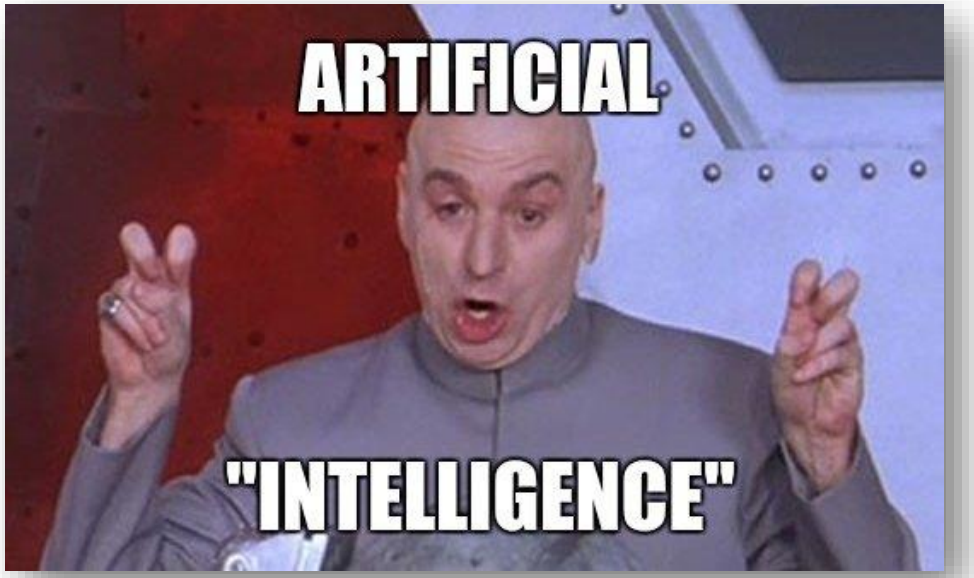
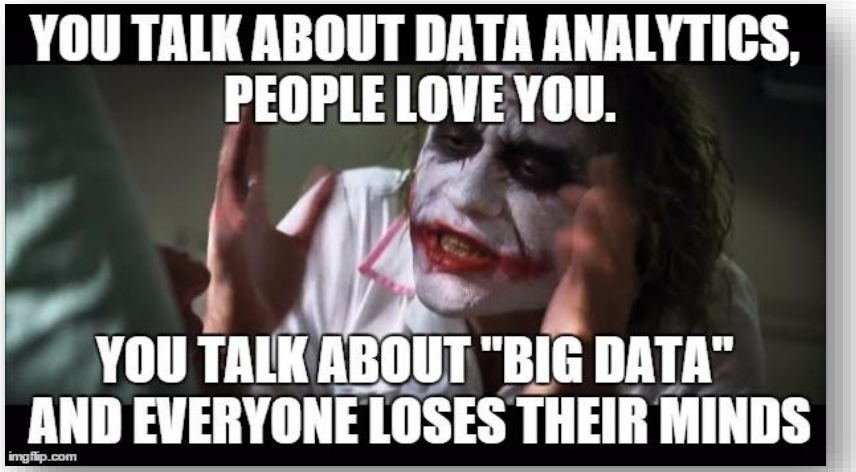


Background, Dimensionality Reduction

MACHINE LEARNING



Topics

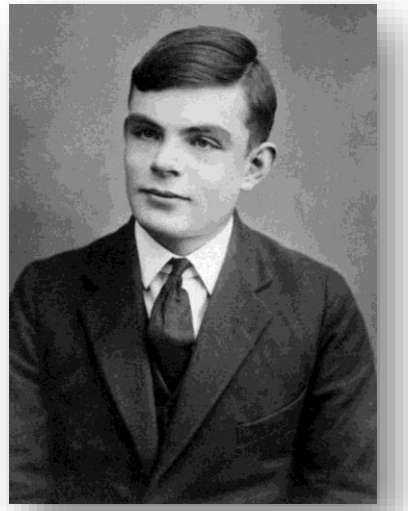
- Introduction: GUI and basic calculations
- Coding 1: Scripts, style, and variable classes
- Coding 2: Control statements and loops
- Visualization 1: Basics, subplots, get and set
- Coding 3: Functions
- Visualization 2: Descriptive plots
- Coding 4: Basic input and output
- Visualization 3: Distribution and 3D plots
- Coding 5: Input and output specials – last lecture before holidays
- Machine Learning 1: Introduction and dimension reduction
- Machine Learning 2: Clustering
- Machine Learning 3: Classification
- Coding 6: Efficiency and debugging basics
- Coding 7: Advanced functions and debugging

Artificial Intelligence

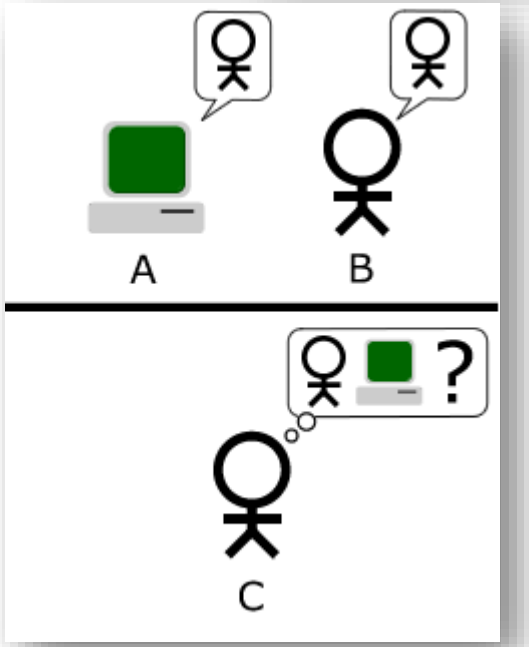
- Intelligence in machines
  - Contains reasoning, planning, language processing, perception and action control (amongst others)
  - „Intelligent“ is not as defined as it is in psychology
- Weak AI (narrow problem solver, expert systems) vs. Strong AI (conscious?, general?)
  - Ongoing debates, no clear conclusions
- AI as a buzzword is used very widely, "AI is whatever hasn't been done yet.“ (Tesler's Theorem)
  - Anything with learning and problem solving
  - Anything to sell to your CEO
- Philosophical considerations about “Artificial Intelligence/Consciousness” often have nothing to do with what’s actually there
- Artificial Intelligence (AI) ~= Machine Learning (ML)
  - ML can be a tool to create AI

The Turing Test (Imitation Game)

- Alan Turing (1912-1954)
  - Father of theoretical computer science and AI
  - Turing Machine (theoretical general purpose computer)
  - Tragic death (cyanide poisoning) after chemical castration (against homosexuality)
  - Rehabilitation by the Queen only in 2013!
- Test whether a machine would exhibit intellgient behavior that is indistinguishable from a human
- Interrogation by a human of two conversation partners
  - One AI and one human
  - Text-only
  - Human judge decides who is the machine
  - If indistinguishable, test is passed
- Pass ~= Strong AI (c.f. Chinese Room)
  - Generalized AI probably
- Rather a theoretical question than serious research!



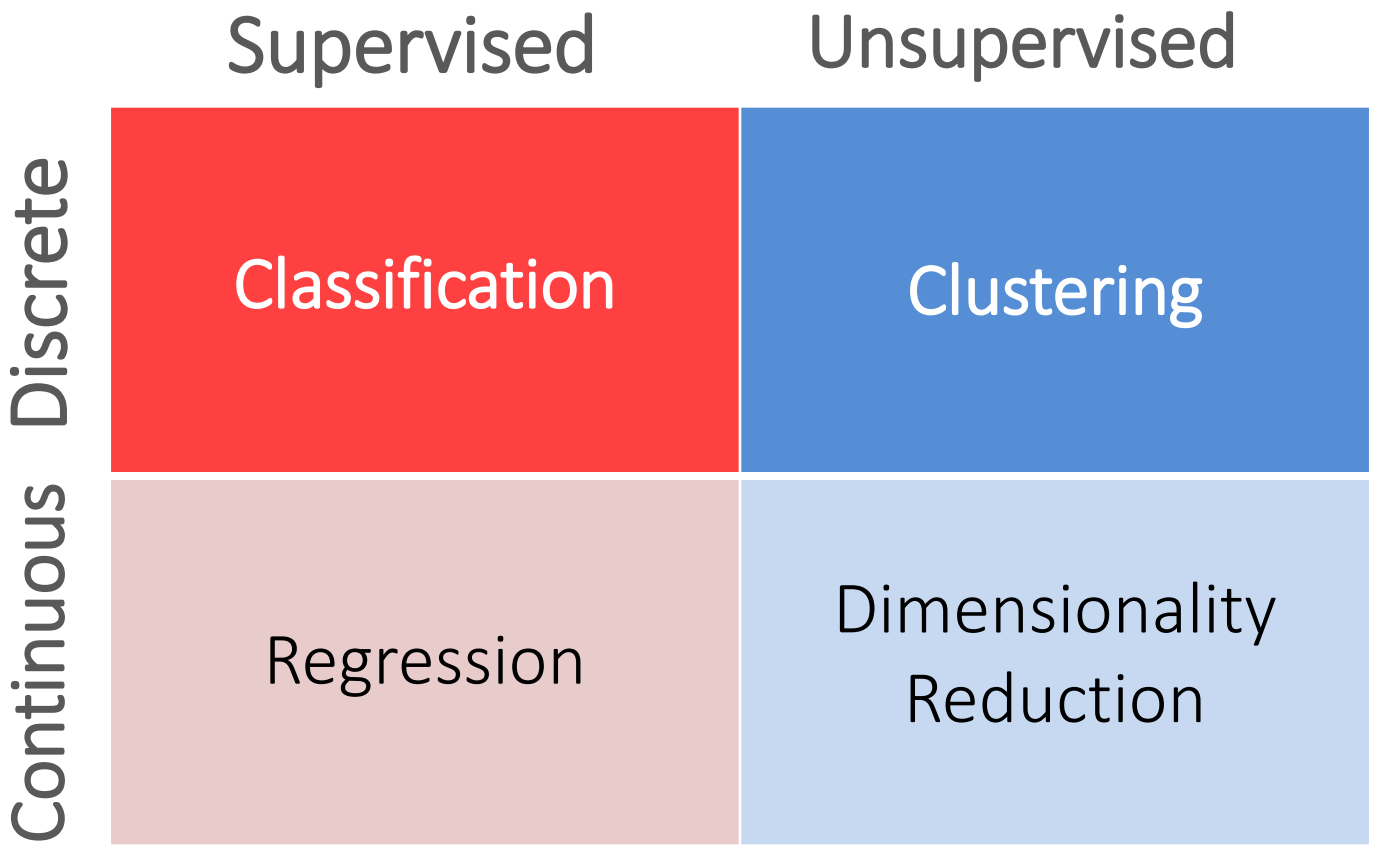
https://de.wikipedia.org/wiki/Alan_Turing



https://de.wikipedia.org/wiki/Turing-Test

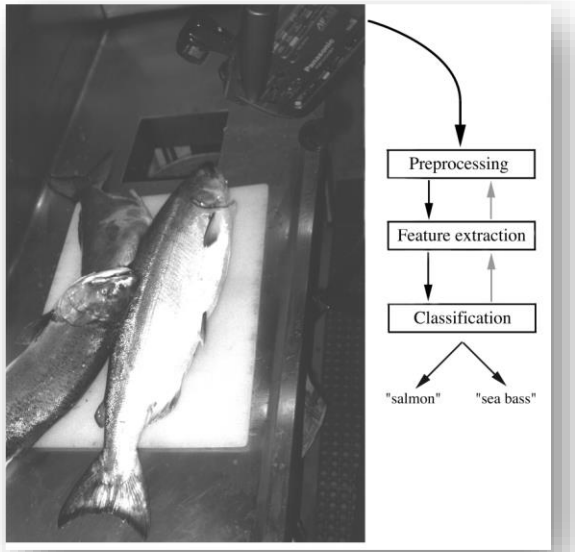
Machine Learning (ML)

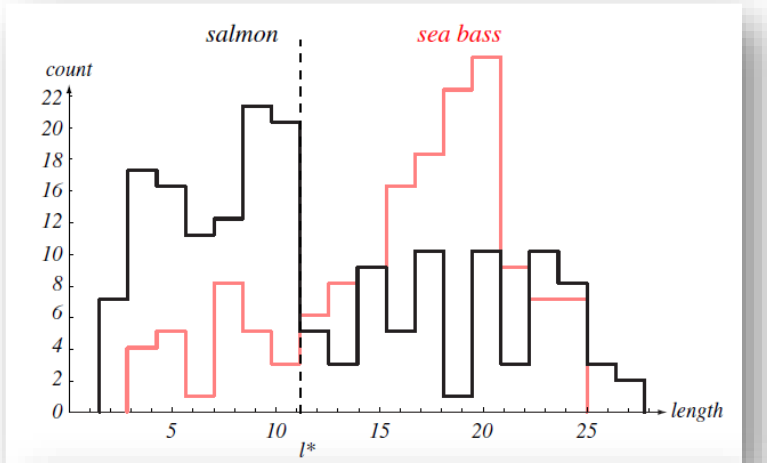
- "A computer program is said to learn from experience *E* with respect to some class of tasks *T* and performance measure *P* if its performance at tasks in *T*, as measured by *P*, improves with experience *E*.“ (Tom Mitchell, 1997)
- Learn from training data and test on new data
- Unsupervised learning (**dimensionality reduction, clustering**)
- Supervised learning (**classification, regression**)
- Reinforcement Learning (more relevant for agents in an environment)

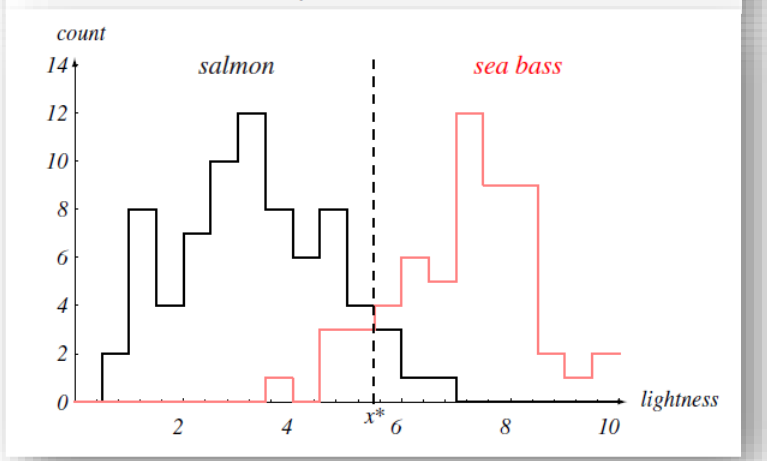

Supervised
Unsupervised
Discrete
Continuous
Classification
Clustering
Regression
Dimensionality Reduction

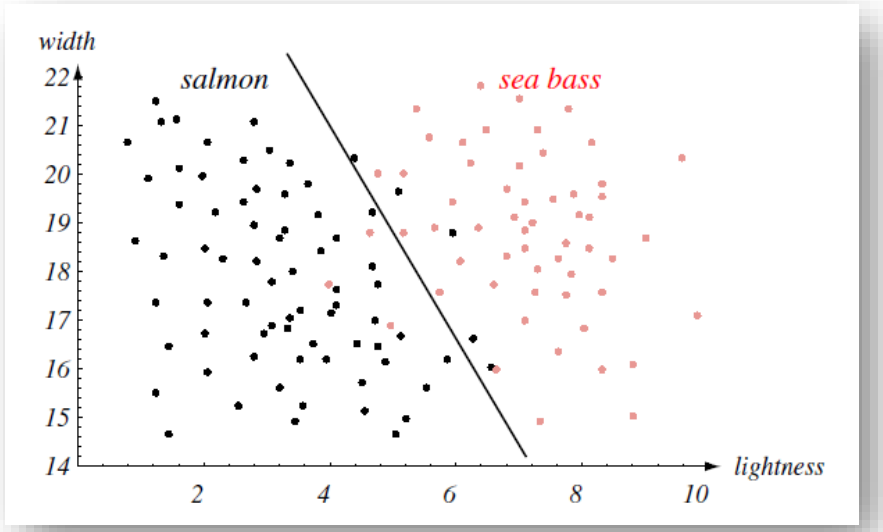
Example

- Discriminating between fish (From: Duda et al., Pattern Classification, 2nd ed., Wiley, 2001)



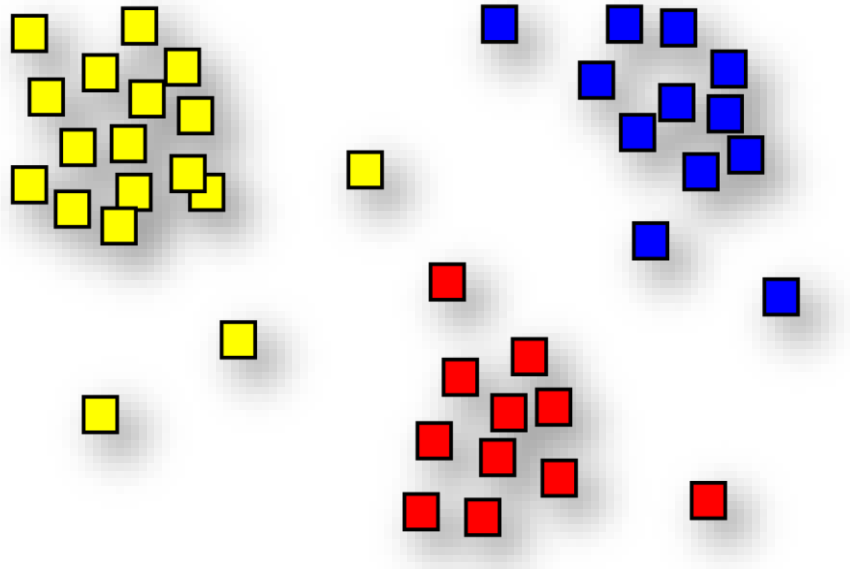




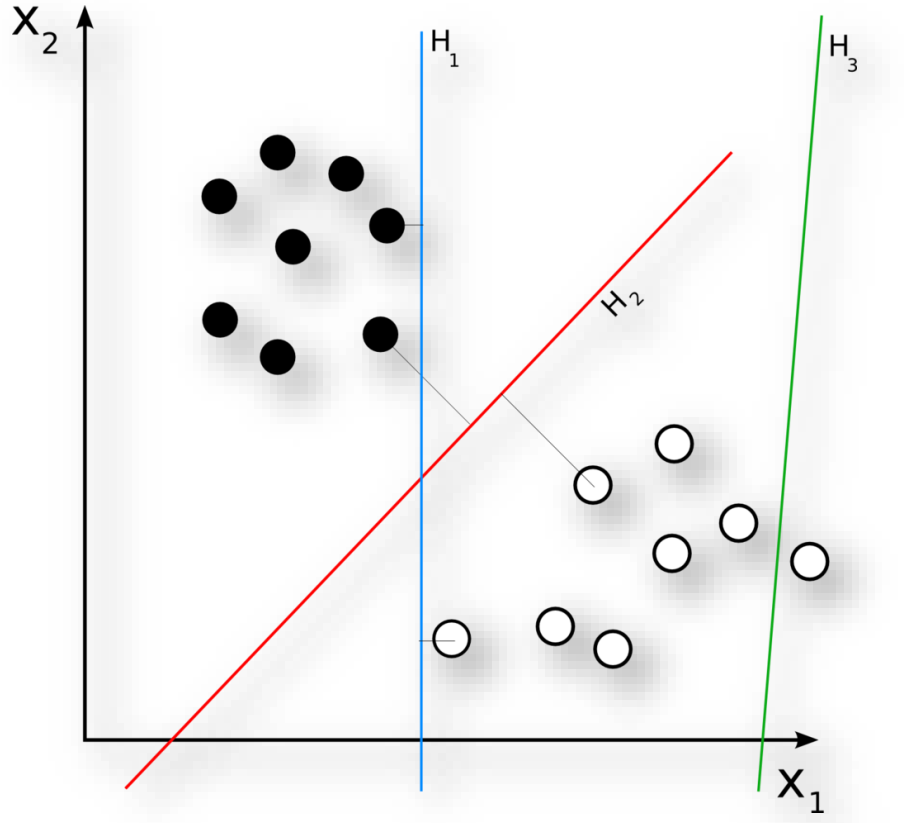


Clustering & Classification

- Clustering: What data points belong together?
- Classification: Which category does this data point belong to?



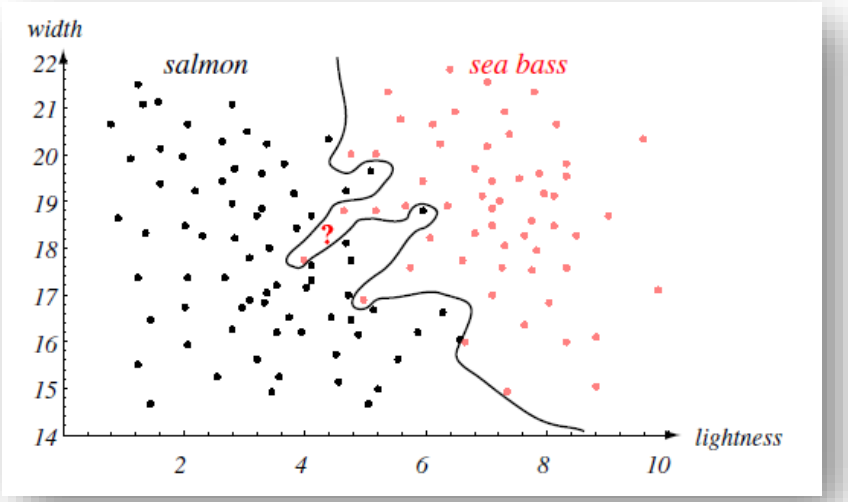
https://en.wikipedia.org/wiki/Cluster_analysis



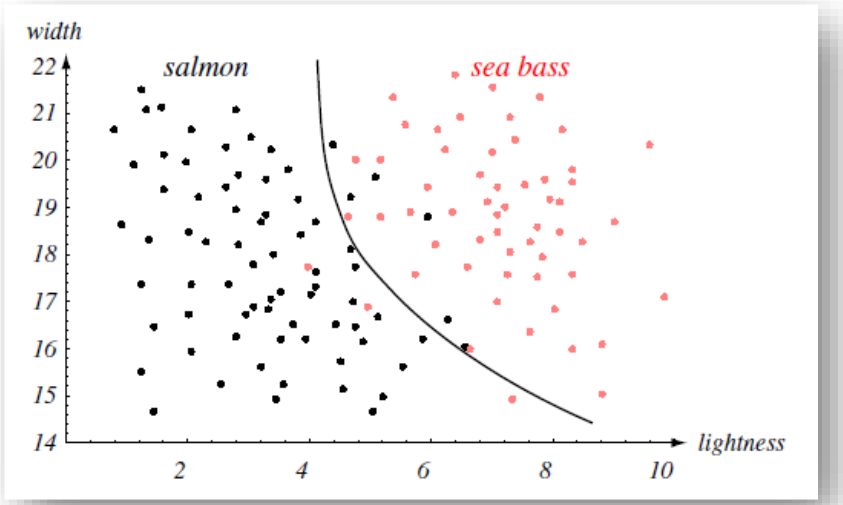
https://en.wikipedia.org/wiki/Linear_classifier

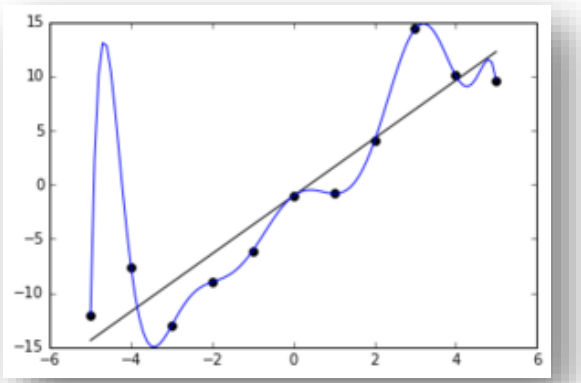
Overfitting

- Your model corresponds too closely to the training data, but does not generalize to new data (bad predictions)
- Training Error ~= Testing Error
- Model complexity << data set size
  - E.g. factor 10



Duda et al., Pattern Classification, 2nd ed., Wiley, 2001

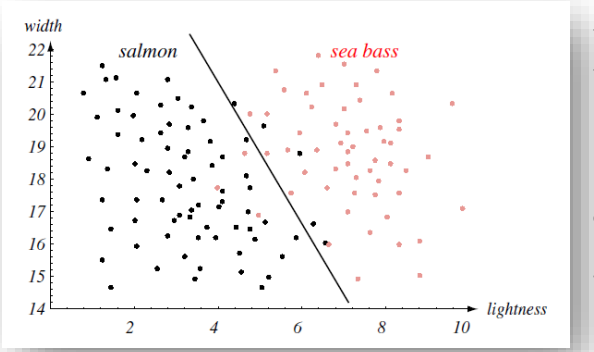




https://en.wikipedia.org/wiki/Overfitting

Features and Dimensions

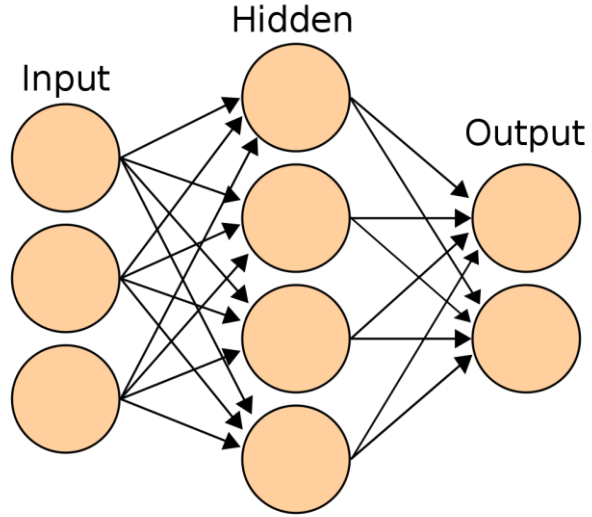
- Feature Extraction
  - You have a raw data set, what are the informative values that you derive from it?
    - Depends on the question!
  - EEG data -> e.g. power in certain frequencies
- Feature Selection
  - Select a subset of relevant features (variables) for your model
  - Less features (variables) -> Faster computation, less overfitting
- Curse of dimensionality: If the feature space has a very large amount of dimensions, problems arise
  - E.g. because the data does not fill the entire volume or distances are hard to define
  - **Dimensionality reduction**: Reducing the number of dimensions of your feature vector while keeping as much information (variance) in your data as possible



Duda et al., Pattern Classification, 2nd ed., Wiley, 2001

Artificial Neural Networks

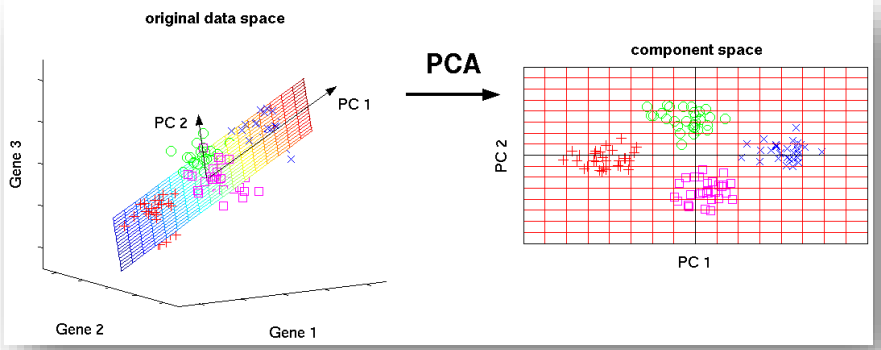
- "neurons" connect with weights to other neurons
  - Input, Anonymous, and output layer
  - Other structures are possible too!
- "Activity" (numeric value) spreads through the net
  - $p_j(t) = \sum_i o_i(t) w_{ij}$ & $a_j(t+1) = f(a_j(t), p_j(t), \theta_j)$
- Errors are backpropagated to the input
  - The contribution of each weight and neuron is computed
  - Weights are changed in the process to minimize the error
- Deep Learning: Many Anonymous Layers
  - Performs feature selection and extraction itself
  - Needs VERY large amount of data
  - Hard to interpret how exactly tasks are solved
  - Think first whether a simple algorithm can work as well!

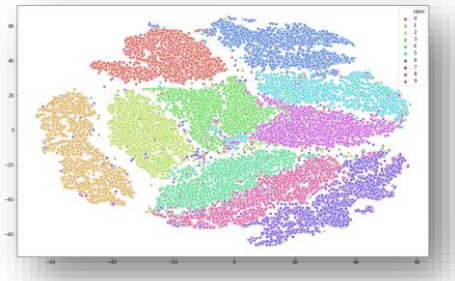


https://en.wikipedia.org/wiki/Artificial_neural_network

Dimensionality Reduction

- Big data -> manageable data (**visualize**, classify)
- Principal Component Analysis (PCA)
  - Very commonly used (invented in 1901)
  - Correlated variables are transformed into a set of uncorrelated variables (principal components)
  - Components are sorted according to their variance
  - Disregarding components with low variance contribution reduces the dimensions while keeping data variability high
- t-Distributed Stochastic Neighbor Embedding (t-SNE)
  - Projects n-dimensional data into 2D or 3D
  - Similar objects are modeled by nearby points and dissimilar objects are modeled by distant points with high probability
  - https://distill.pub/2016/misread-tsne/





https://chanzuckerberg.github.io/scRNA-python-workshop/preprocessing/02-normalization.html

https://towardsdatascience.com/dimensionality-reduction-using-t-distributed-stochastic-neighbor-embedding-t-sne-on-the-mnist-9d36a3dd4521