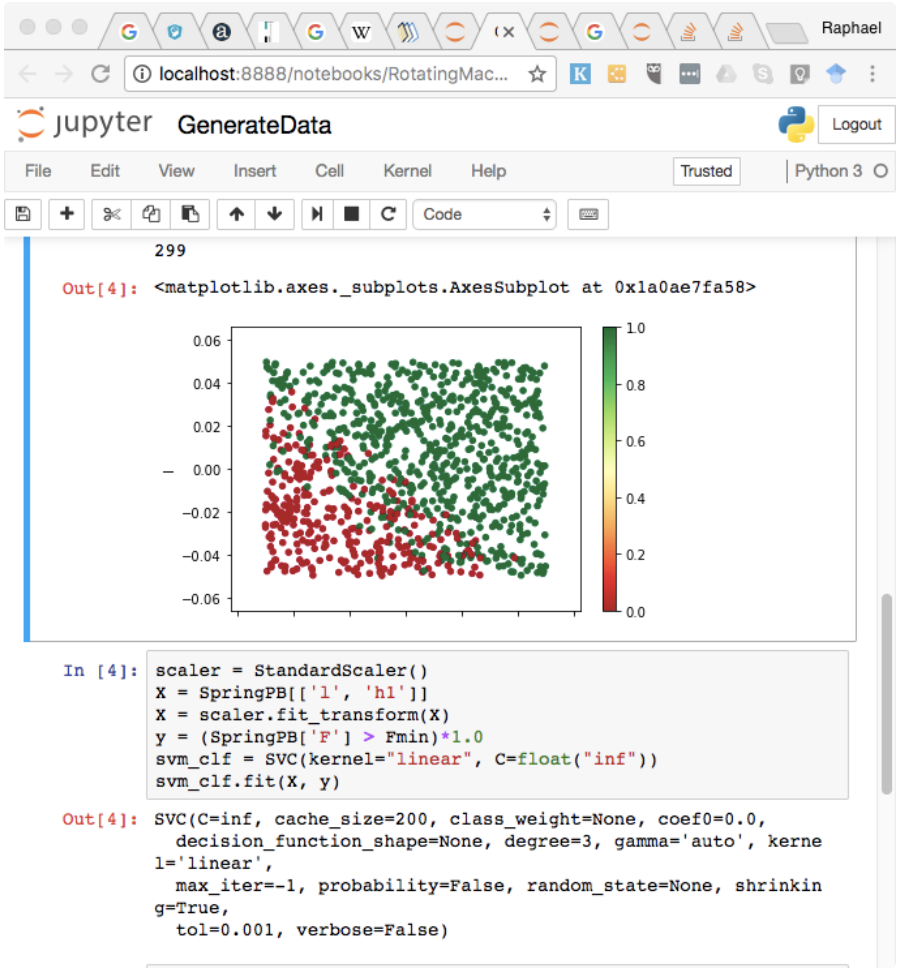
- Set up your system.
  - Install Anaconda to obtain Python/Jupyter
  - Set up a free education account with github.com
  - Install the app for your operating system
- Acquire data: start with popular open data sets. Get your company to make data accessible.
- Ingest and transform: figure out the formats and sizes of your data. Find appropriate ways to import or access them.
- Explore the data. Do you already find patterns from just plotting them?
- Try your "toolbox" of methods (or a subset of it that sounds promising).
- Visualise the results. Make your findings convincing to others: colleagues, managers, customers etc.

Abschnitt 2

Data Science: Tools and Techniques

Tools for Data Scientist

- Programming languages:
  - R
  - Python
  - SAS
  - ...
- Visualisation app:
  - Tableau
- Development Environment (IDE):
  - Jupyter
  - Spyder
- Also potentially:
  - Matlab
  - Scilab
  - ...



For installation, visit:
https://www.anaconda.com/
Data folder:
http://bit.ly/ifv122019

Python Introduction

- Python
  - is interpreted
  - can be run on terminal or IDE
- Indentation is strict, defines program structures (no braces)
- Dynamically typed
  - No variable declaration required
- Syntax
  - Refer to example worksheets
- Data structures
  - List: `l = [1, 2, ä"]`
  - Tuples: `t = (1, 2, ä")`
  - Dictionaries: `d = {ä":1, "b":2}`
  - Sets: `s = set(`

    $1, 2, 3, 4$

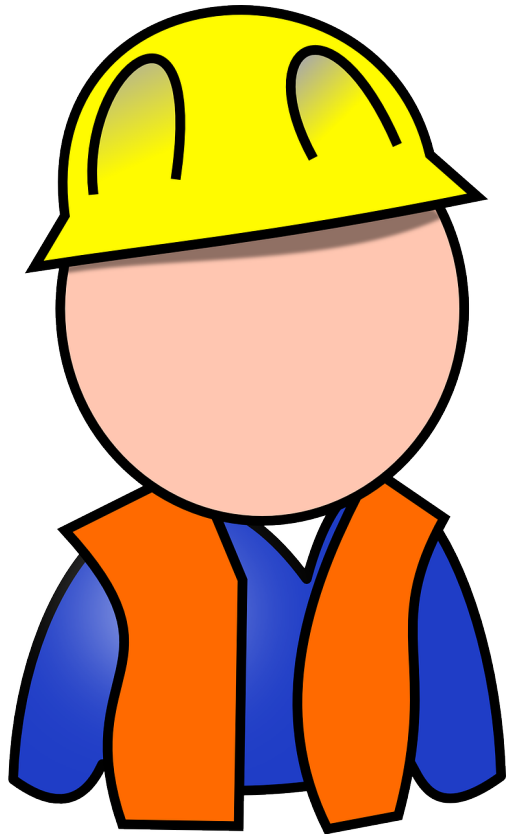
    `)`

Pandas Introduction

- Pandas is a framework for manipulation and handling of series and rectangular data
- Syntax
  - Refer to example worksheets
- Data structures:
  - Series (`pd.Series`(from e.g. array))
  - DataFrame (`pd.DataFrame`(from e.g. dicts))

Get to work!

Load and try `0-Basics.ipynb` (`http://bit.ly/ifv122019`)



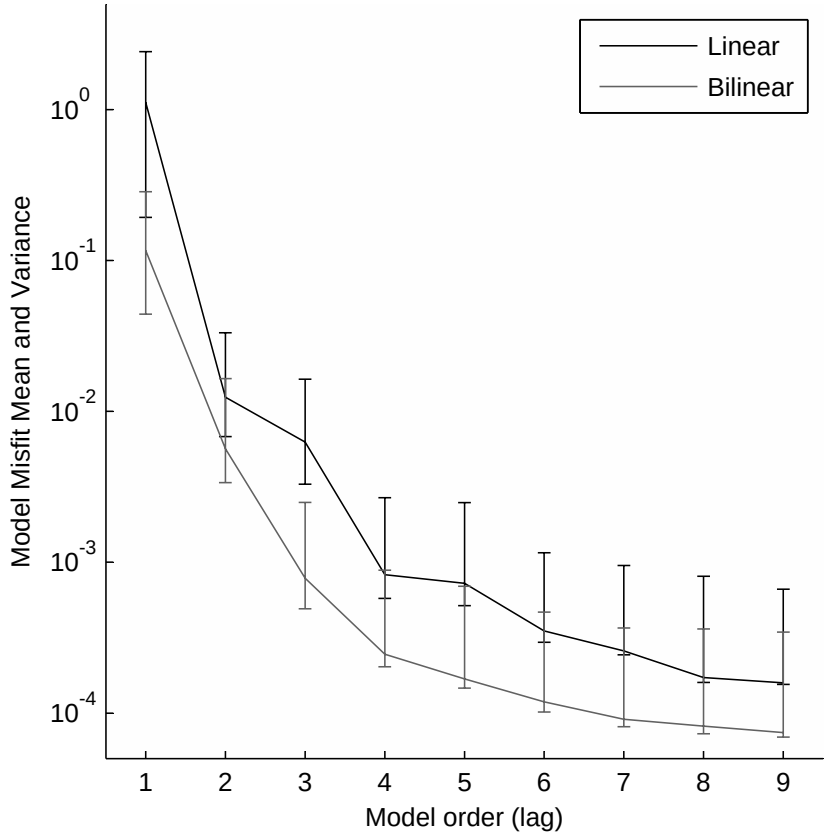
Selected Techniques applied in Data Science

- **Visualisation**
- Regression: Linear, Logistic
- **Density Estimation**
- Confidence Intervals
- Test of Hypotheses
- Pattern Recognition
- Time Series
- **Unsupervised Learning (Clustering)**
- Supervised Learning
- Decision Trees
- **Monte-Carlo-Simulation**
- Bayesian Statistics
- **Principal Component Analysis**
- **Support Vector Machines**

Visualisation approaches

Likely, you will all know Excel plots - there is much more to it...

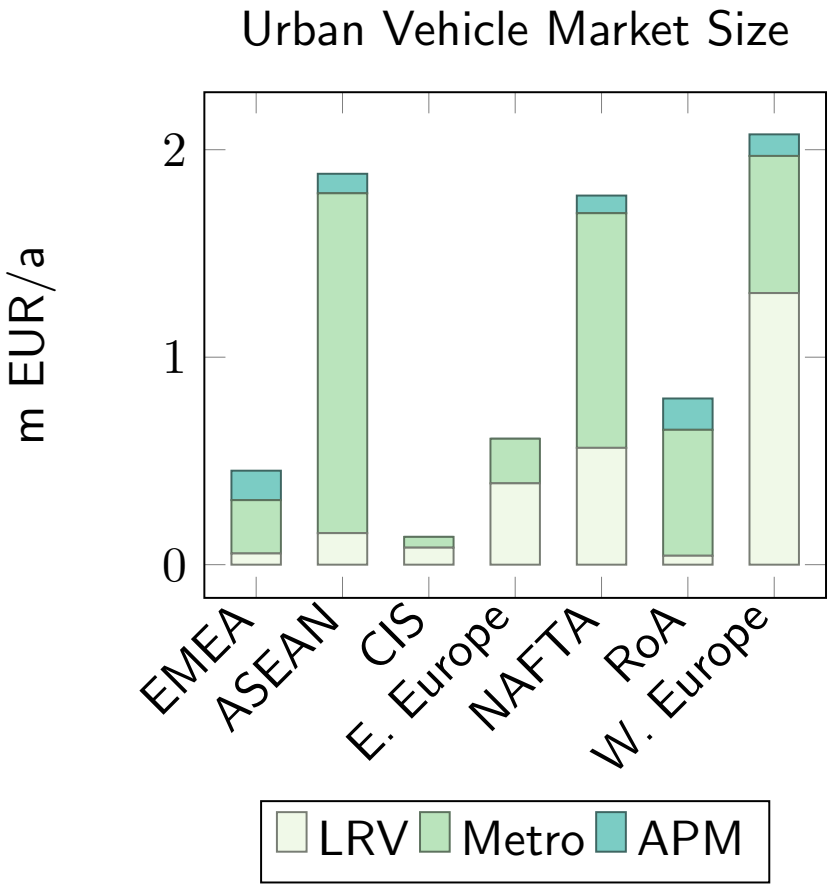
- Line plot with / w/o confidence
- Bar plots (stacked/grouped)
- Histogram
- Scatter plot
  - Blob sizes
  - Colors
- Box whisker
- Bubble chart
- Geospatial plots
- Surface plot
- Heat map
- Tree map



Visualisation approaches

Likely, you will all know Excel plots - there is much more to it...

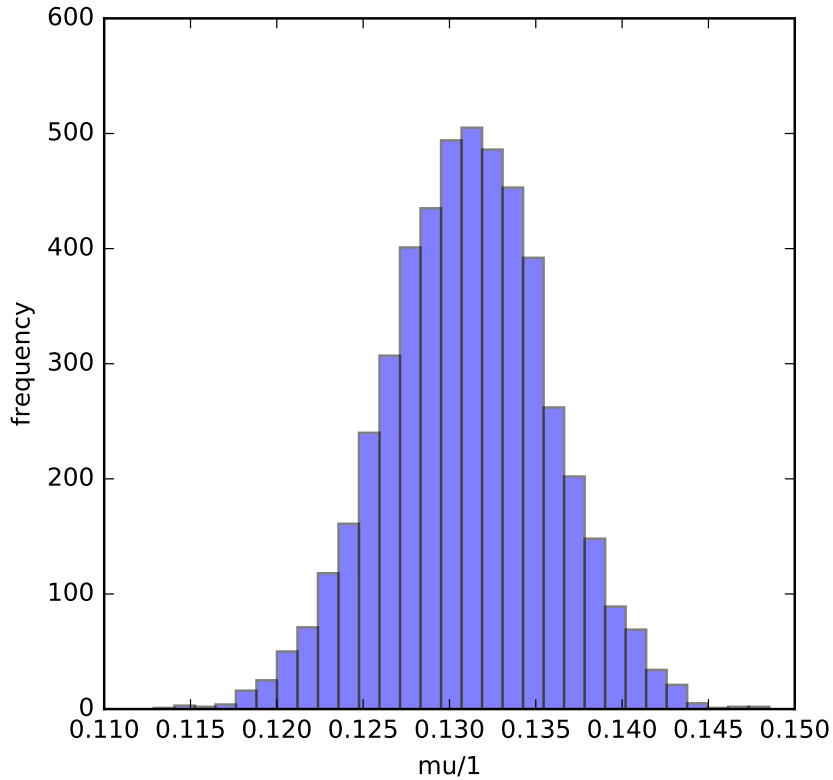
- Line plot with / w/o confidence
- Bar plots (stacked/grouped)
- Histogram
- Scatter plot
  - Blob sizes
  - Colors
- Box whisker
- Bubble chart
- Geospatial plots
- Surface plot
- Heat map
- Tree map



Visualisation approaches

Likely, you will all know Excel plots - there is much more to it...

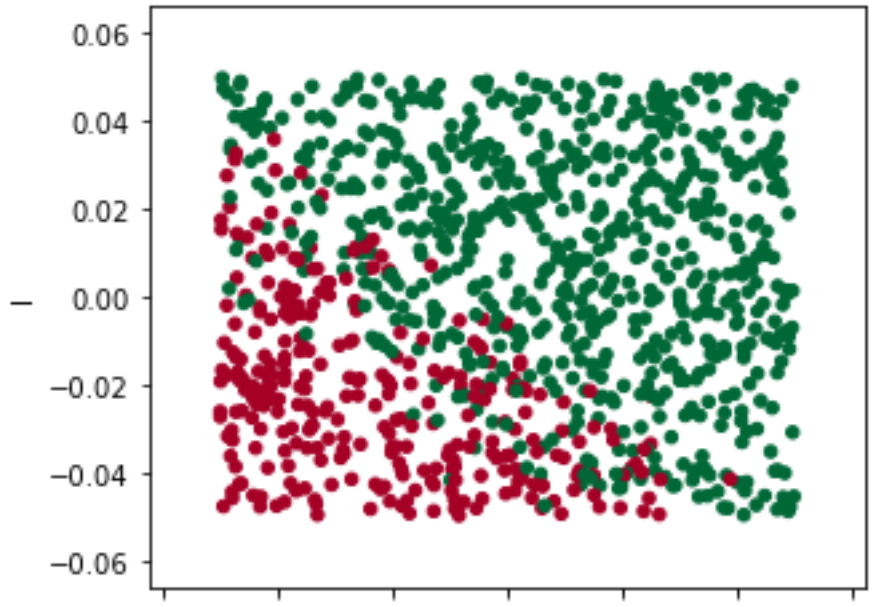
- Line plot with / w/o confidence
- Bar plots (stacked/grouped)
- Histogram
- Scatter plot
  - Blob sizes
  - Colors
- Box whisker
- Bubble chart
- Geospatial plots
- Surface plot
- Heat map
- Tree map



Visualisation approaches

Likely, you will all know Excel plots - there is much more to it...

- Line plot with / w/o confidence
- Bar plots (stacked/grouped)
- Histogram
- Scatter plot
  - Blob sizes
  - Colors
- Box whisker
- Bubble chart
- Geospatial plots
- Surface plot
- Heat map
- Tree map



Visualisation approaches

Likely, you will all know Excel plots - there is much more to it...

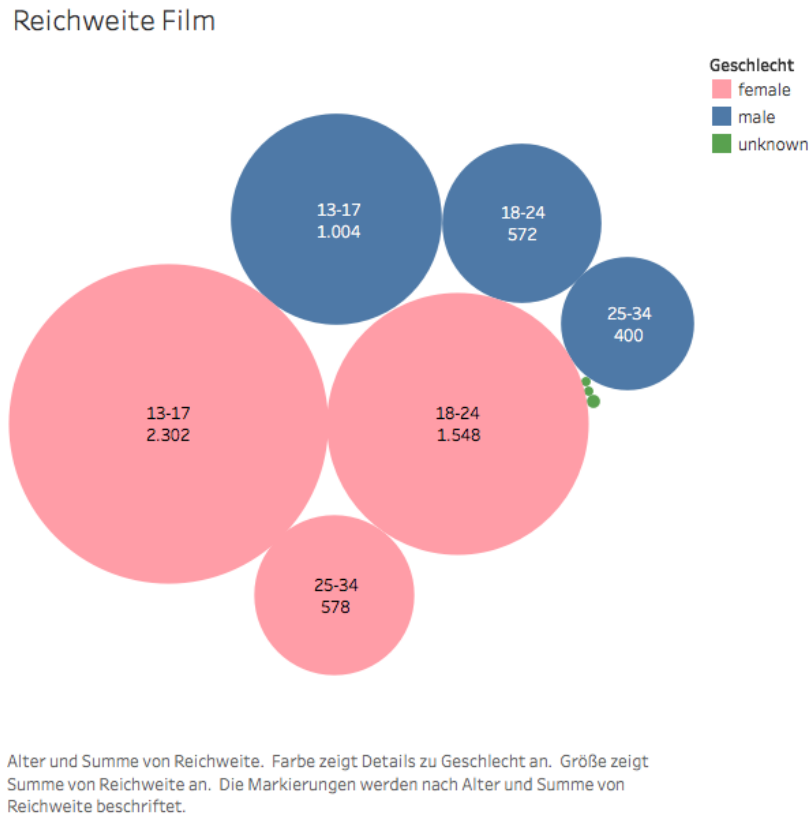
- Line plot with / w/o confidence
- Bar plots (stacked/grouped)
- Histogram
- Scatter plot
  - Blob sizes
  - Colors
- Box whisker
- Bubble chart
- Geospatial plots
- Surface plot
- Heat map
- Tree map

Intentionally left blank

Visualisation approaches

Likely, you will all know Excel plots - there is much more to it...

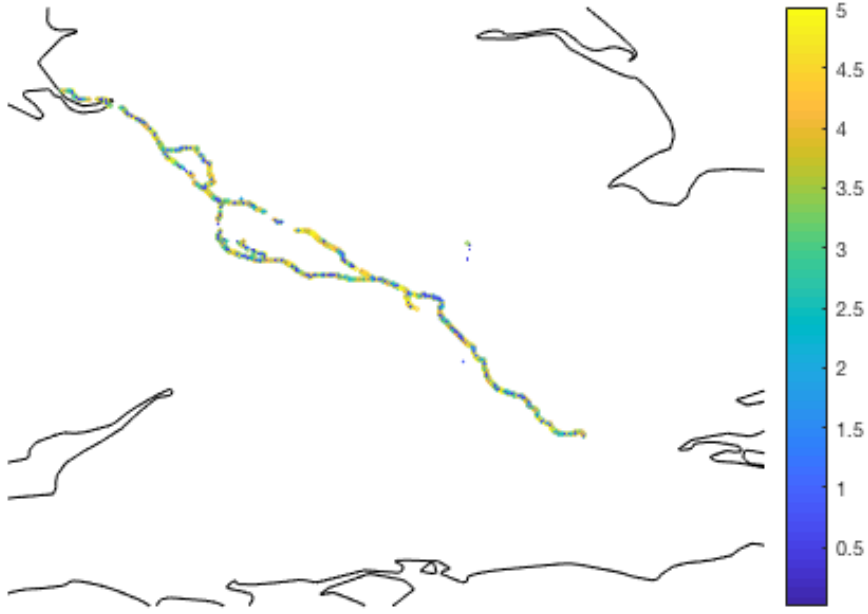
- Line plot with / w/o confidence
- Bar plots (stacked/grouped)
- Histogram
- Scatter plot
  - Blob sizes
  - Colors
- Box whisker
- Bubble chart
- Geospatial plots
- Surface plot
- Heat map
- Tree map



Visualisation approaches

Likely, you will all know Excel plots - there is much more to it...

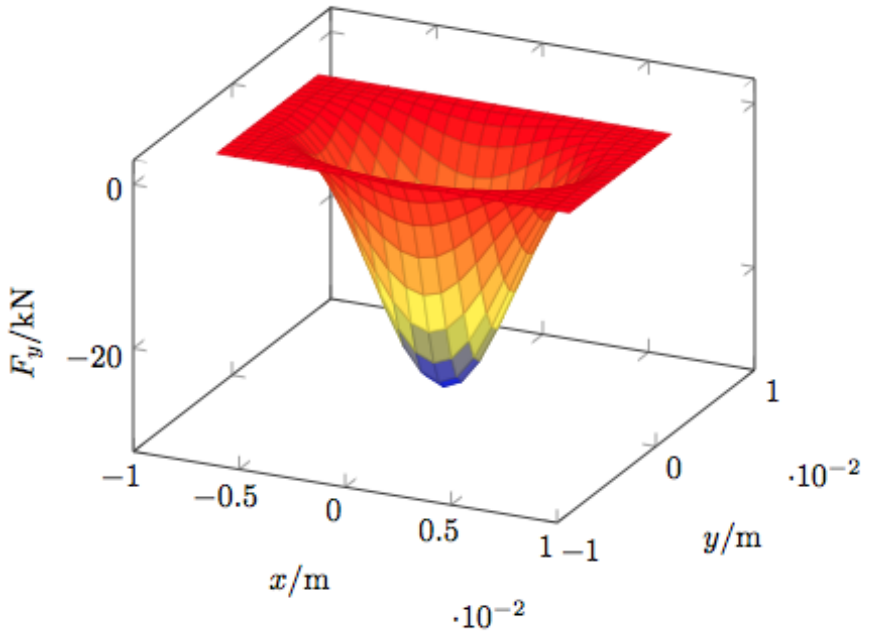
- Line plot with / w/o confidence
- Bar plots (stacked/grouped)
- Histogram
- Scatter plot
  - Blob sizes
  - Colors
- Box whisker
- Bubble chart
- Geospatial plots
- Surface plot
- Heat map
- Tree map



Visualisation approaches

Likely, you will all know Excel plots - there is much more to it...

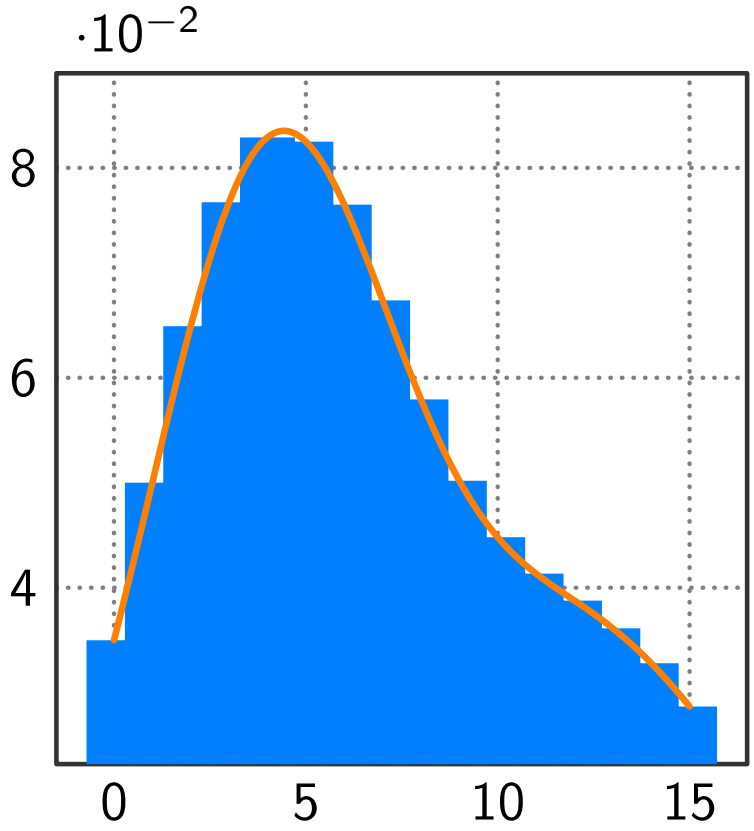
- Line plot with / w/o confidence
- Bar plots (stacked/grouped)
- Histogram
- Scatter plot
  - Blob sizes
  - Colors
- Box whisker
- Bubble chart
- Geospatial plots
- Surface plot
- Heat map
- Tree map



Density estimation

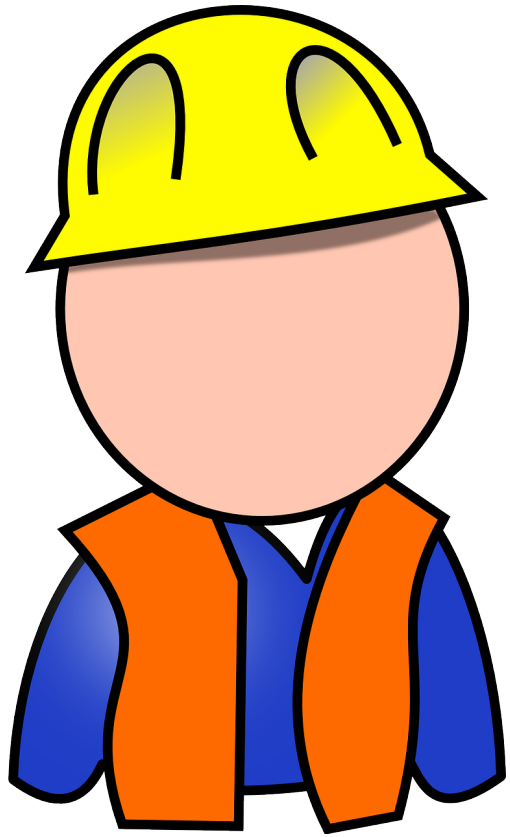
Estimate underlying probability density function from observed data

- Analysis of distribution properties from sampled data
- Fit an appropriate kernel to the samples



Get to work!

Load and try `1-Visualisation.ipynb` (http://bit.ly/ifv122019)



Machine Learning

Using MIT OpenCourseware Slides

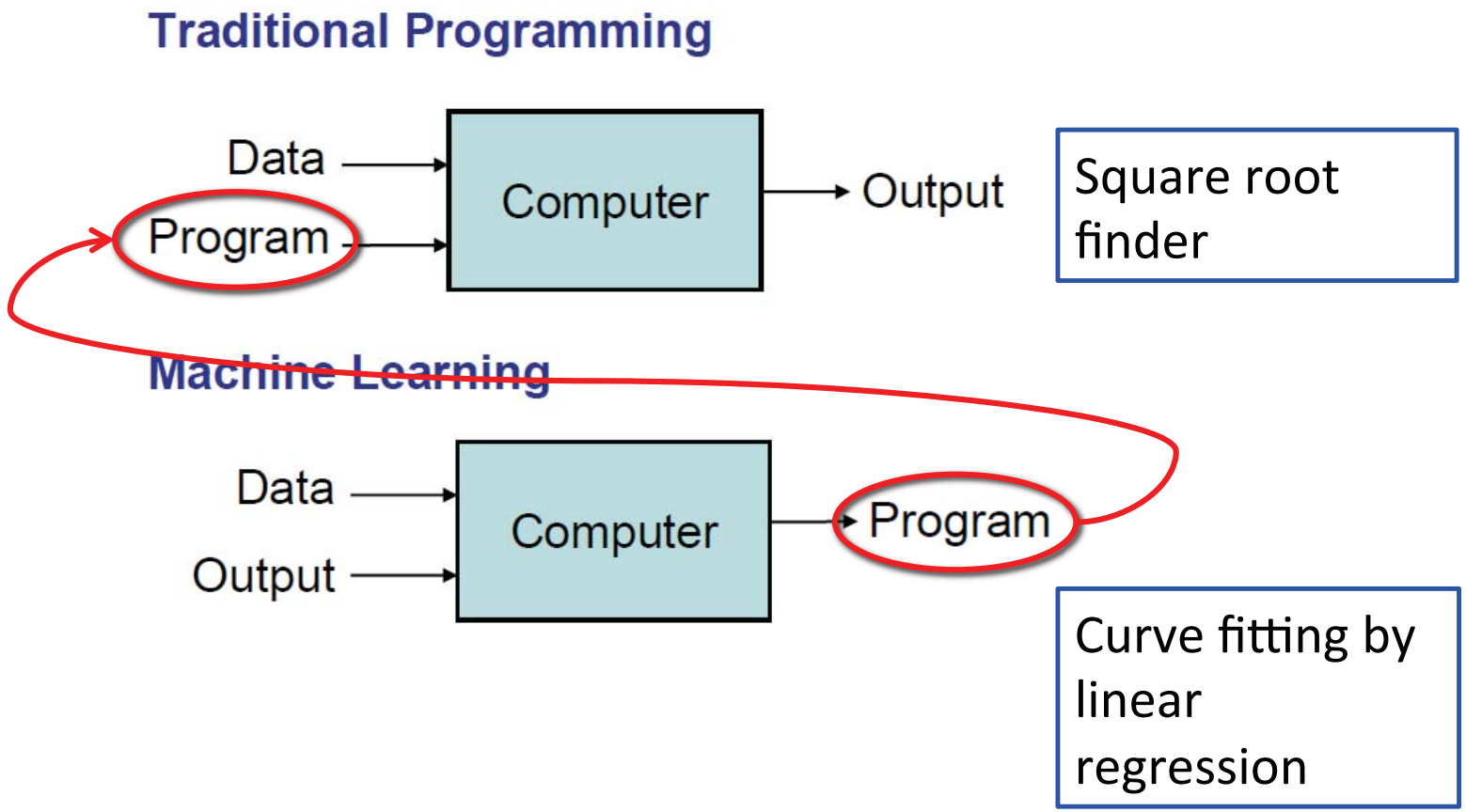
Source:

Eric Grimson, John Guttag, and Ana Bell. 6.0002 Introduction to Computational Thinking and Data Science. Fall 2016. Massachusetts Institute of Technology: MIT OpenCourseWare, https://ocw.mit.edu. License: Creative Commons BY-NC-SA.

What Is Machine Learning?

- All useful programs "learn" something
- In the first lecture of 6.0001 we looked at an algorithm for finding square roots
- Last week we looked at using linear regression to find a model of a collection of points
- Early definition of machine learning:
  - *"Field of study that gives computers the ability to learn without being explicitly programmed."* Arthur Samuel (1959)
    - Computer pioneer who wrote first self-learning program, which played checkers – learned from "experience"
    - Invented alpha-beta pruning – widely used in decision tree searching

What Is Machine Learning?



How Are Things Learned?

- Memorization
  - Accumulation of individual facts
  - Limited by
    - Time to observe facts
    - Memory to store facts

Declarative knowledge

- Generalization
  - Deduce new facts from old facts
  - Limited by accuracy of deduction process
    - Essentially a predictive activity
    - Assumes that the past predicts the future

Imperative knowledge

- Interested in extending to programs that can infer useful information from **implicit** patterns in data

Clustering

Using MIT OpenCourseware Slides

Source:

Eric Grimson, John Guttag, and Ana Bell. 6.0002 Introduction to Computational Thinking and Data Science. Fall 2016. Massachusetts Institute of Technology: MIT OpenCourseWare, https://ocw.mit.edu. License: Creative Commons BY-NC-SA.

Machine Learning Paradigm

- Observe set of examples: training data
- Infer something about process that generated that data
- Use inference to make predictions about previously unseen data: test data
- Supervised: given a set of feature/label pairs, find a rule that predicts the label associated with a previously unseen input
- *Unsupervised*: given a set of feature vectors (without labels) group them into "natural clusters"

Clustering Is an Optimization Problem

$$variability(c) = \sum_{e \in c} distance(mean(c), e)^2$$

$$dissimilarity(C) = \sum_{c \in C} variability(c)$$

- Why not divide variability by size of cluster?
  - Big and bad worse than small and bad
- Is optimization problem finding a C that minimizes *dissimilarity(C)*?
  - No, otherwise could put each example in its own cluster
- Need a constraint, e.g.,
  - Minimum distance between clusters
  - Number of clusters

Two Popular Methods

- Hierarchical clustering
- K-means clustering

Hiearchical Clustering

1. Start by assigning each item to a cluster, so that if you have N items, you now have N clusters, each containing just one item.

2. Find the closest (most similar) pair of clusters and merge them into a single cluster, so that now you have one fewer cluster.

3. Continue the process until all items are clustered into a single cluster of size N.

What does distance mean?

Linkage Metrics

- S*ingle-linkage:* consider the distance between one cluster and another cluster to be equal to the shortest distance from any member of one cluster to any member of the other cluster
- C*omplete-linkage*: consider the distance between one cluster and another cluster to be equal to the greatest distance from any member of one cluster to any member of the other cluster
- A*verage-linkage:* consider the distance between one cluster and another cluster to be equal to the average distance from any member of one cluster to any member of the other cluster

Example of Hierarchical Clustering

	BOS	NY	CHI	DEN	SF	SEA
BOS	0	206	963	1949	3095	2979
NY		0	802	1771	2934	2815
CHI			0	966	2142	2013
DEN				0	1235	1307
SF					0	808
SEA						0

{BOS} {NY} {CHI} {DEN} {SF} {SEA}
{BOS, NY} {CHI} {DEN} {SF} {SEA}
{BOS, NY, CHI} {DEN} {SF} {SEA}
{BOS, NY, CHI} {DEN} {SF, SEA}
{BOS, NY, CHI, DEN} {SF, SEA} Single linkage

or

{BOS, NY, CHI} {DEN, SF, SEA} Complete linkage

Clustering Algorithms

- Hierarchical clustering
  - Can select number of clusters using dendogram
  - Deterministic
  - Flexible with respect to linkage criteria
  - Slow
    - Naïve algorithm n^3
    - n^2 algorithms exist for some linkage criteria
- K-means a much faster greedy algorithm
  - Most useful when you know how many clusters you want

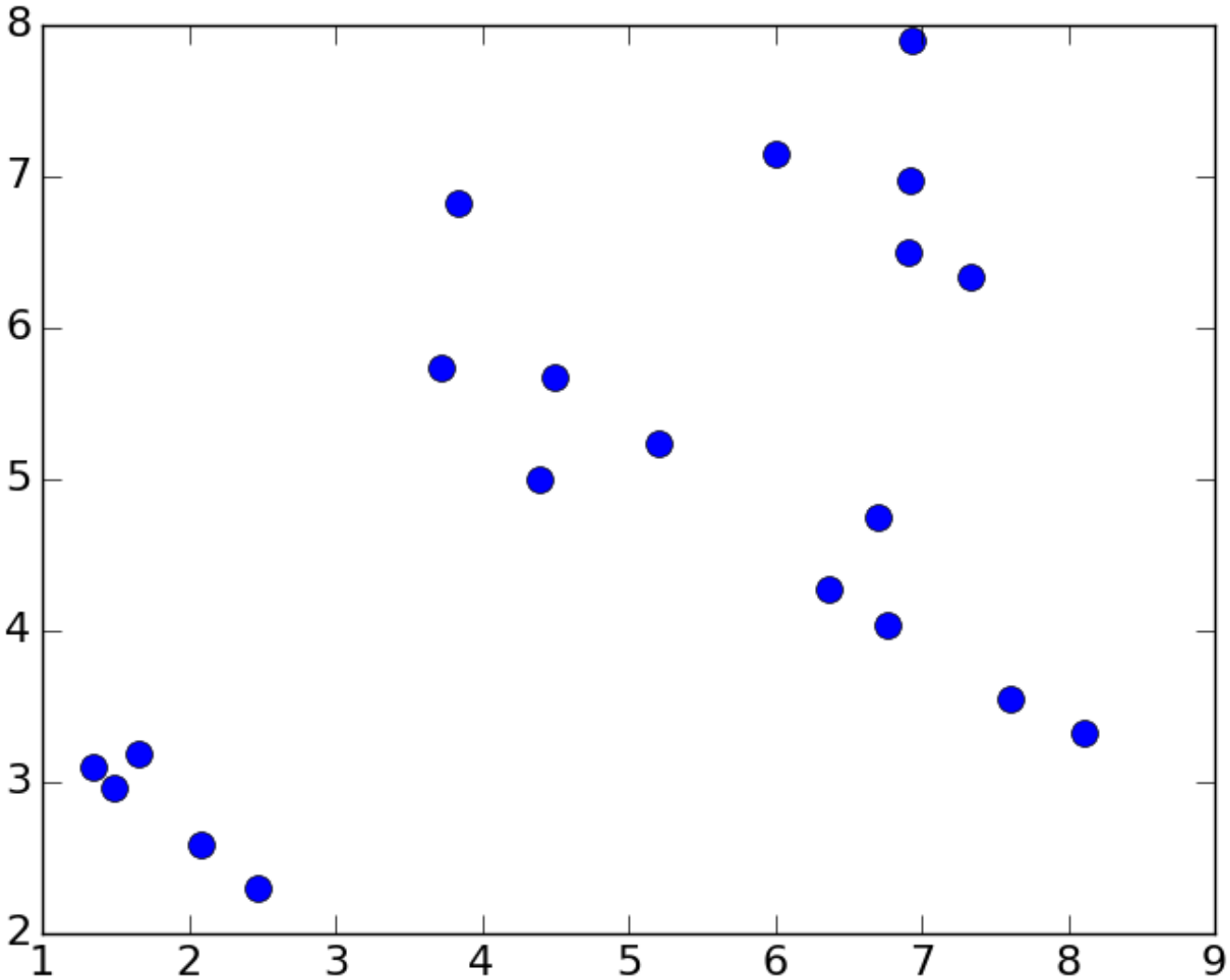
K-means Algorithm

```
randomly chose k examples as initial centroids
while true:
    create k clusters by assigning each
        example to closest centroid
    compute k new centroids by averaging
        examples in each cluster
    if centroids don't change:
        break
```

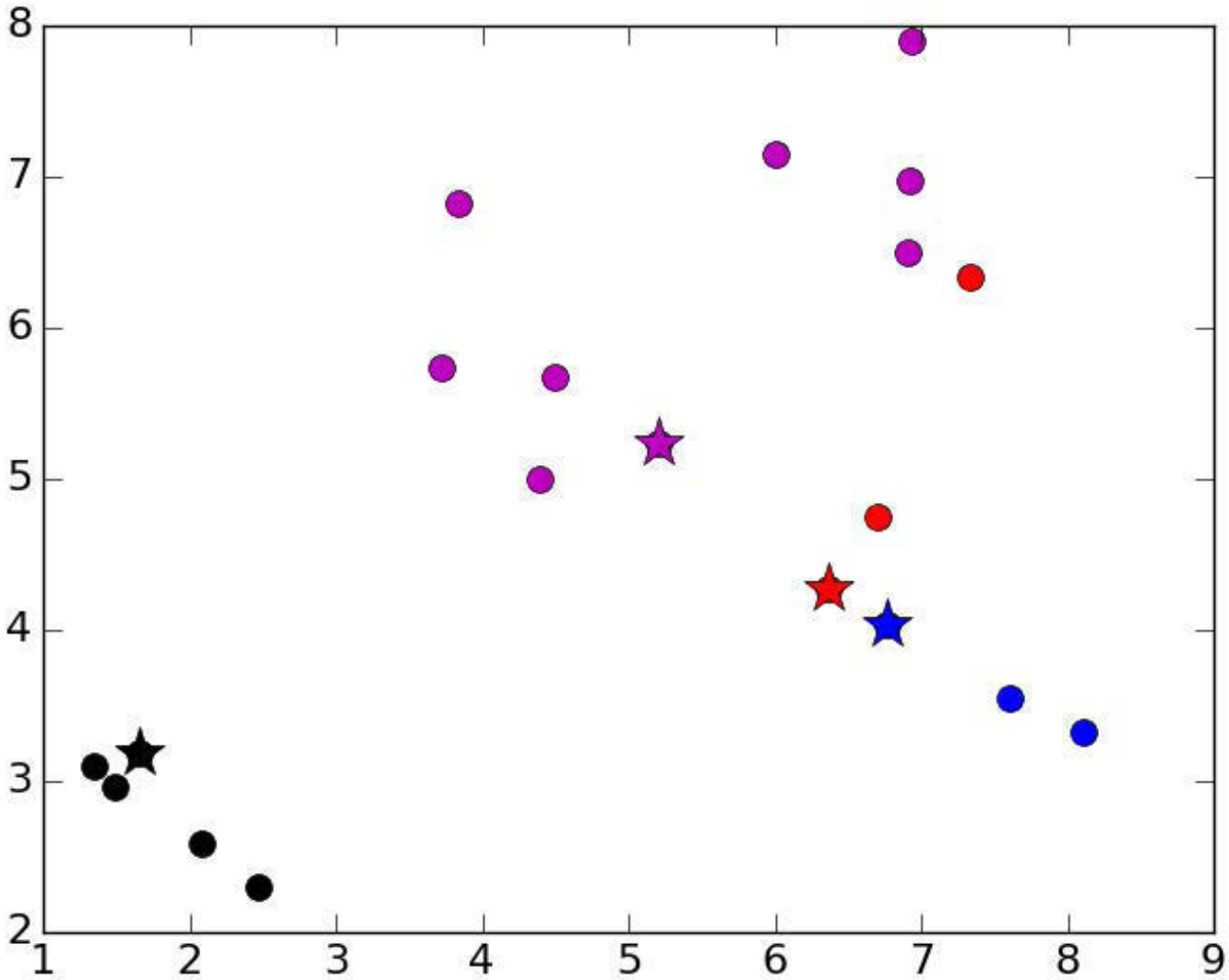
What is complexity of one iteration?

k*n*d, where n is number of points and d time required to compute the distance between a pair of points

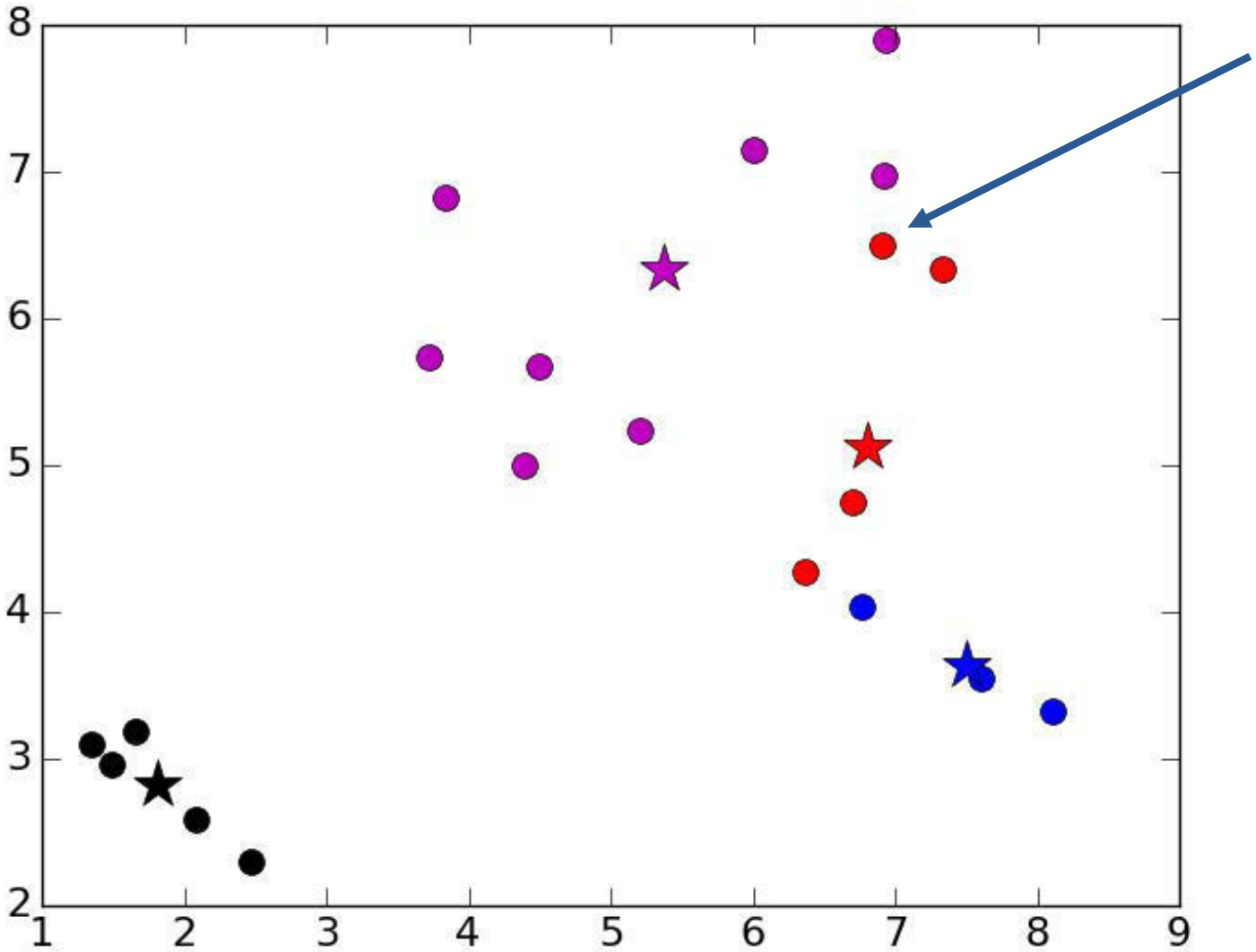
An Example



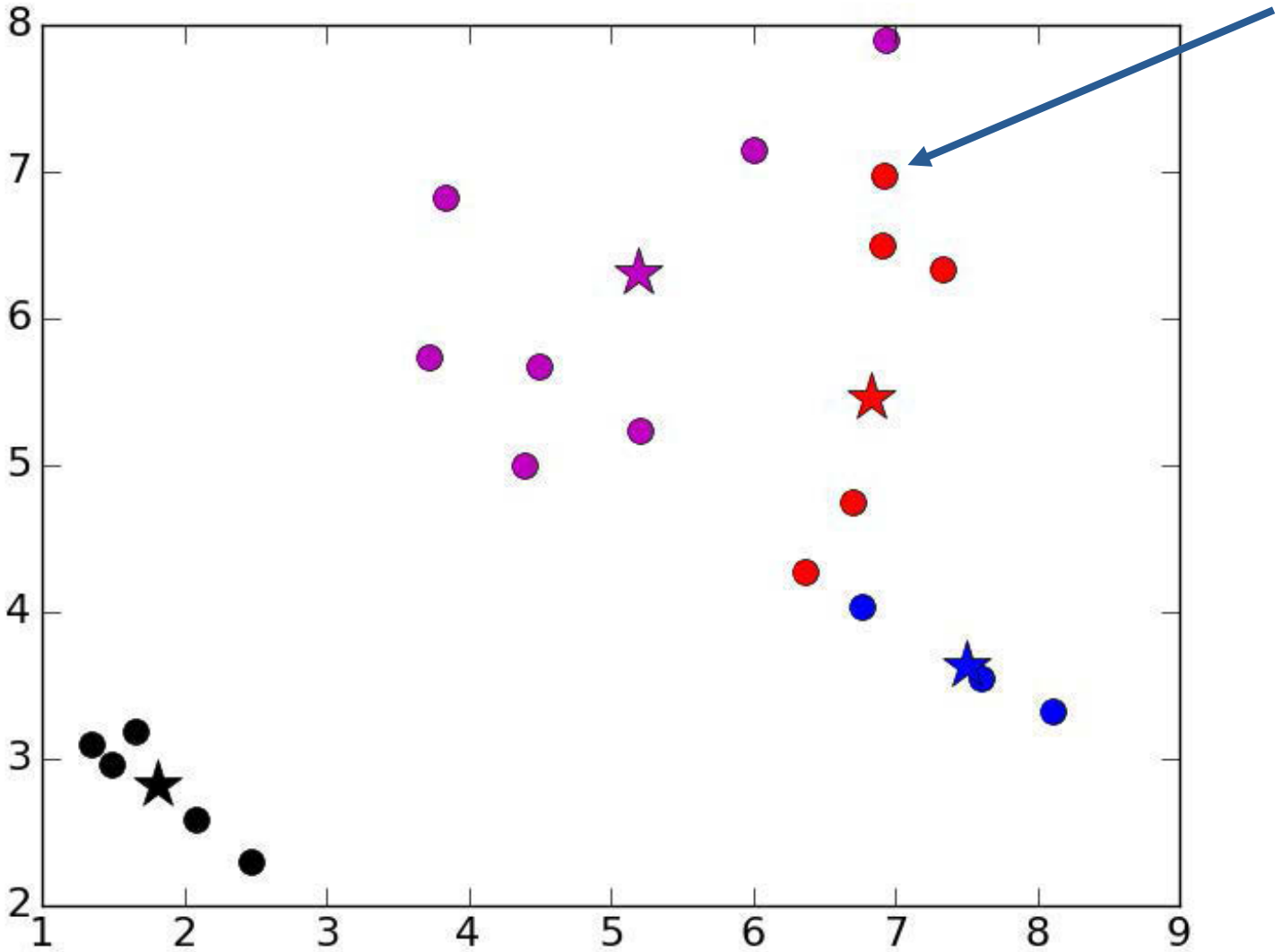
K = 4, Initial Centroids



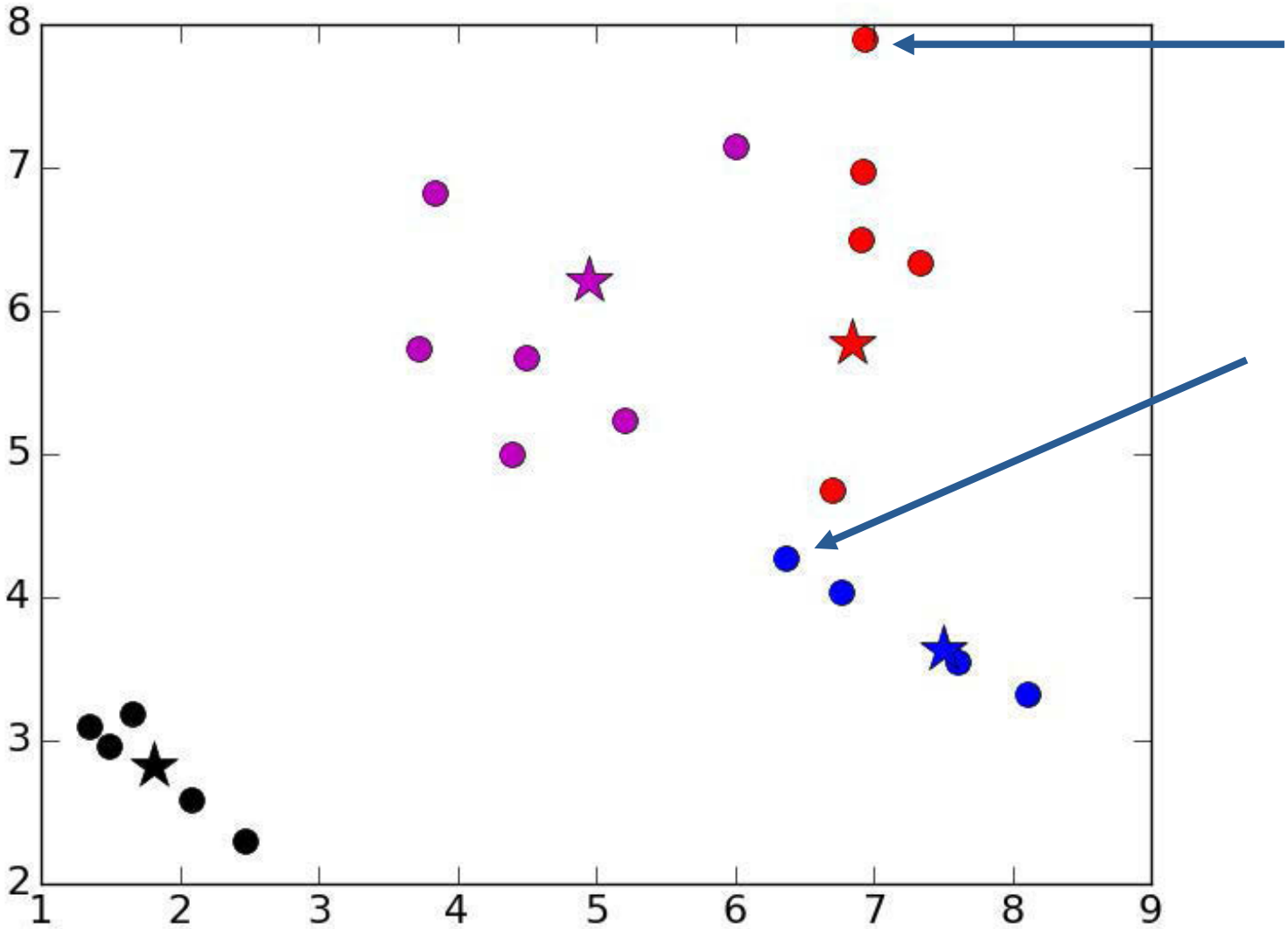
Iteration 1



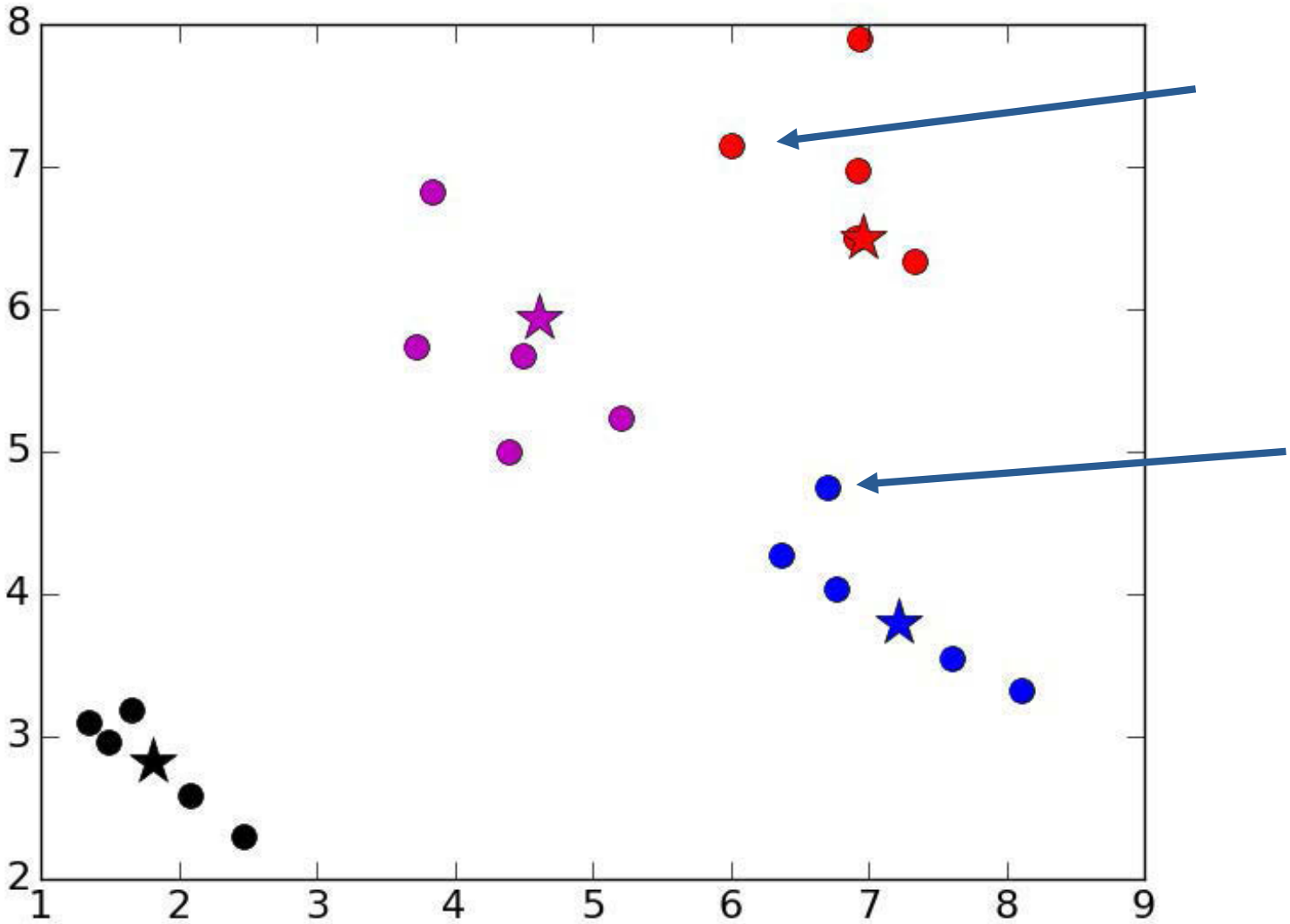
Iteration 2



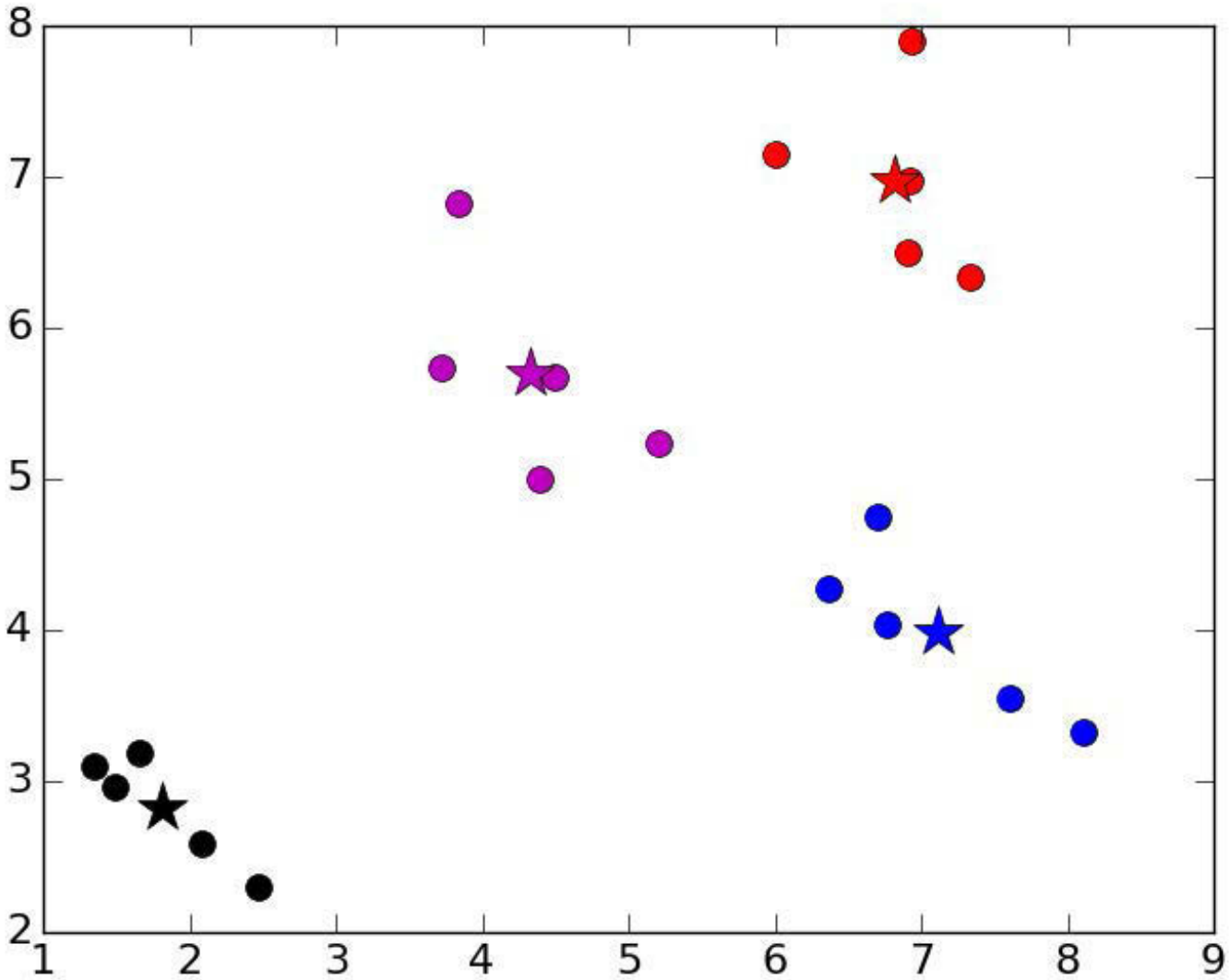
Iteration 3



Iteration 4

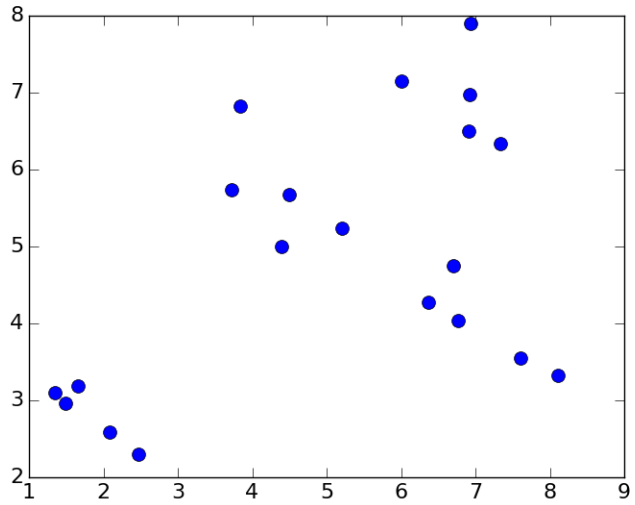


Iteration 5



Issues with k-means

- Choosing the "wrong" k can lead to strange results
  - Consider k = 3

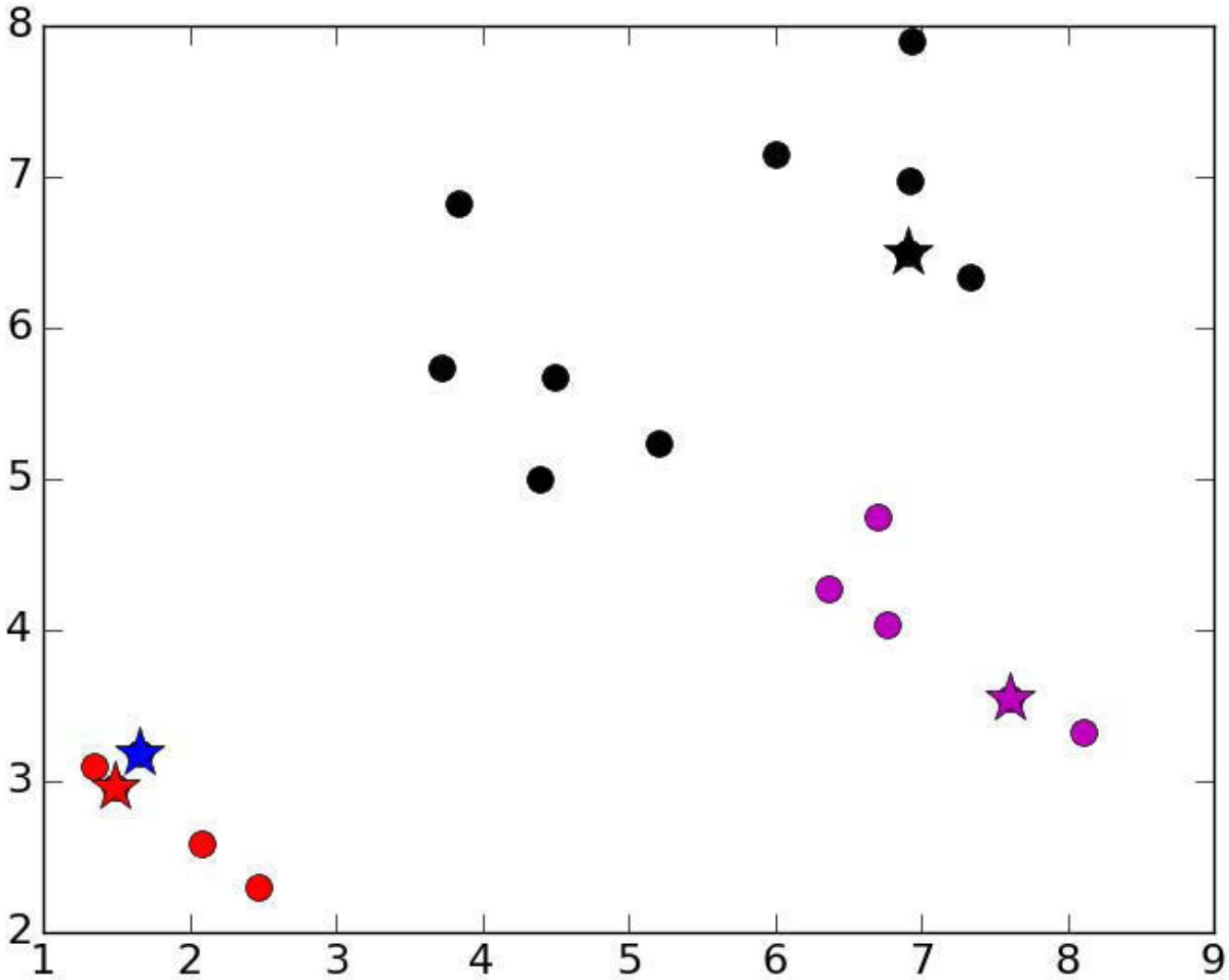


- Result can depend upon initial centroids
  - Number of iterations
  - Even final result
    - Greedy algorithm can find different local optimas

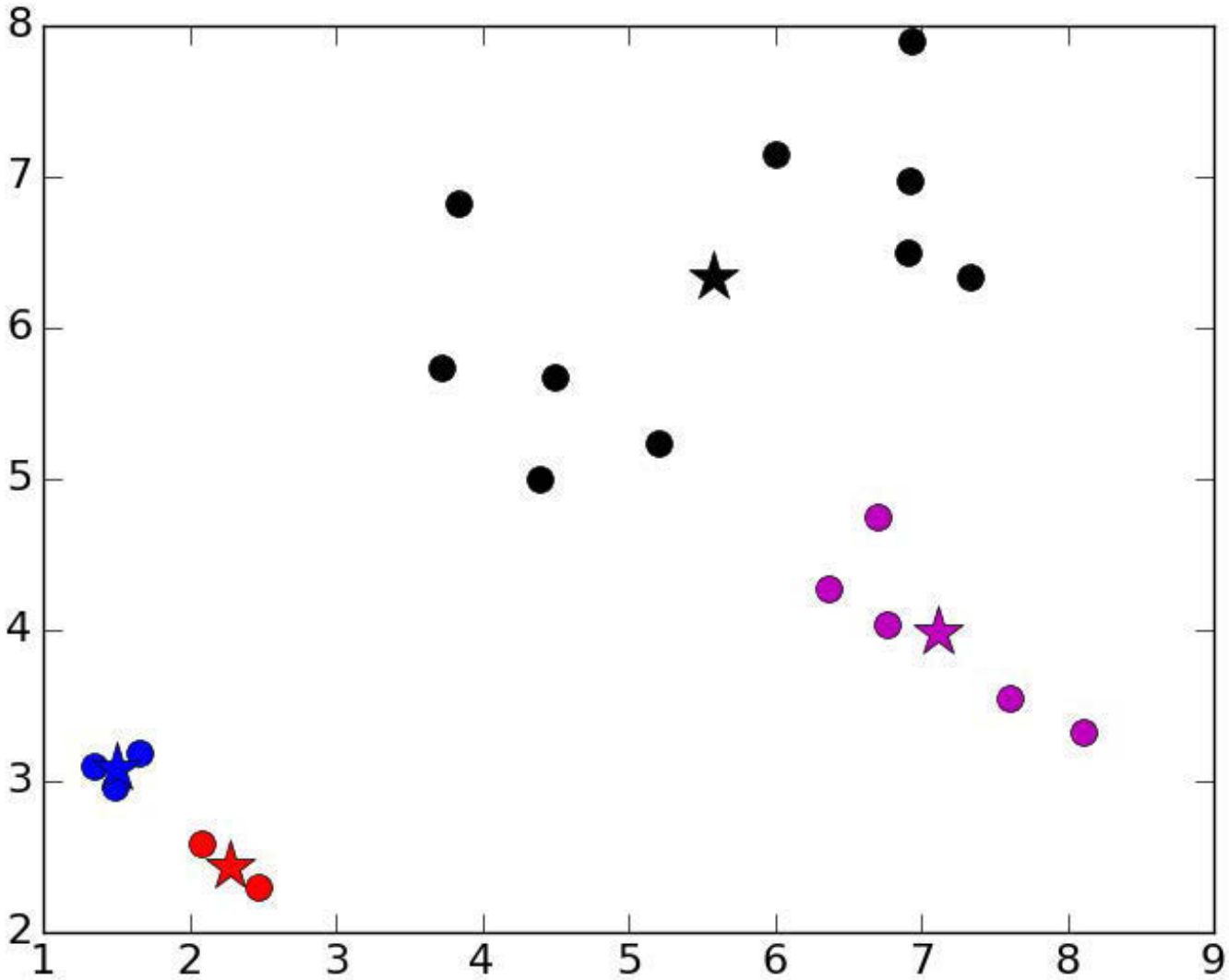
How to Choose K

- *A priori* knowledge about application domain
  - There are two kinds of people in the world: k = 2
  - There are five different types of bacteria: k = 5
- Search for a good k
  - Try different values of k and evaluate quality of results
  - Run hierarchical clustering on subset of data

Unlucky Initial Centroids



Converges On



Monte-Carlo-Simulation (MC-Simulation) recap

- Method of estimating value of unknown quantity using inferential statistics
- Inferential statistics terms:
  - Population: set of examples
  - Sample: Proper subset of population
- Random sample tends to exhibit same properties as population it is drawn from

Example (Flip coins)

Let's estimate the probabilities of heads vs. tails for an infinite number of coin flips:

- One flip (heads): 100% heads?
- Two flips (h, h): still 100 % heads? Confidence level?
- 100 flips (52 h, 48 t): Probability of next coin coming up heads 52/100.

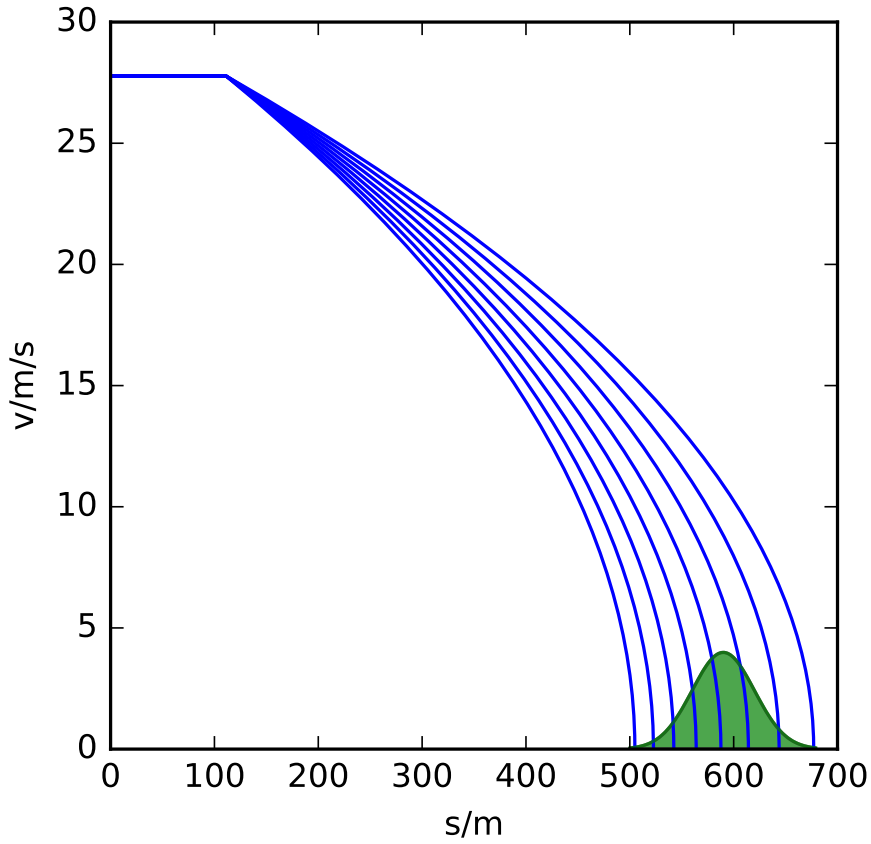
Key aspects to be asked in MC-Simulations

- Never possible to guarantee perfect accuracy through sampling
- Never assume that an estimate is precisely correct
- How many samples do we need to look at before we can have justified confidence in our answers?
  - Answer depends on underlying distribution
  - Especially hard for defects “Rare event simulation”

MC-Simulation application example: Braking curve

CCS systems rely on braking curves to describe the train's braking capability.

- To supervise train velocity, CCS systems predict the future braking capability of the train
- However, there is not *the* braking capability
- Braking curves exhibit a randomised behaviour



Principal Component Analysis (PCA)

Using MIT OpenCourseware Slides

Source:

Philippe Rigollet. 18.650 Statistics for Applications . Fall 2016. Massachusetts Institute of Technology: MIT OpenCourseWare, https://ocw.mit.edu. License: Creative Commons BY-NC-SA.

Multivariate statistics and review of linear algebra (1)

- Let $\mathbf{X}$ be a d-dimensional random vector and $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be n independent copies of $\mathbf{X}$.

- Write $\mathbf{X}_i = (X_i^1, \ldots, X_i^d)^\top, \quad i = 1, \ldots, n.$
- Denote by $\mathbb{X}$ the random $n \times d$ matrix

$$\mathbb{X} = \begin{pmatrix} \cdots & \mathbf{X}_1^\top & \cdots \\ & \vdots & \\ \cdots & \mathbf{X}_n^\top & \cdots \end{pmatrix}.$$

Multivariate statistics and review of linear algebra (2)

Assume that $\mathbb{E}[\|\mathbf{X}\|_2^2] < \infty$.

Mean of $\mathbf{X}$:

$$\mathbb{E}[\mathbf{X}] = \Big(\mathbb{E}[X^1], \ldots, \mathbb{E}[X^d]\Big)^\top.$$

Covariance matrix of $\mathbf{X}$: the matrix $\Sigma = (\sigma_{j,k})_{j,k=1,\ldots,d}$, where

$$\sigma_{j,k} = cov(\mathbf{X}^j, \mathbf{X}^k).$$

It is easy to see that

$$\Sigma = \mathbb{E}[\mathbf{X}\mathbf{X}^\top] - \mathbb{E}[\mathbf{X}]\mathbb{E}[\mathbf{X}]^\top = \mathbb{E}\Big[(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{X} - \mathbb{E}[\mathbf{X}])^\top\Big].$$

Multivariate statistics and review of linear algebra (3)

Empirical mean of $\mathbf{X}_1, \ldots, \mathbf{X}_n$:

$$\bar{\mathbf{X}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_i = \left(\bar{X}^1, \ldots, \bar{X}^d \right)^\top.$$

Empirical covariance of $\mathbf{X}_1, \ldots, \mathbf{X}_n$: the matrix $S = (s_{j,k})_{j,k=1,\ldots,d}$ where $s_{j,k}$ is the empirical covariance of the X_i^j, X_i^k, $i = 1 \ldots, n$.

It is easy to see that

$$S = \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_i \mathbf{X}_i^\top - \bar{\mathbf{X}} \bar{\mathbf{X}}^\top = \frac{1}{n} \sum_{i=1}^{n} \left(\mathbf{X}_i - \bar{\mathbf{X}} \right) \left(\mathbf{X}_i - \bar{\mathbf{X}} \right)^\top.$$

Multivariate statistics and review of linear algebra (4)

Note that $\bar{\mathbf{X}} = \frac{1}{n}\mathbb{X}^\top \mathbb{1}$, where $\mathbb{1} = (1, \ldots, 1)^\top \in \mathbb{R}^d$.

Note also that

$$S = \frac{1}{n}\mathbb{X}^\top \mathbb{X} - \frac{1}{n^2}\mathbb{X}\mathbb{1}\mathbb{1}^\top \mathbb{X} = \frac{1}{n}\mathbb{X}^\top H \mathbb{X},$$

where $H = I_n - \frac{1}{n}\mathbb{1}\mathbb{1}^\top$.

H is an orthogonal projector: $H^2 = H, H^\top = H$. *(on what subspace ?)*

If $\mathbf{u} \in \mathbb{R}^d$,

- $\mathbf{u}^\top \Sigma \mathbf{u}$ is the variance of $\mathbf{u}^\top \mathbf{X}$;
- $\mathbf{u}^\top S \mathbf{u}$ is the sample variance of $\mathbf{u}^\top \mathbf{X}_1, \ldots, \mathbf{u}^\top \mathbf{X}_n$.

Multivariate statistics and review of linear algebra (5)

In particular, $\mathbf{u}^\top S\mathbf{u}$ measures how spread (i.e., diverse) the points are in direction $\mathbf{u}$.

If $\mathbf{u}^\top S\mathbf{u} = 0$, then all $\mathbf{X}_i$' s are in an affine subspace orthogonal to $\mathbf{u}$.

If $\mathbf{u}^\top \Sigma\mathbf{u} = 0$, then $\mathbf{X}$ is almost surely in an affine subspace orthogonal to $\mathbf{u}$.

If $\mathbf{u}^\top S\mathbf{u}$ is large with $\|\mathbf{u}\|_2 = 1$, then the direction of $\mathbf{u}$ explains well the spread (i.e., diversity) of the sample.

Multivariate statistics and review of linear algebra (6)

In particular, Σ and S are symmetric, positive semi-definite.

Any real symmetric matrix $A \in \mathbb{R}^{d \times d}$ has the decomposition

$$A = PDP^\top,$$

where:

P is a $d \times d$ orthogonal matrix, i.e., $PP^\top = P^\top P = I_d$;

D is diagonal.

The diagonal elements of D are the *eigenvalues* of A and the columns of P are the corresponding *eigenvectors* of A.

A is semi-definite positive iff all its eigenvalues are nonnegative.

Principal Component Analysis: Heuristics (1)

The sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$ makes a cloud of points in $\mathbb{R}^d$.

In practice, d is large. If $d > 3$, it becomes impossible to represent the cloud on a picture.

Question: Is it possible to project the cloud onto a linear subspace of dimension $d' < d$ by keeping as much information as possible ?

Answer: PCA does this by keeping as much covariance structure as possible by keeping orthogonal directions that discriminate well the points of the cloud.

Principal Component Analysis: Heuristics (2)

Idea: Write $S = PDP^\top$, where

$P = (\mathbf{v}_1, \ldots, \mathbf{v}_d)$ is an orthogonal matrix, i.e., $\|\mathbf{v}_j\|_2 = 1, \mathbf{v}_j^\top \mathbf{v}_k = 0, \forall j \neq k$.

$$D = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & 0 \\ & & \ddots & \\ & 0 & & \lambda_d \end{pmatrix}, \text{ with } \lambda_1 \geq \ldots \geq \lambda_d \geq 0.$$

Note that D is the empirical covariance matrix of the $P^\top \mathbf{X}_i$'s, $i = 1, \ldots, n$.

In particular, λ_1 is the empirical variance of the $\mathbf{v}_1^\top \mathbf{X}_i$'s; λ_2 is the empirical variance of the $\mathbf{v}_2^\top \mathbf{X}_i$'s, etc...

Principal Component Analysis: Heuristics (3)

So, each λ_j measures the spread of the cloud in the direction $\mathbf{v}_j$.

In particular, $\mathbf{v}_1$ is the direction of maximal spread.

Indeed, $\mathbf{v}_1$ maximizes the empirical covariance of $\mathbf{a}^\top \mathbf{X}_1, \ldots, \mathbf{a}^\top \mathbf{X}_n$ over $\mathbf{a} \in \mathbb{R}^d$ such that $\|\mathbf{a}\|_2 = 1$.

Proof: For any unit vector $\mathbf{a}$, show that

$$\mathbf{a}^\top \Sigma \mathbf{a} = \left(P^\top \mathbf{a}\right)^\top D \left(P^\top \mathbf{a}\right) \leq \lambda_1,$$

with equality if $\mathbf{a} = \mathbf{v}_1$.

Principal Component Analysis: Main principle

Idea of the PCA: Find the collection of orthogonal directions in which the cloud is much spread out.

Theorem

$$\mathbf{v}_1 \in \underset{\|\mathbf{u}\|=1}{\operatorname{argmax}} \ \mathbf{u}^\top S\mathbf{u},$$

$$\mathbf{v}_2 \in \underset{\|\mathbf{u}\|=1, \mathbf{u}\perp\mathbf{v}_1}{\operatorname{argmax}} \ \mathbf{u}^\top S\mathbf{u},$$

$$\dots$$

$$\mathbf{v}_d \in \underset{\|\mathbf{u}\|=1, \mathbf{u}\perp\mathbf{v}_j, j=1,\dots,d-1}{\operatorname{argmax}} \ \mathbf{u}^\top S\mathbf{u}.$$

Hence, the k orthogonal directions in which the cloud is the most spread out correspond exactly to the eigenvectors associated with the k largest values of S.

Principal Component Analysis: Algorithm (1)

1. Input: $\mathbf{X}_1, \ldots, \mathbf{X}_n$: cloud of n points in dimension d.

2. Step 1: Compute the empirical covariance matrix.

3. Step 2: Compute the decomposition $S = PDP^\top$, where $D = Diag(\lambda_1, \ldots, \lambda_d)$, with $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d$ and $P = (\mathbf{v}_1, \ldots, \mathbf{v}_d)$ is an orthogonal matrix.

4. Step 3: Choose $k < d$ and set $P_k = (\mathbf{v}_1, \ldots, \mathbf{v}_k) \in \mathbb{R}^{d \times k}$.

5. Output: $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$, where

$$\mathbf{Y}_i = P_k^\top \mathbf{X}_i \in \mathbb{R}^k, \quad i = 1, \ldots, n.$$

Question: How to choose k ?

Principal Component Analysis: Algorithm (2)

Question: How to choose k ?

Experimental rule: Take k where there is an inflection point in the sequence $\lambda_1, \ldots, \lambda_d$ (scree plot).

Define a criterion: Take k such that

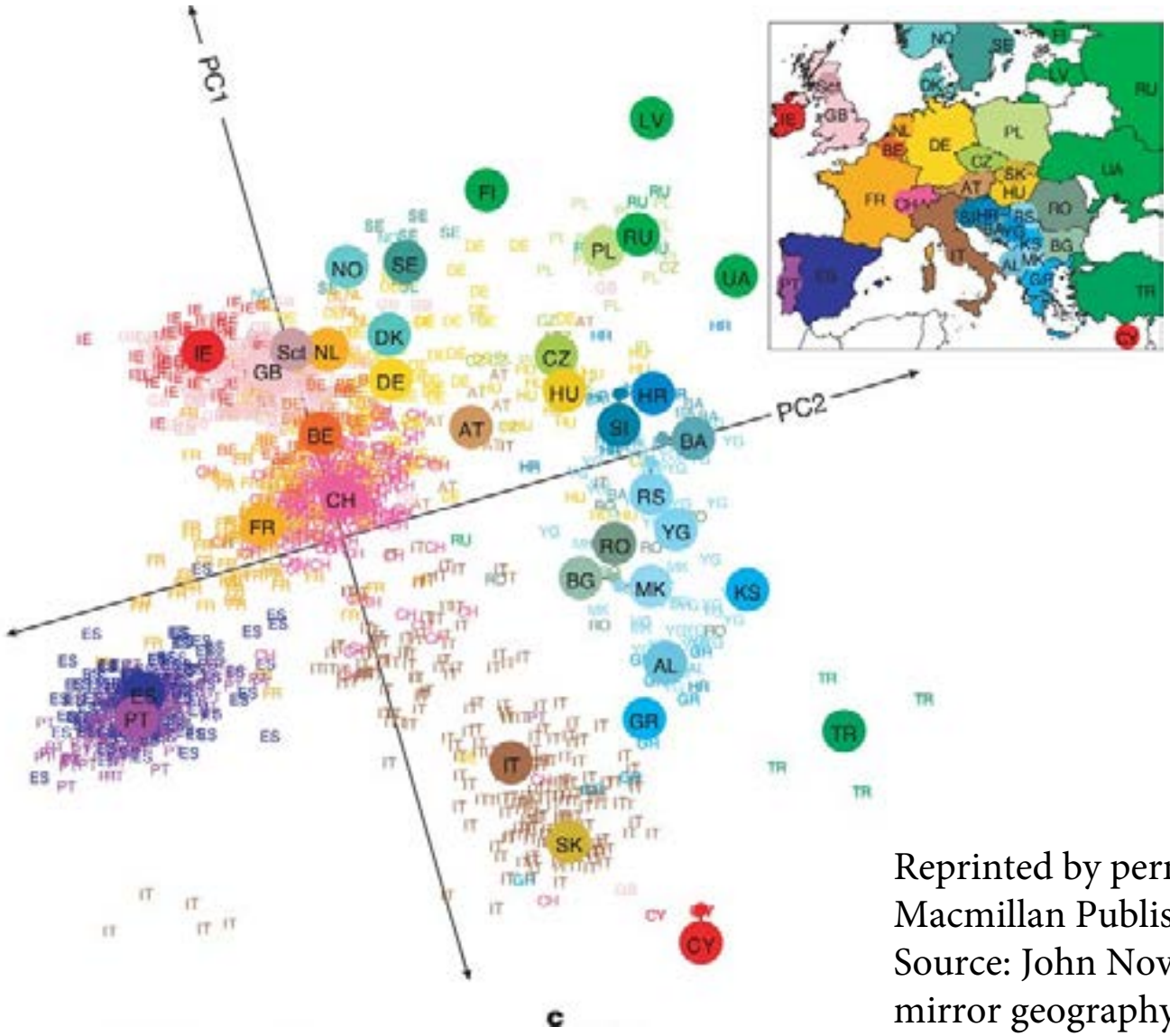
$$\frac{\lambda_1 + \ldots + \lambda_k}{\lambda_1 + \ldots + \lambda_d} \geq 1 - \alpha,$$

for some $\alpha \in (0, 1)$ that determines the approximation error that the practitioner wants to achieve.

Remark: $\lambda_1 + \ldots + \lambda_k$ is called *the variance explained by the PCA* and $\lambda_1 + \ldots + \lambda_d = Tr(S)$ is *the total variance*.

Data visualization: Take $k = 2$ or 3.

Example: Expression of 500,000 genes among 1400 Europeans



Reprinted by permission from Macmillan Publishers Ltd: Nature. Source: John Novembre, et al. "Genes mirror geography within Europe." Nature 456 (2008): 98-101. © 2008.

Principal Component Analysis - Beyond practice (1)

PCA is an algorithm that reduces the dimension of a cloud of points and keeps its covariance structure as much as possible.

In practice this algorithm is used for clouds of points that are not necessarily random.

In statistics, PCA can be used for estimation.

If $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are i.i.d. random vectors in $\mathbb{R}^d$, how to estimate their population covariance matrix Σ ?

If $n \gg d$, then the empirical covariance matrix S is a consistent estimator.

In many applications, $n \ll d$ (e.g., gene expression). Solution: sparse PCA

Principal Component Analysis - Beyond practice (2)

It may be known beforehand that Σ has (almost) low rank.

Then, run PCA on S: Write $S \approx S'$, where

$$S' = P \begin{pmatrix} \lambda_1 & & & & & & \\ & \lambda_2 & & & & 0 & \\ & & \ddots & & & & \\ & & & \lambda_k & & & \\ & & & & 0 & & \\ & 0 & & & & \ddots & \\ & & & & & & 0 \end{pmatrix} P^\top.$$

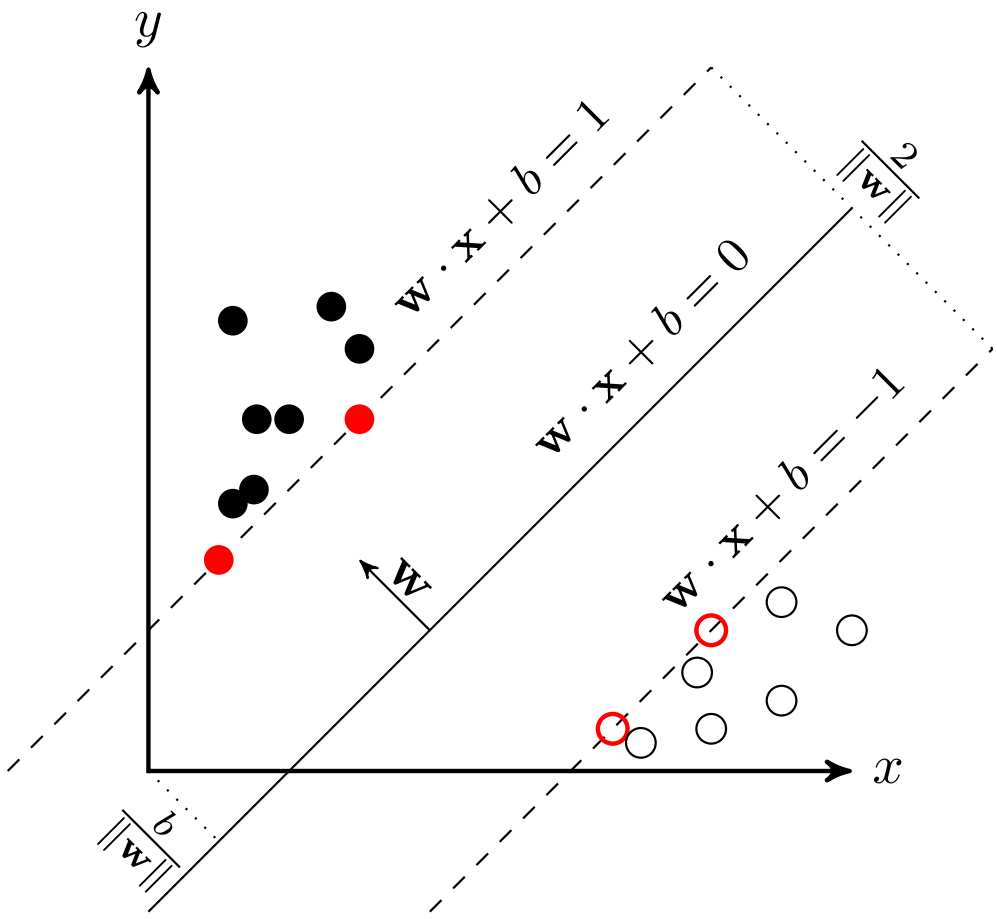
S' will be a better estimator of S under the low-rank assumption.

A theoretical analysis would lead to an optimal choice of the tuning parameter k.

Support Vector Machine

Separate data into subsets according to their nD-coordinates.

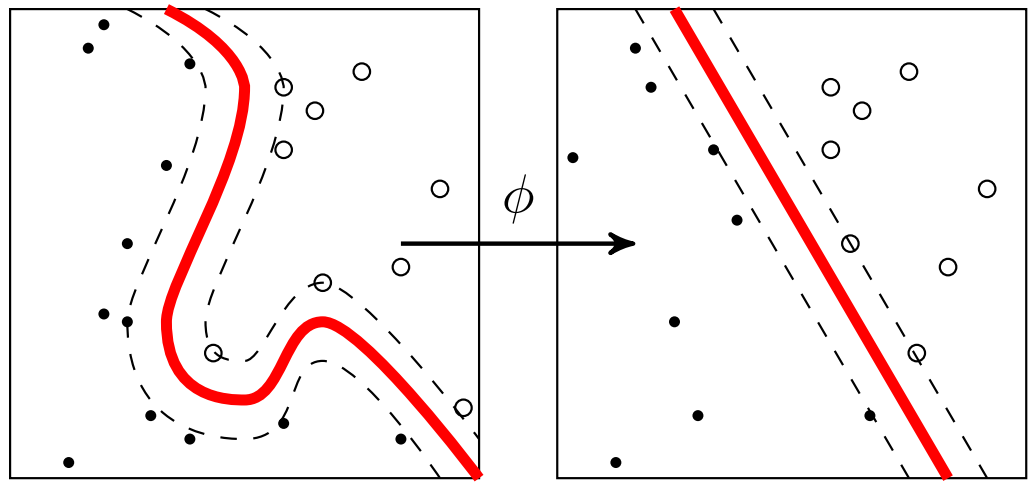
- Idea:
  - Find separating hyperplane maximising the distance between borderline instances
  - For data mixed in feature space, slack parameter (and penalty C) is used
  - If no separating hyperplane can be found in feature space, dimension is increased "Kernel trick"
- Robust:
  - High dimensionality
  - Small datasets
- Simple to complex models



Support Vector Machine

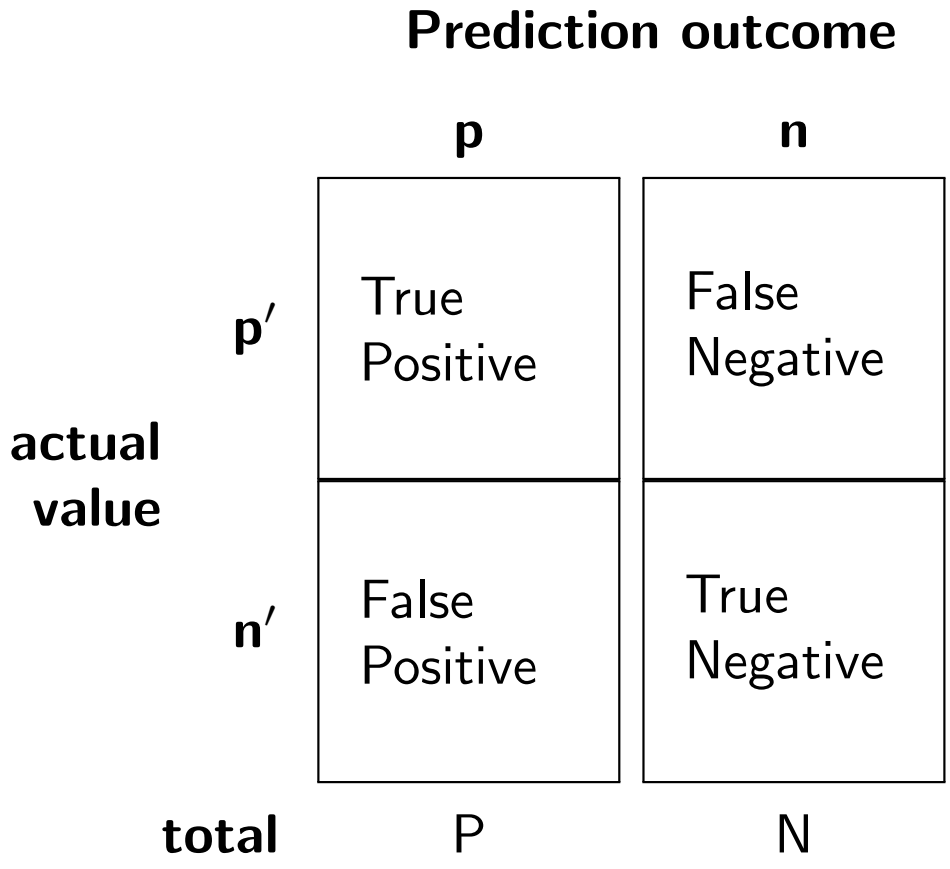
Separate data into subsets according to their nD-coordinates.

- Idea:
  - Find separating hyperplane maximising the distance between borderline instances
  - For data mixed in feature space, slack parameter (and penalty C) is used
  - If no separating hyperplane can be found in feature space, dimension is increased "Kernel trick"
- Robust:
  - High dimensionality
  - Small datasets
- Simple to complex models



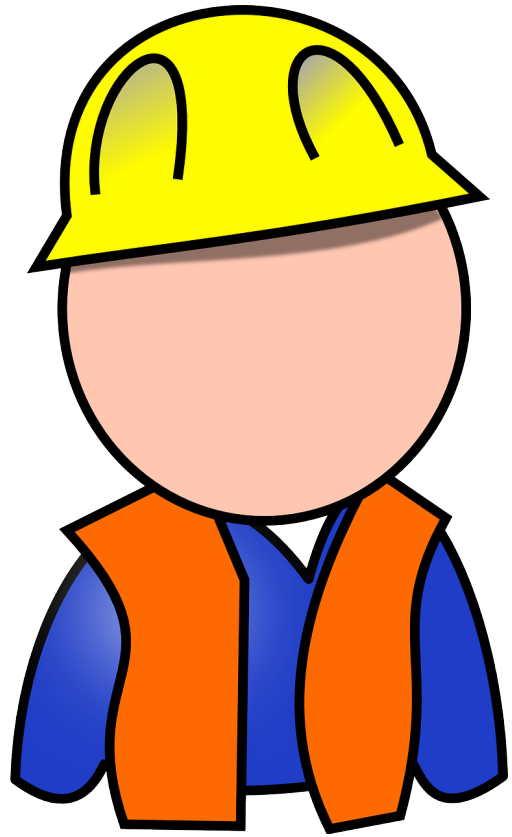
Practical hints for application of data science methods

- Get to know your data
- Use training and validation sets
- For clustering and ML:
  - Regularise
  - Remove outliers, NaN
- When building models:
  - Don't expect perfect fit
  - Inspect confusion matrix



Get to work!

Load and try `2-Container-SVM.ipynb` (http://bit.ly/ifv122019)



Abschnitt 3

Recap Fertigungsmesstechnik

Definition Fertigungsmesstechnik

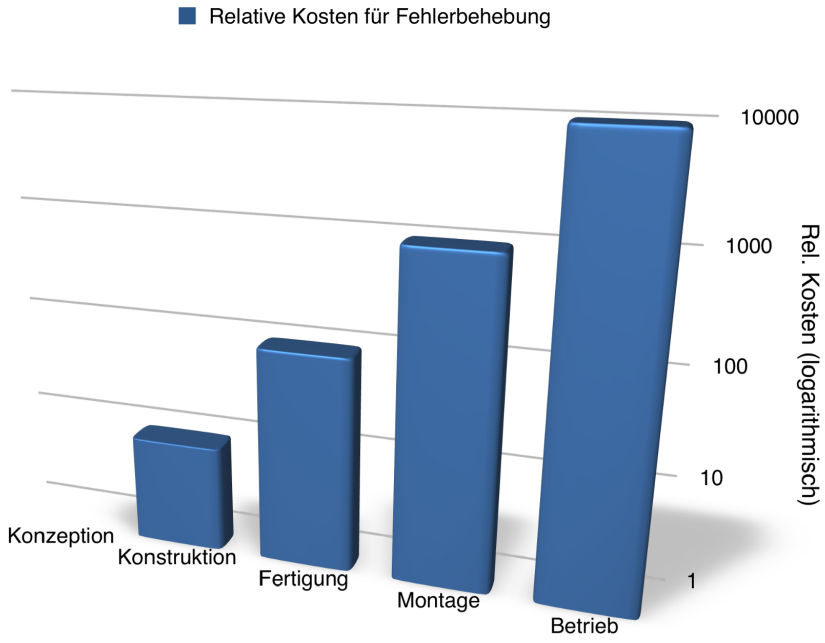
*[...] Oberbegriff für alle mit Mess- und Prüfaufgaben verbundenen Tätigkeiten, die beim industriellen Entstehungsprozess eines Produktes zu erbringen sind. [**?**, S. 1]*

- Im Entstehungsprozess
- Messen und Prüfen
- Aspekte der Definition:
  - Geringe Fertigungstiefe
  - Automatisierung
  - gestiegene Qualitätsforderungen
- Von Kontrollinstanz zu Komponente des QM

Frühe Entwicklung und Rolle der Fertigungsmesstechnik

- Messtechnik seit ca. 4000 v. Chr.: Vergleich mit natürlichen Maßen
- Festlegung Urmeter 1799
- Beschleunigt durch Austauschbau und Massenfertigung
- Elektronische Messtechnik ca. seit 1970

- Rolle der Messtechnik:
  - 20er Jahre: Sortierung
  - 30er Jahre: Prozessüberwachung und Regelung
  - 80er Jahre: Planung zur Fehlervermeidung
  - 90er Jahre: Gesamtheitliches Qualitätsdenken



Aufgaben und Ziele der Fertigungsmesstechnik

- Erfassung von Qualitätsmerkmalen an Messobjekt
  - Werkstoffeigenschaften (z.B. Gefüge, Härte)
    - Als Eingangsprüfung
    - Nach thermischer Behandlung
    - auch für Schweißnahtgüte
  - Geometrie (z.B. Maß, Form, Lage)
    - Dominierende Prüfung
    - Gestalt des Werkstücks
    - Oberflächeneigenschaft
  - Funktion (z.B. Kraft, Geschwindigkeit)
    - Produkt oder Baugruppe
    - Von Sichtprüfung bis zur vollautomatisierten Funktionsprüfung

Größenordnungen und Definitionen

- Messgröße Länge: üblicherweise $\left(10^{-9}...10^{2}\right)\,\mathrm{m}$
- Toleranzen zunehmend enger

Definition (Messgröße)

Die Messgröße ist die physikalische Größe, der die Messung gilt.

Definition (Messen)

Messen ist das Ausführen von geplanten Tätigkeiten zum Vergleich der Messgröße mit einer Einheit.

Definition (Prüfen)

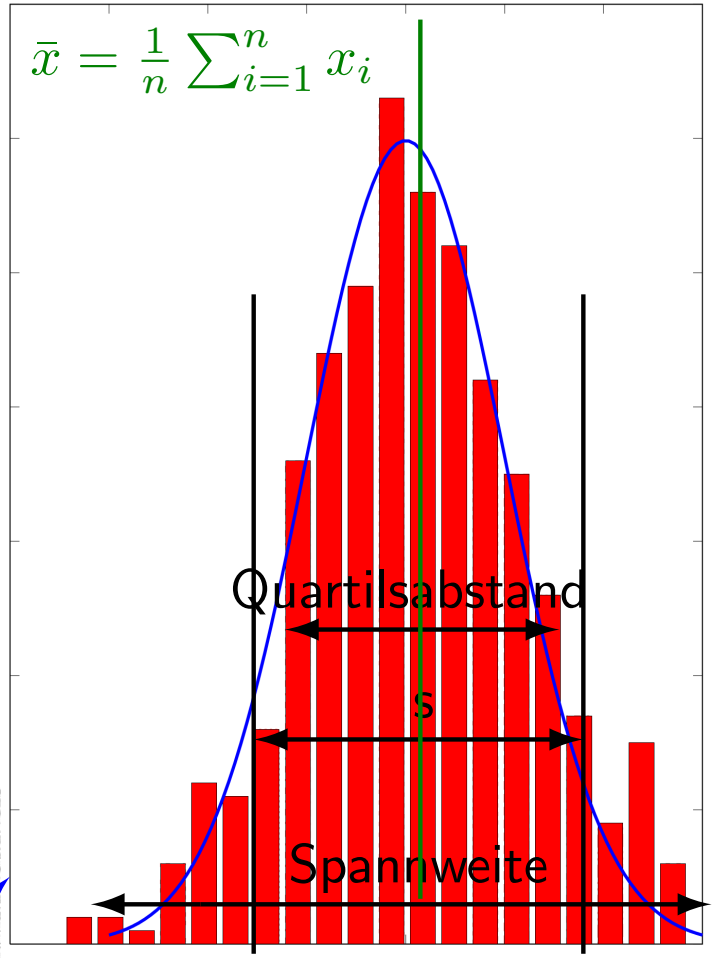
Prüfen heißt feststellen, inwieweit ein Prüfobjekt eine Forderung erfüllt.

Abschnitt 4

Prüfdatenauswertung

Beschreibung

Beschreibung der Verteilung durch Lage und Streuungsparameter verdichtet die information.

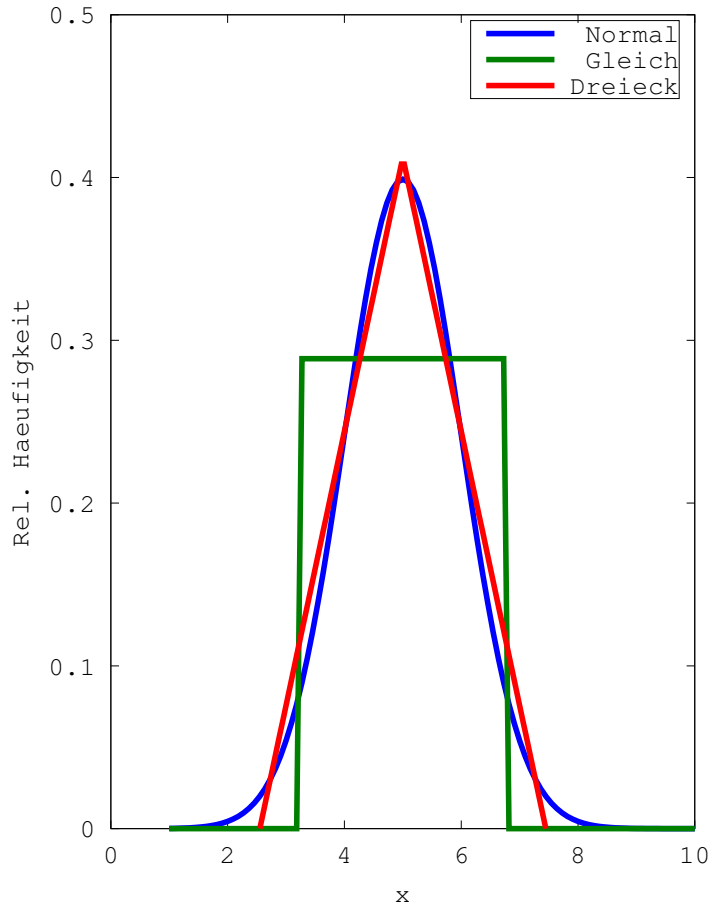


- Lageparameter:
  - Arithmetisches Mittel $\bar{x}$
  - Modalwert: am häufigsten angenommene Klasse
  - Median: je 50% der Messwerte größer bzw. kleiner
- Streuungsparameter
  - Spannweite
  - Quartilsabstand: Spannweite der zentralen 50% der Messwerte
  - Standardabweichung s
  - Quantile: Merkmalsausprägung, für die ein Anteil α kleiner ist

Verteilungen

Zufallsmodelle können helfen, beobachtete Phänomene zu beschreiben und über die Modellbildung vorhersagen zu treffen.

- Näherungsweise Beschreibung
- I.d.R. Konvergenz für große Stichproben
- Wahrscheinlichkeit kann als relative Häufigkeit interpretiert werden
- Verteilung stetiger und diskreter Merkmale unterschiedlich zu modellieren



Verteilungen diskreter Merkmale

- Poisson-Verteilung:

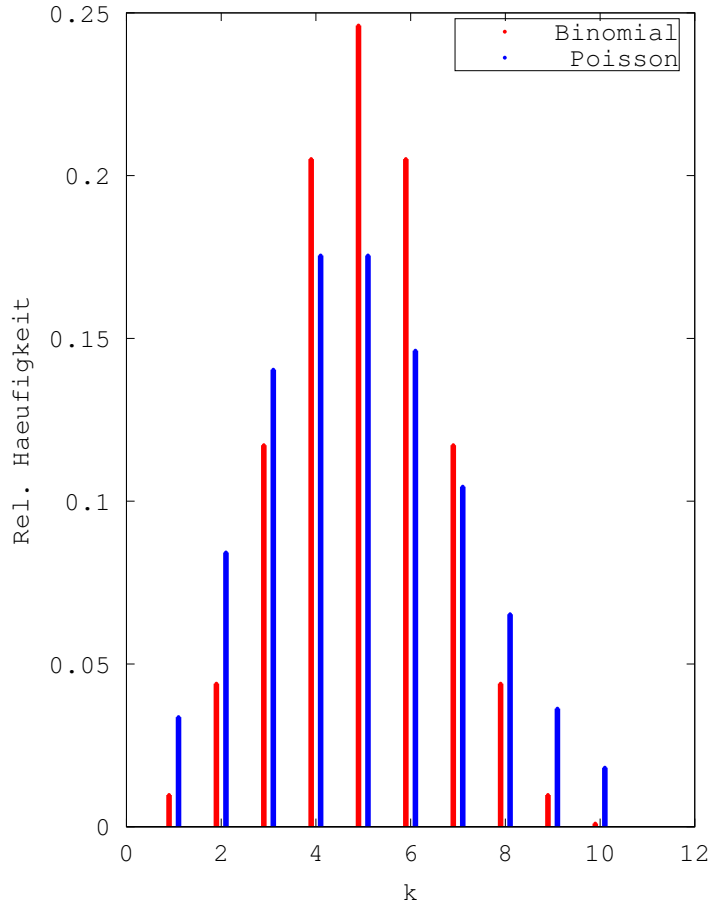
$$p_P(k) = \frac{\lambda^k}{k!} \exp(-\lambda)$$

$\rightarrow$ Geringe Wahrscheinlichkeiten, z.B. $p \leq 0,1$

- Binomial-Verteilung:

$$p_B(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$\rightarrow$ Höhere Wahrscheinlichkeiten, z.B. $p > 0,1$



Verteilungen kontinuierlicher Merkmale

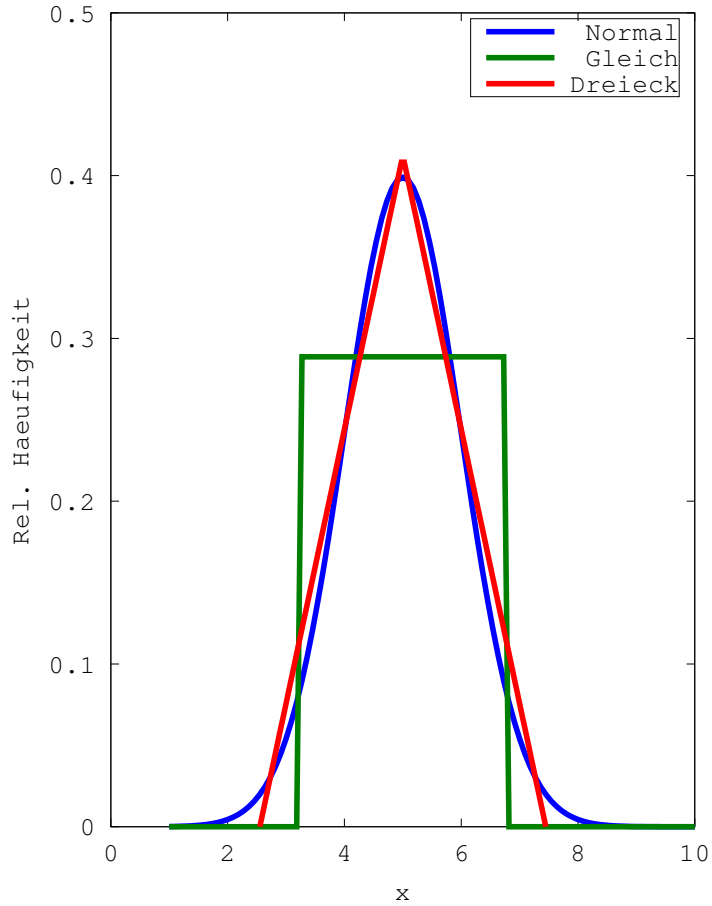
- Normalverteilung:

$$p_N(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$\rightarrow$ Faltung vieler Verteilungen

- Gleichverteilung
- Dreieckverteilung

$\rightarrow$ Faltung zweier Rechteckverteilungen



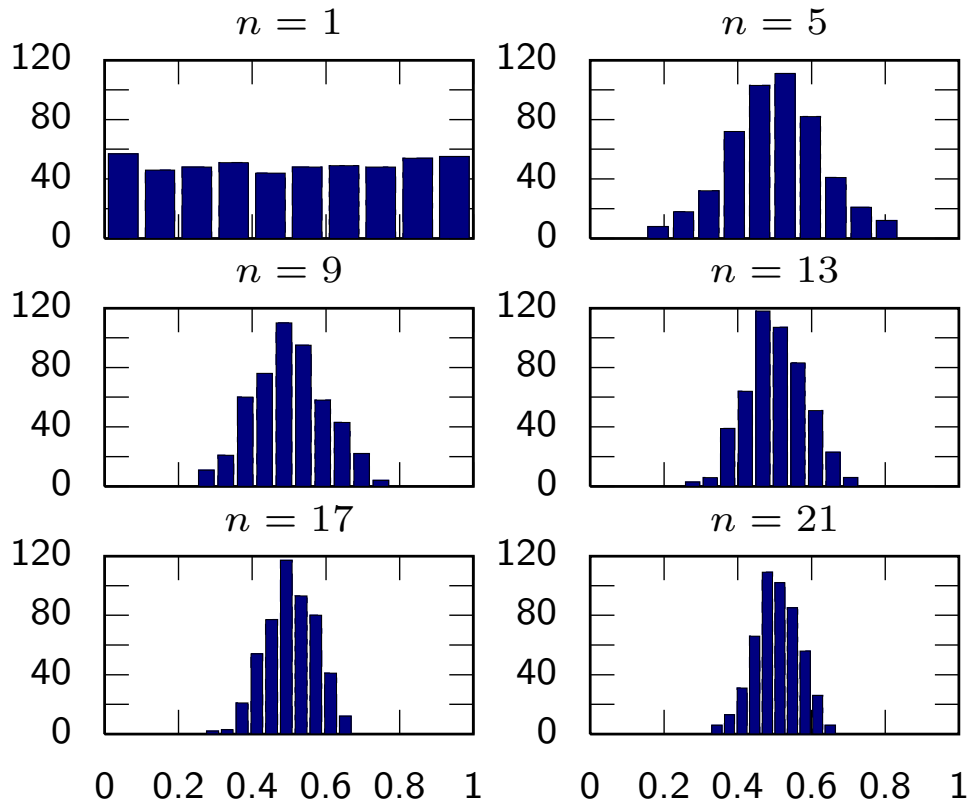
Zentraler Grenzwertsatz

Konvergenz der Summe von Zufallsvariablen gegen die Standardnormalverteilung.

- Sei $X_1, X_2, X_3, \ldots$ eine Folge von Zufallsvariable, die auf demselben Wahrscheinlichkeitsraum unabhängig und identisch verteilt sind.
- Sei weiterhin $S_n = X_1 + X_2 + \ldots + X_n$ die n-te Teilsumme, eine Zufallsvariable mit $E(S_n) = n\mu$ und $\mathrm{Var}(S_n) = n\sigma^2$
- Dann konvergiert die Verteilungsfunktion der standardisierten Zufallsvariablen

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

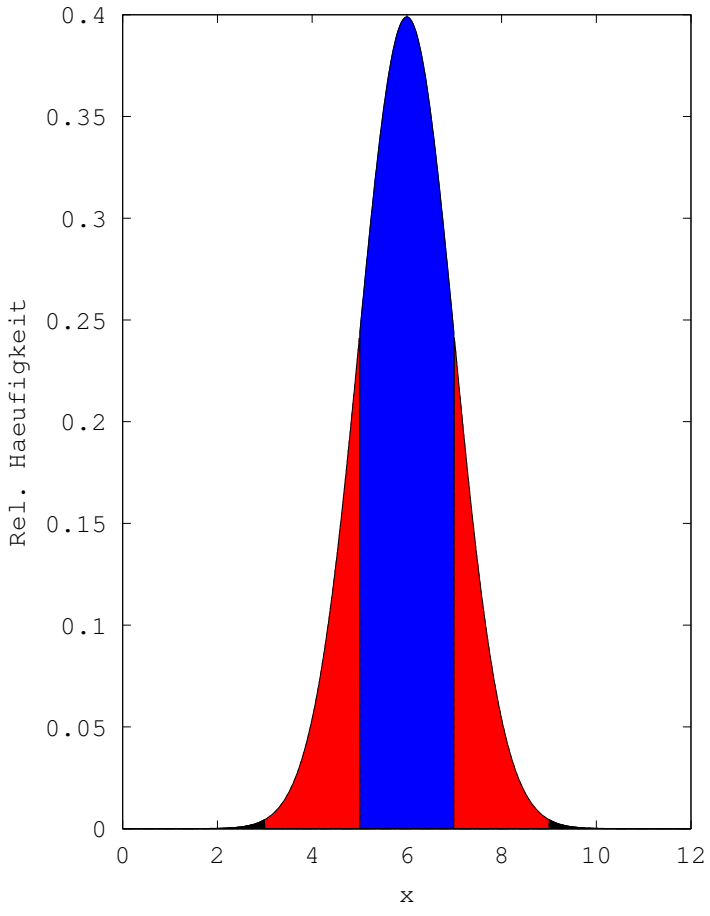
für $n \to \infty$ punktweise gegen die Standardnormalverteilung $\mathcal{N}(0, 1)$.



Summe aus n gleichverteilten Zufallsvariablen

Eigenschaften der Normalverteilung

Viele Phänomene in Technik und Naturwissenschaft lassen sich durch eine Normalverteilung annähern.



- Mittelwert μ
- Standardabweichung σ
- σ-Umgebungen:

k	% der Realisierungen
1	68,3
2	95,5
3	99,4

- Quantile:

% der Realisierungen	k
50	0,675
90	1,65
95	1,96
99	2,58

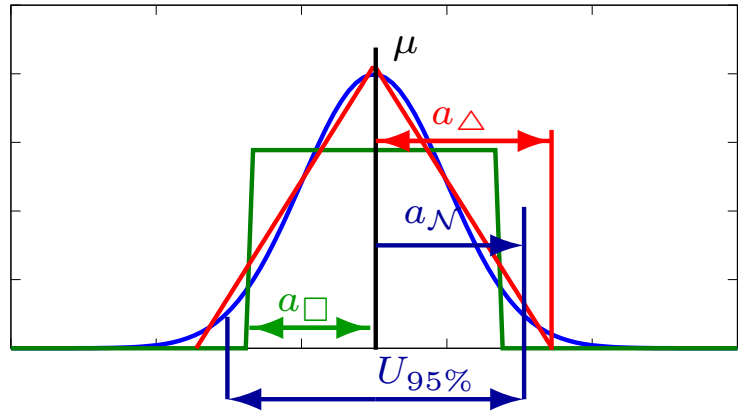
Vertiefung: Konfidenzintervall, Bestimmung u

Konfidenzintervalle sind Intervalle in denen der wahre Wert einer Merkmalsausprägung mit Wahrscheinlichkeit p liegt.

- Größe und Lage abhäbhängig von
  - Verteilung
  - Konfidenzniveau (z.B. 95%)
- Für Normalverteilung:

k	% der Realisierungen
1	68,3
2	95,5
3	99,4

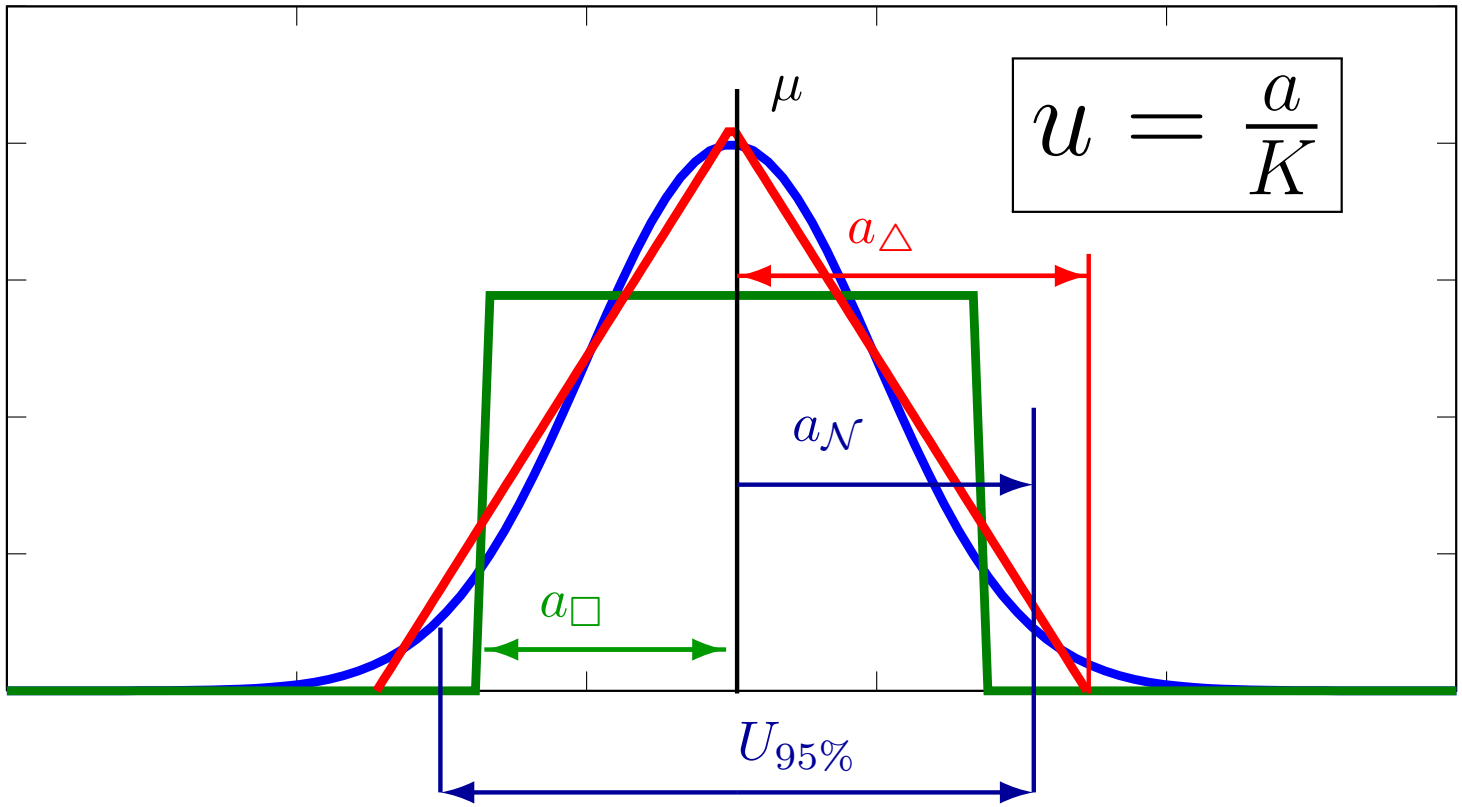
- Fertigungsmesstechnik: 95%-Konfidenzintervall $U_{95\%}$ für $k = 2$



$$u = \frac{a}{K}$$

Normalverteilung: $k = \sqrt{4}$, a Abmaße von $U_{95\%}$
Gleichverteilung: $k = \sqrt{3}$
Dreiecksverteilung: $k = \sqrt{6}$

Vertiefung: Konfidenzintervall, Bestimmung u



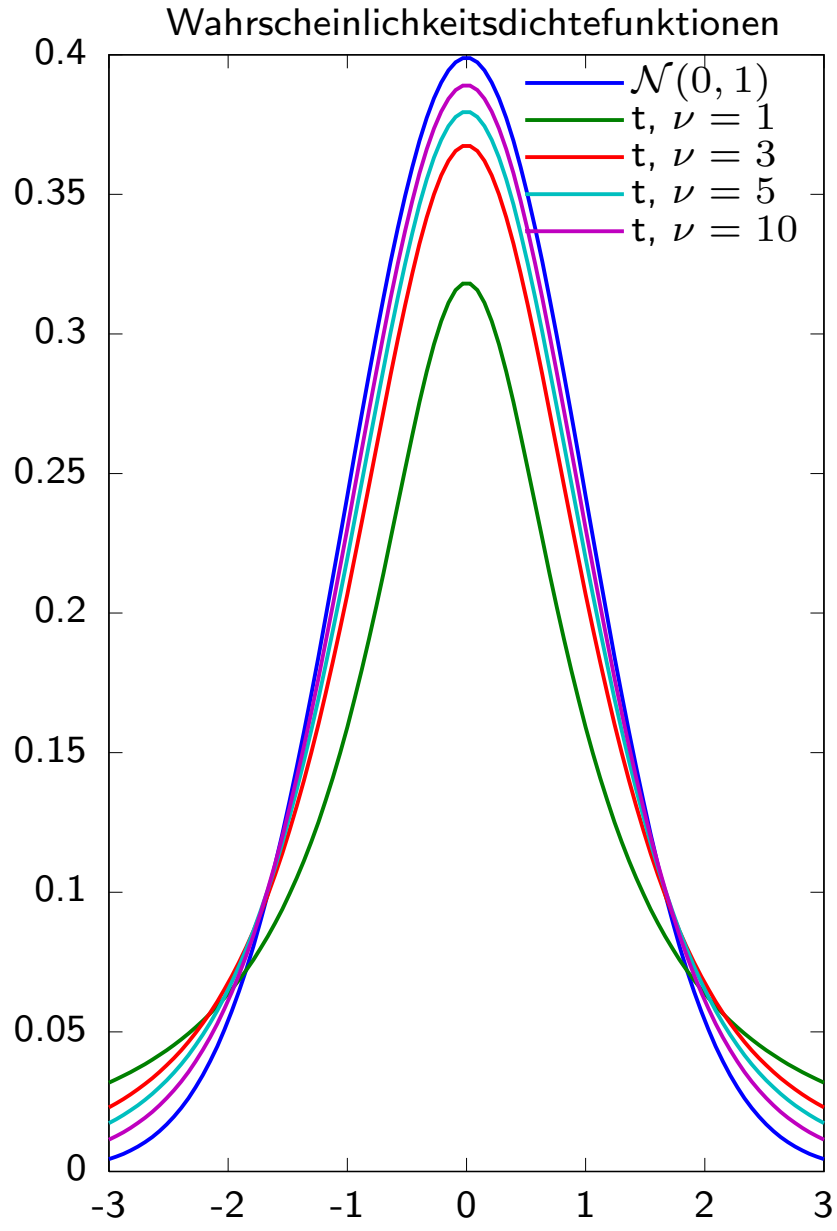
- Normalverteilung: $k = \sqrt{4}$, a Abmaße von $U_{95\%}$
- Gleichverteilung: $k = \sqrt{3}$
- Dreiecksverteilung: $k = \sqrt{6}$

Student'sche t-Verteilung

- Normalverteilung (Hypothese):
  - Endliche Stichprobe
  - Wahrscheinlichkeit, "seltene" Werte zu realisieren?
- Stichprobenvarianz fällt zu klein aus
- "schlanke" Normalverteilung nicht konservativ
- Abhilfe: Student'sche t-Verteilung:

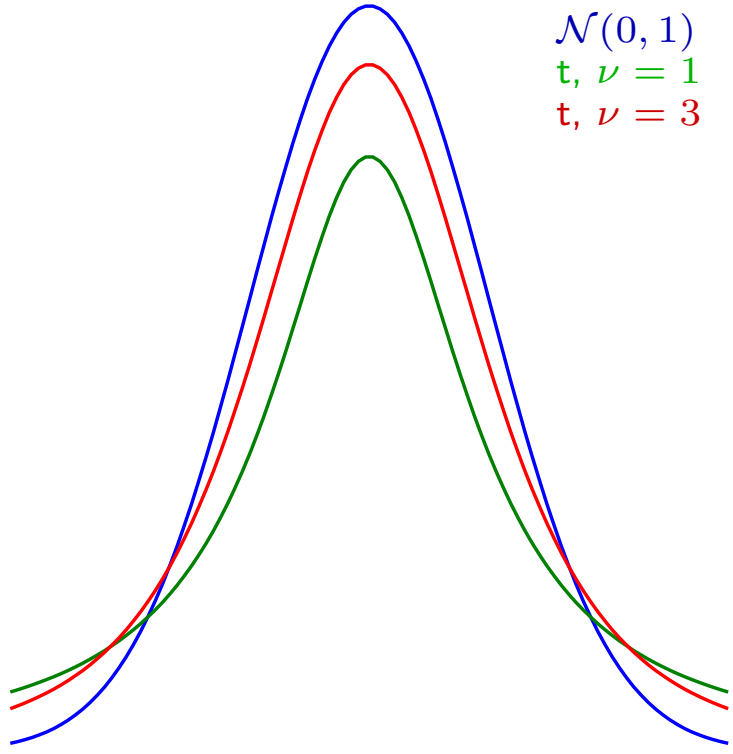
$$p_n(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{2}\right)^{\frac{-n+1}{2}}$$

  - für n Freiheitsgrade



Praktische Anwendung der Student'schen t-Verteilung

- t-Verteilung unterstellt "breitere" Wahrscheinlichkeitsdichte
- Schätzwert der Stichprobenvarianz wird korrigiert
- Bestimmung Korrekturfaktor aus Konfidenz und Stichprobengröße n
- Freiheitsgrade $\nu = n - m$ für m zu schätzende Parameter

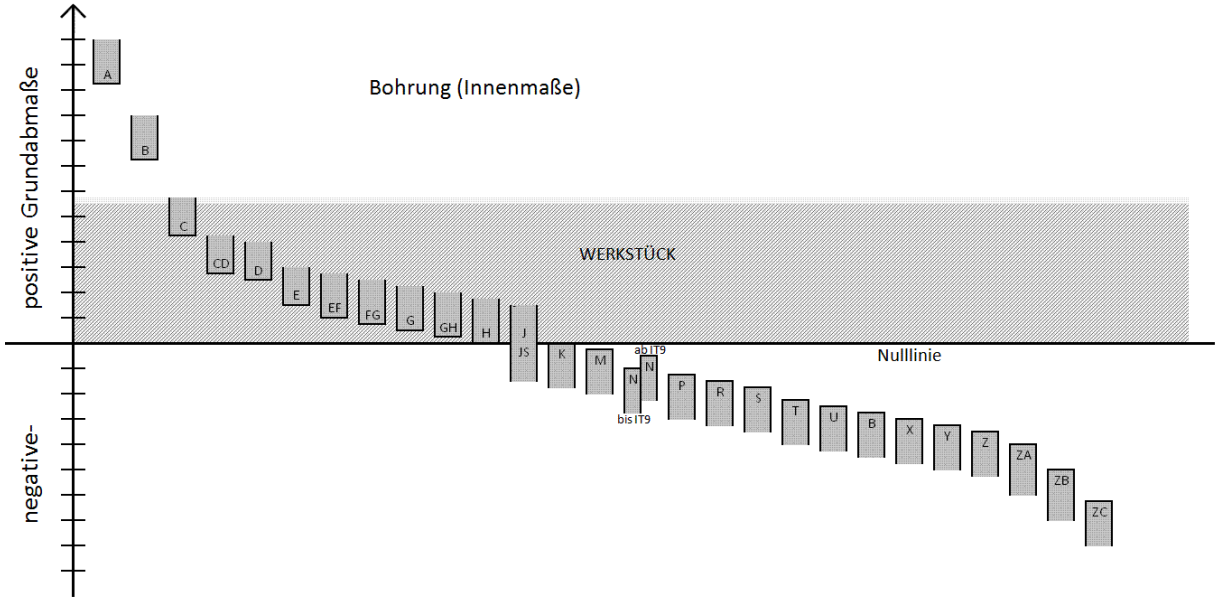


ν	1	2	3	4	5	10	20	50	∞
k	13,97	4,53	3,31	2,87	2,65	2,28	2,13	2,05	2,00

k-Werte zur Bestimmung der erweiterten Messunsicherheit $U_{95\%}$

Definitionen

- Maß: Bestimmung einer Länge
- Nennmaß: Zeichnungangabe
- Istmaß: tatsächliches Maß
- Oberes/unteres Abmaß: zulässige Abweichung - Achtung Vorzeichen!
- Mindestmaß: Nennmaß - unteres Abmaß
- Höchstmaß: Nennmaß + oberes Abmaß



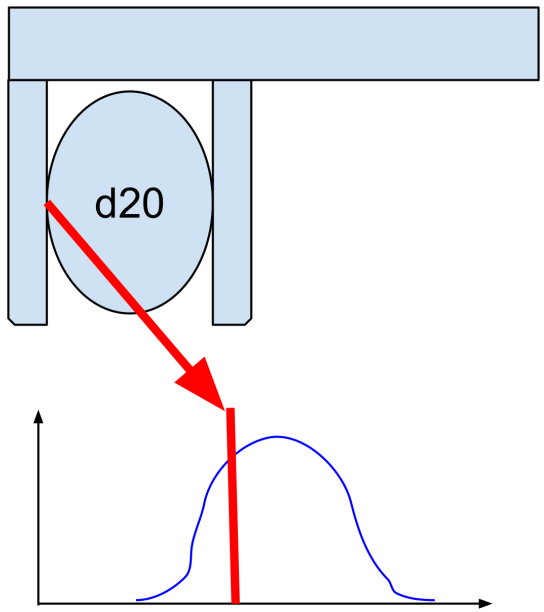
Abschnitt 5

Messunsicherheit

Messergebnis und Messunsicherheit

Der Vergleich einer Messgröße mit einer Einheit gelingt nicht fehlerfrei.

- Messgeräteabweichungen
- Instabilität der Messgröße
- Umwelteinflüsse (z.B. Temperatur)
- Beobachtereinflüsse
- Es werden unterschieden:
  - Systematische Messabweichungen
    - bekannt o. unbekannt
  - Zufällige Messabweichungen



Definition (Messergebnis)

Das Messergebnis ist der Schätzwert des wahren Wertes einer Messgröße.

Verfahren zur Abschätzung der Messunsicherheit

Verfahren A

- Abschätzung aus vorliegender Stichprobe
- n Messwerte liegen vor und können statistisch ausgewertet werden
- Die Messwerte sind näherungsweise normalverteilt

$$u = \frac{s}{\sqrt{n}}$$

s: Standardabweichung der Stichprobe

Verfahren B

- Ermittlung der minimalen/maximalen Messabweichung a durch Modellbildung
- Ermittlung von Grenzen und Form der Verteilung der Messabweichungen

$$u = \frac{a}{k}$$

$k = \sqrt{4}$ für Normalverteilung

$k = \sqrt{3}$ für Gleichverteilung

$k = \sqrt{6}$ für Dreiecksverteilung

Schätzfunktionen

Eine Schätzfunktion (Schätzer) dient zur Ermittlung eines Parameter-Schätzwertes aus empirischen Daten

- Grundlage: endlich viele Beobachtungen (Stichprobe)
  - Schätzer selbst fehlerbehaftet
  - Häufig Zufallsvariable
- Schluß auf Grundgesamtheit
- Schätzen einzelner Parameter der Verteilung
  - Mittelwert
  - Median
  - Standardabweichung

Definition (Zufallsvariable)

Als Zufallsvariable bezeichnet man eine messbare Funktion von einem Wahrscheinlichkeitsraum in einen Messraum.

Definition (Schätzfunktion)

Eine Schätzfunktion dient dazu, aufgrund von empirischen Daten einer Stichprobe einen Schätzwert zu ermitteln und dadurch Informationen über unbekannte Parameter einer Grundgesamtheit zu erhalten.

Schätzfunktionen und Eigenschaften

Gängige Schätzfunktionen und wünschenswerte Eigenschaften

- Mittelwert

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

$$\hat{\mu} = \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

- Varianz

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

$$\hat{\sigma}^2 = s_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2$$

- Erwartungstreue:
  - Erwartungswert der Schätzfunktion gleich wahrem Parameter
  - Kein systematischer Fehler (Bias).
- Konsistenz:
  - Unsicherheit des Schätzers nimmt für $n \to \infty$ ab
- Effizienz:
  - Minimale Varianz des Schätzers
- BLUE: Best Linear Unbiased Estimator

Bestimmung der Messunsicherheit durch Modellbildung

- Ermitteln der Einflussgrößen:
  - Ermittlung systematischer Abweichungen, ggf. kompensieren.
  - Einfluss des Normals
  - Wiederholpräzision (Messunsicherheit bei wiederholter Messung an einem Messobjekt)
  - Temperatureinfluss
- Modellbildung Messgröße $y = f(x_1, x_2, \ldots, x_n)$
- Bestimmung kombinierte Standardunsicherheit u_c

$$u_c = \sqrt{\left(\frac{\partial f}{\partial x_1} u_{x_1}\right)^2 + \left(\frac{\partial f}{\partial x_2} u_{x_n}\right)^2 + \ldots + \left(\frac{\partial f}{\partial x_n} u_{x_n}\right)^2}$$

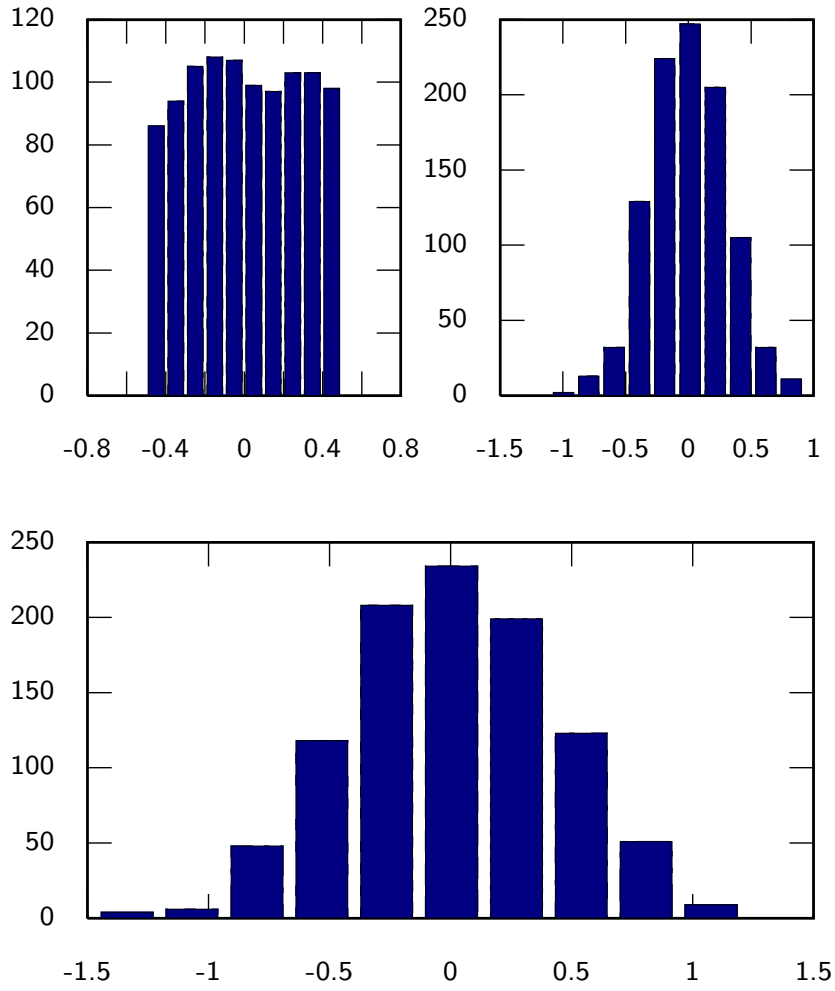
- Vereinfachung für Linearität von f und gleichem Gewicht der x_i

$$u_c = \sqrt{u_{x_1}^2 + u_{x_2}^2 + \ldots + u_{x_n}^2}$$

Monte-Carlo-Simulation: Motivation

Monte-Carlo-Simulationen (MC-Simulationen) können zur Ermittlung kombinierter Wahrscheinlichkeiten eingesetzt werden.

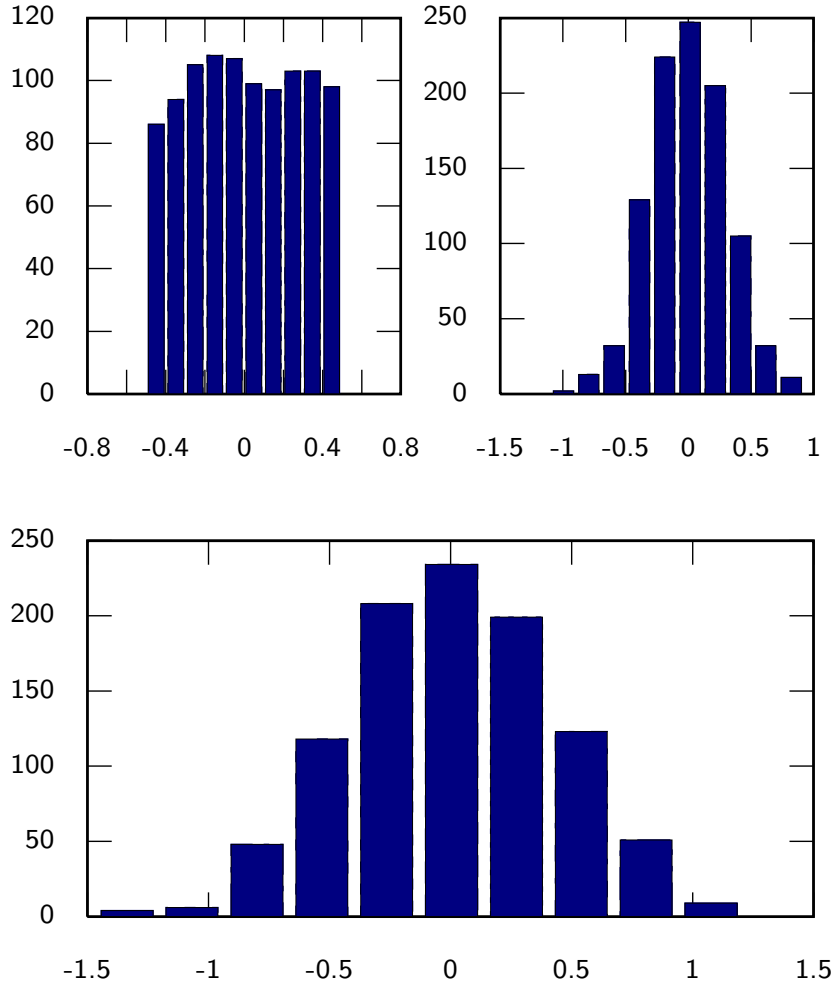
- Statistische Verfahren:
  - Hohe Stichprobenzahlen für Validität nötig
  - Zeitaufwändig
  - Teuer
- Fehlerfortpflanzung:
  - Analytisch teils komplex
  - Konservativ bei nicht normalverteilten Zufallsvariablen:
    - Annäherung Gleichverteilung durch Normalverteilung



Monte-Carlo-Simulation: Grundlagen

MC-Simulationen basieren auf der häufig wiederholten Simulation von Zufallsexperimenten.

- Grundlage:
  - Gesetz der großen Zahlen
  - Pseudozufallsverteilungen der Parameter erzeugen
  - Resultierendes Ergebnis bilden
  - Analyse der Lage- und Streuparameter
- Vorteile:
  - Einfache Modellierung
  - Einfache Durchführung
- Nachteil:
  - Korrektheit der Lösungen nicht immer einfach zu beurteilen



Vergleich der Verfahren des GUM und MC-Simulationen

MC-Simulation

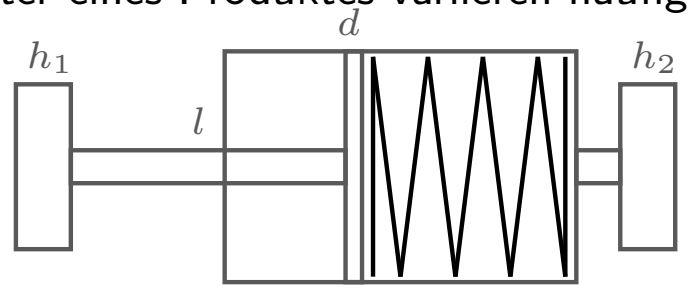
- Eingangsgrößen x_i werden explizit W-Dichtefunktionen zugeordnet
- Keine partiellen Ableitungen benötigt, können bereitgestellt werden
- Keine Einschränkung in Form der Verteilung von y
- Allgemeine, kürzest mögliche, Konfidenzintervalle

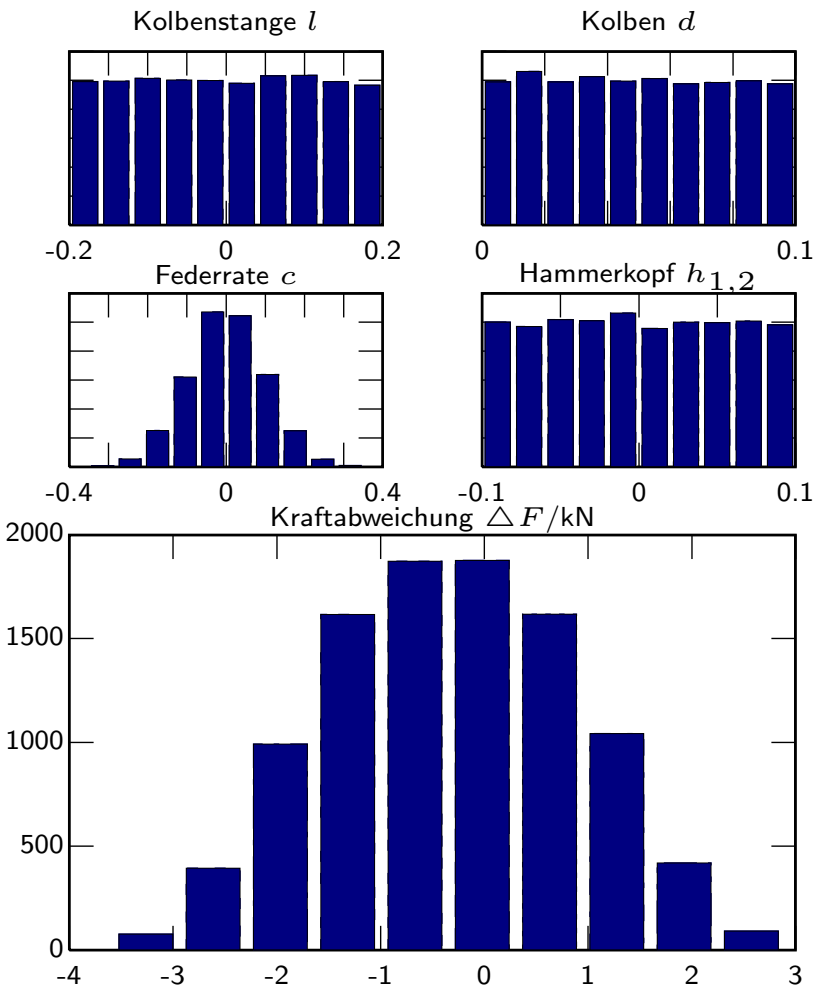
GUM Verfahren A/B

- Eingangsgrößen x_i werden Schätzwert und Standardunsicherheit zugeordnet
- Empfindlichkeitskoeffizienten und Taylor-Approximation benötigt
- Beschränkung auf Gauß- oder t-Verteilung
- Um den Schätzwert symmetrische Konfidenzintervalle

Monte-Carlo-Simulation: Beispiel

Parameter eines Produktes variieren häufig in Form von Gleich- oder Dreiecksverteilungen.





- Kraft eines Federspeicherzylinders

$$\triangle F = c\,(\triangle h_1 + \triangle l + \triangle d + \triangle h_2)$$

- MC-Simulation mit N = 10000

MC-Simulation zur Bestimmung der Messunsicherheit

Nicht normatives Beiblatt 1 zu DIN V ENV 13005

- MC-Simulation als Ergänzung zu Verfahren A und B des "GUM", z.B. falls
  - Linearisierung des Modells zu unangemessener Darstellung führt, z.B. durch Sensibilität
  - Partielle Ableitungen schwierig oder unmöglich zu finden sind
  - die Wahrscheinlichkeitsdichte merklich von Gauß- oder t-Verteilung abweicht , z.B. durch Asymmetrie
  - Wiederholte Versuche zur Bestimmung der Messunsicherheit nicht möglich sind
  - Modell des Messprozesses nicht in explizite Form gebracht werden kann
  - Unsicherheitsbeiträge nicht näherungsweise von der gleichen Größenordnung sind
- MC-Simulation ist im Einklang mit GUM
  - Ermittlung von Erwartungswert, Standardunsicherheit und Überdeckungsintervall der Messgröße y

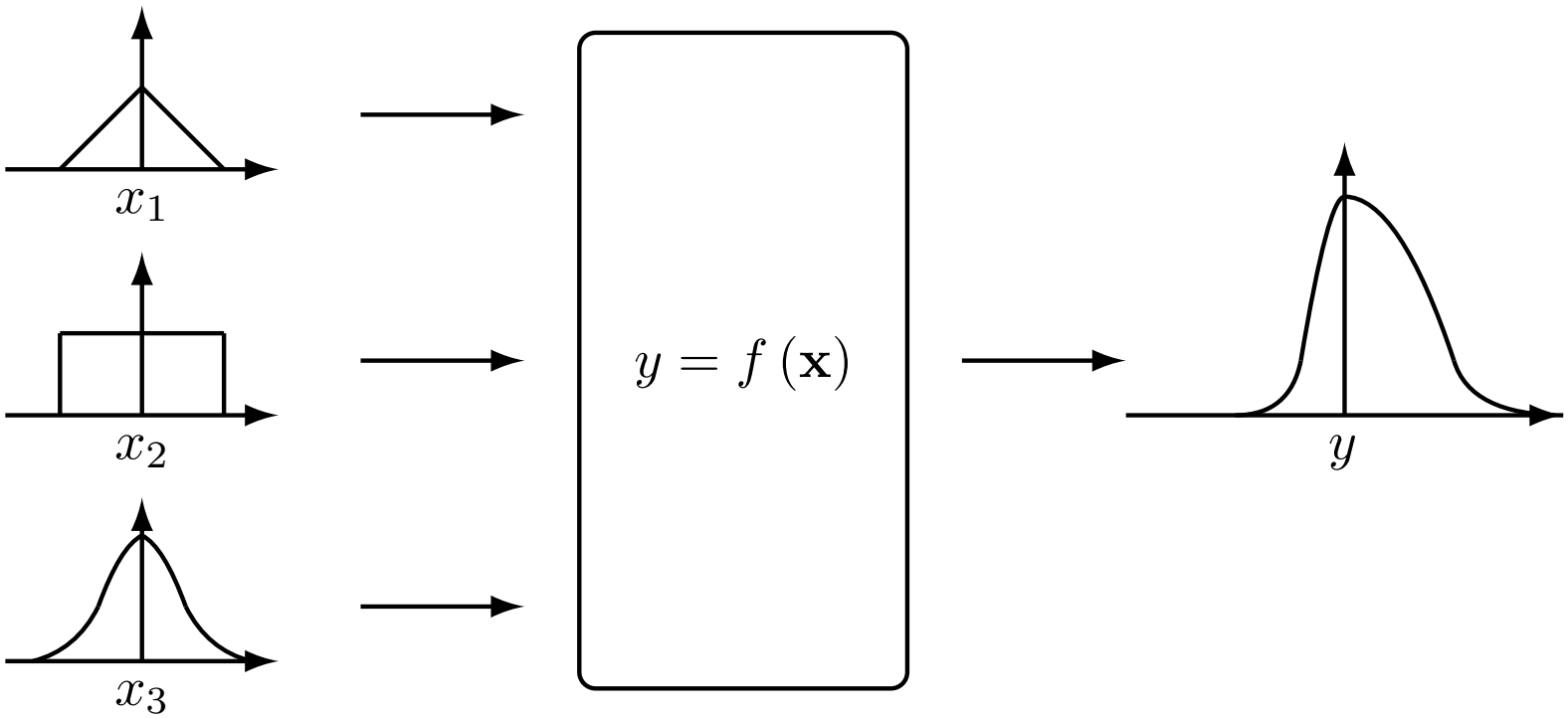
Ablauf Erstellung einer MC-Simulation

1) Modellierung des Systems:
   a) Ausgangsgröße y
   b) Eingangsgrößen $\mathbf{x} = (x_1, \ldots, x_i)^T$
   c) Modell als Beziehung zwischen y und $\mathbf{x}$, nicht zwingend explizit
   d) Zuordnung von Wahrscheinlichkeitsdichtefunktionen zu den x_i
      - x_i unabhängig: individuelle W-Verteilungen, z.B. Gauß-, Gleichverteilung etc.
      - x_i, x_j abhängig: Gemeinsame W-Verteilung
2) Fortpflanzung: Simulation des Systems
   a) Bestimmung des Wertes y aus den x_i
   b) Werte der x_i als Pseudozufallszahlen
3) Zusammenfassung:
   a) Bestimmung des Erwartungswerts $\mathrm{E}(y)$[1]
   b) Bestimmung der Standardabweichung ${\sigma_y}^2$
   c) Bestimmung des Konfidenzintervalls (Überdeckungsintervalls), das y mit einer festgelegten Wahrscheinlichlichkeit enthält

[1] Nicht alle W-Verteilungen weisen einen Erwartungswert auf.
[2] Nicht alle W-Verteilungen weisen eine Standardabweichung auf.

Bestimmung der Fortpflanzung

Neben analytischen und statistischen Methoden sind MC-Simulationen zulässig und effizient.
Verschiedene W-Verteilungen und nichtlineare Systemfunktionen machen sie notwendig.



W-Dichtefunktionen für die Eingangsgrößen

- Allgemein: $\mathbf{x}$ wird gemeinsame PDF zugeordnet
- Unabhängigen x_i werden einzelne PDF zugewiesen
- Allgemein: Bestimmung PDF gemäß Bayes-Theorem oder Prinzip der maximalen Entropie
- Praktisch häufig auftretende Fälle: [**?**, S. 34f.]
  - Rechteckverteilung: Toleranzbänder
  - $\arcsin$-Verteilung (U-förmig): Sinusförmige periodische Schwingungen
  - Gauß-Verteilung: Schwankung durch viele Einflussfaktoren (Zentraler GWS)
  - t-Verteilung: Modellierung aus endlich vielen Anzeigewerten
  - Exponential-Verteilung: Linkssteile Verteilung für positive Größen

Voraussetzungen für Anwendbarkeit und Gültigkeit der MC-Simulation

- Es ist f stetig bezüglich der x_i in der Nachbarschaft der besten Schätzwerte $\hat{x}_i$
- Die Verteilungsfunktion für y ist streng wachsend
- Die Wahrscheinlichkeitsdichtefunktion (PDF) für y ist
  - stetig über dem Intervall, für das die PDF streng positiv ist
  - unimodal
  - streng wachsend (oder 0) links vom Modalwert und streng fallend (oder 0) rechts vom Modalwert
- Es existieren $\mathrm{E}(y)$ und $\mathrm{Var}(y)$

  $\rightarrow$ Gewährleistet stochastische Konvergenz mit steigendem M
- Es ist die Zahl der Monte-Carlo-Versuche M ausreichend groß gewählt

  $\rightarrow$ Gewährleistet Zuverlässigkeit der stochastischen Information

Praktische Umsetzung einer MC-Simulation

Die hohe Anzahl Versuche M macht effiziente Software nötig.

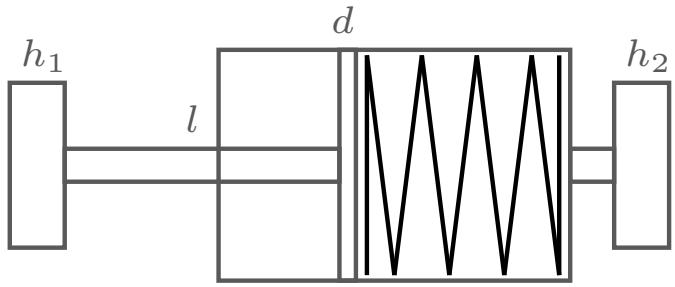
1. M wählen:
   - Da MC-Simulationen Zufallsexperimente sind, kann kein festes M Korrektheit des Algorithmus garantieren
   - Für Signifikanzniveau α: $M \gg \frac{1}{1-\alpha}$, z.B. $M \geq 10^4 \frac{1}{1-\alpha}$
   - Alternativ: Adaptive Verfahren
2. Erzeugung von Zufallszahlen für $\mathbf{x}_r$ aus den gewählten PDF
   - Pseudozufallszahlengenerator muss über gewisse Eigenschaften verfügen, z.B. Sequenzlänge
3. Bestimmung der Werte $y_r = f(\mathbf{x}_r)$ für $r = 1, \ldots, M$
4. Diskrete Darstellung der Verteilungsfunktion für die Ausgangsgröße
5. Schätzung der Ausgangsröße und der beigeordneten Standardunsicherheit
   - $\bar{y} = \frac{1}{M} \sum_{i=1}^{M} y_i$
   - $s^2(\bar{y}) = \frac{1}{M-1} \sum_{i=1}^{M} (y_i - \bar{y})^2$

Teil 1

Manufacturing data

Scenario

Please assume that you are working as a freelance quality consultant to a manufacturing company in the railway sector. Your client manufactures spring parking brake cylinders as depicted below.



A spring parking brake is used to maintain a rail vehicle in a stationary position, for movements, the brake is released by applying compressed air to the cylinder. For this purpose, one of the determining parameters of a spring parking brake the force is the force applied to the so called hammerheads of the brake cylinder. This force has to maintain the vehicle stationary. For the product under consideration, the required minimum force is $18\,\mathrm{kN}$.

Tasks

Data import and graphical representation

You receive a call from you client because in todays production shift, more than 10% of the production had to be scrapped due to insufficient force. The client sends you a .csv file containing the quality records of the shift. The file contains the following data:

Time, Date	F/N	$\triangle d$/mm	$\triangle h_1$/mm	$\triangle h_2$/mm	$\triangle l$/mm	x/mm

$$x = h_1 + h_2 + d + l$$

In order to start your analysis, you import it and inspect it qualitatively for randomness and systematic influences. Your customer is rather fluent in Python, so you decide to exchange Jupyter Notebooks highlighting your findings.

Tasks

Optional: Monte-Carlo simulation

The client considers a systematic behaviour of the spring stiffness as a possible root cause of the force loss. Each cylinder contains 24 springs of three types, each being supplied with a $\pm 10\,\%$ tolerance on the nominal value:

- 8 small springs: $c_{1,\text{nom}} = 100\,\mathrm{N/mm}$
- 8 medium springs: $c_{2,\text{nom}} = 150\,\mathrm{N/mm}$
- 8 large springs: $c_{3,\text{nom}} = 200\,\mathrm{N/mm}$

Perform a Monte Carlo Simulation of the combined spring stiffness c and its variation. Assume two cases:

C_1 The supplier ships springs equally distributed within the tolerance range.

C_2 The supplier ships large springs within tolerance, however the distribution is skewed towards the lower end of the spectrum, yielding forces with $c_3 \in [180, 190]\ \mathrm{N/mm}$. All other springs are as described above.

Hint: remember to draw the individual springs of a set from an independent and identically distributed (i.i.d.) random variable.

Tasks

Using support vector machine to predict an unobserved product property

esting of the cylinders for their effective force is a costly task. The customer would prefer to sort the components such that only capable cylinders are assembled. They propose to use the hammerhead dimensions h_1, h_2 as well as the length l to predict the force based on learned data.
For this purpose:

- Import the existing dataset
- Scale the data using a standard scaler
- Train a support vector machine based on this data
- Use the new data set (unknown to the SVM) "SpringPBDataValidation.csv" to test the performance
  - You may also generate test data using a Monte-Carlo simulation
- Plot a confusion matrix of the results

How do you like the performance of your classifier? Can you manufacture without end of line testing? Is your data basis sufficient to exclude any non-conforming cylinders from being shipped to the customer?

Literatur I

[1] *DIN V ENV 13005: Beiblatt 1 - Fortpflanzung von Verteilungen unter Verwendung einer Monte-Carlo-Methode*.
DIN Deutsches Institut für Normung e.V., 2012.

[2] T. Pfeifer and R. Schmitt.
Fertigungsmesstechnik.
Oldenbourg Verlag, München, 2010.