Final exam

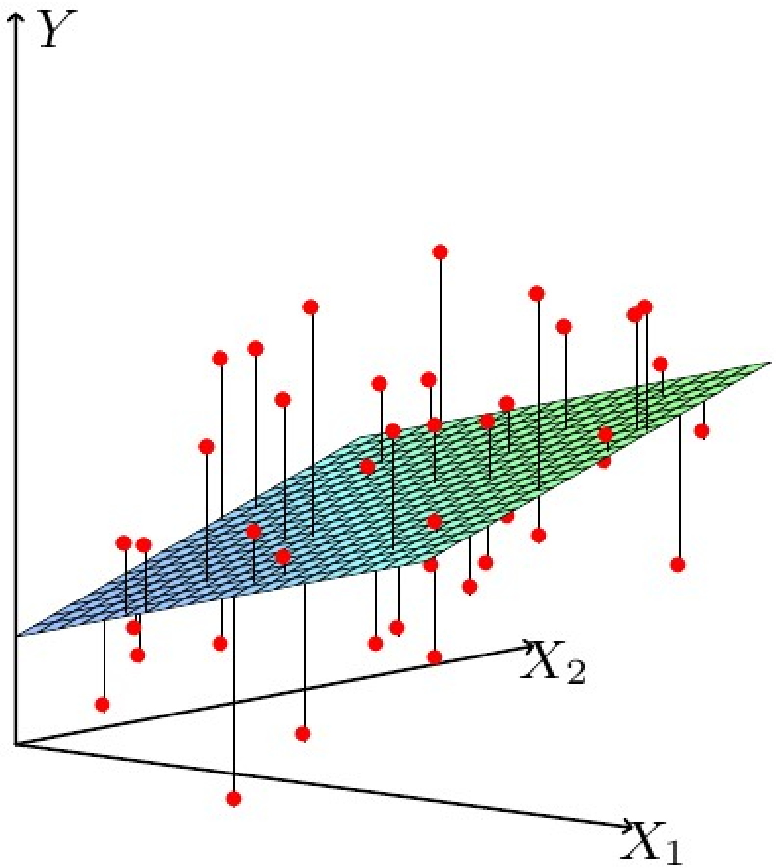
- Do not forget to hand in your projects by **January 15 or 2 days** before the final exam, whichever comes first.
- You can bring a cheat sheet (2 sheets of A4, both sides)
  + theoretical computer science cheat sheet
- **Written final exam:** model exam will be available next week
  40% of you total grade, you need to get at least 50% of points on the exam
  regular exam dates: 16.1.2020 (Thursday) 9:00-12:00
  22.1.2020(Wednesday) 9:00-12:00
  sign up in AIS2 latest 2 days before the exam
- Dates for repeat exams will be scheduled as needed.
- **To finalize your grade, we may ask you to come in to discuss and demonstrate your project.**

Tento rok
účasť
120%!
anketa
.fmph.uniba.sk

Course Summary

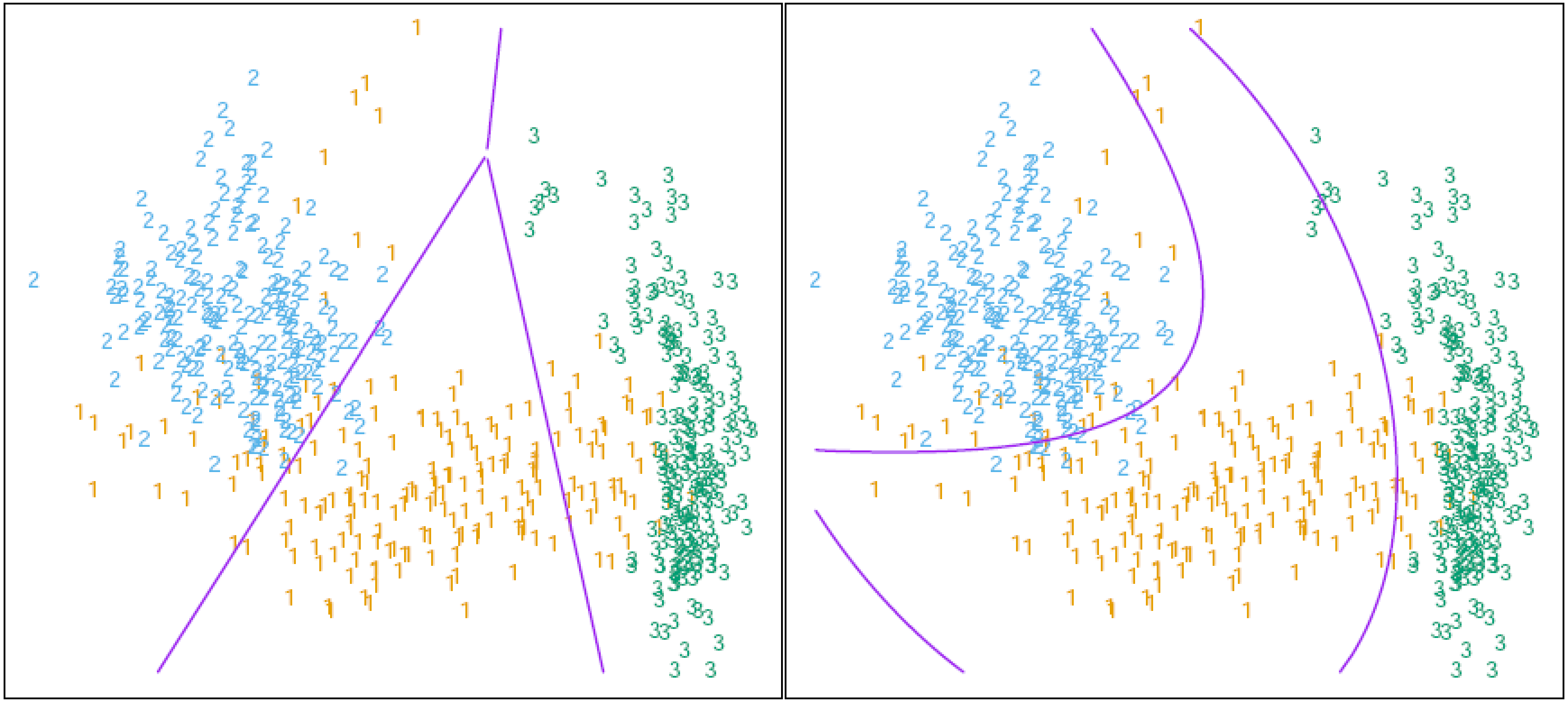
- **Supervised learning**
  - regression, classification
- **Unsupervised learning**
  - clustering, dimensionality reduction
- **Machine learning theory**
  - bias and variance, PAC learning, VC dimension
- **On-line learning and reinforcement learning**

Regression



- Linear regression
- Solving normal equations in O(n^3)
- Gradient descent
- Expansion of underlying vector space through non-linear transformation => generalized linear regression

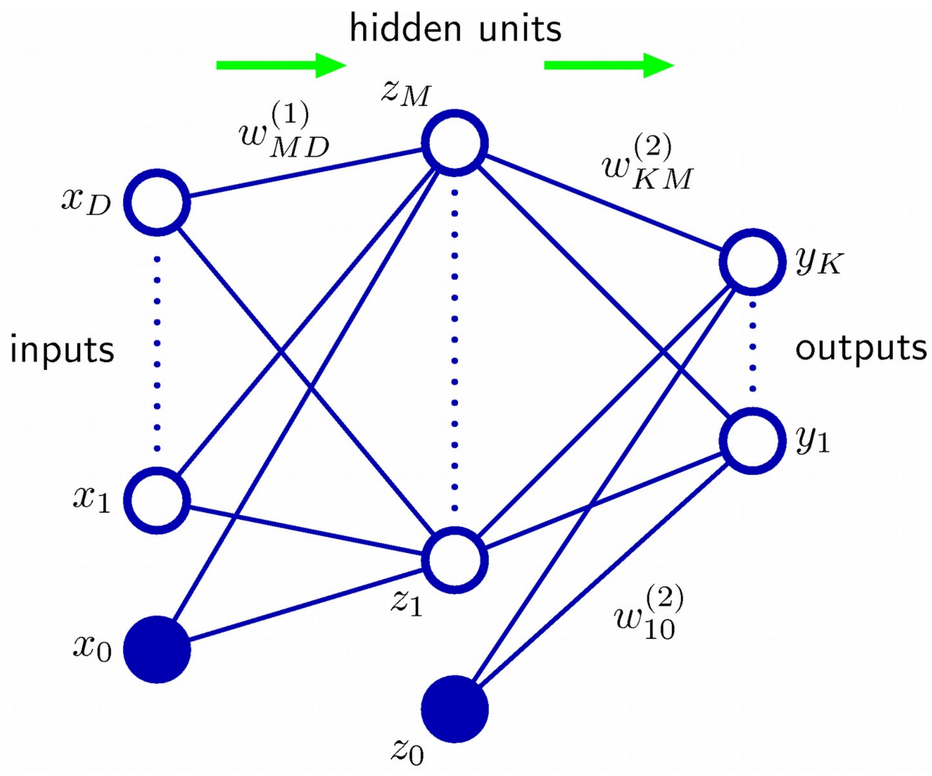
Classification



Linear classification

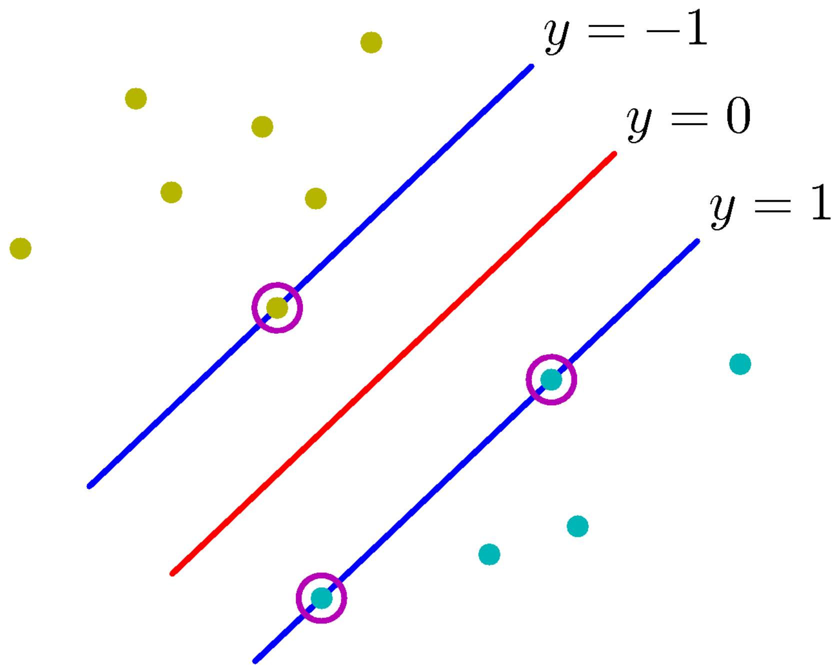
Using non-linear expansions

Neural Networks



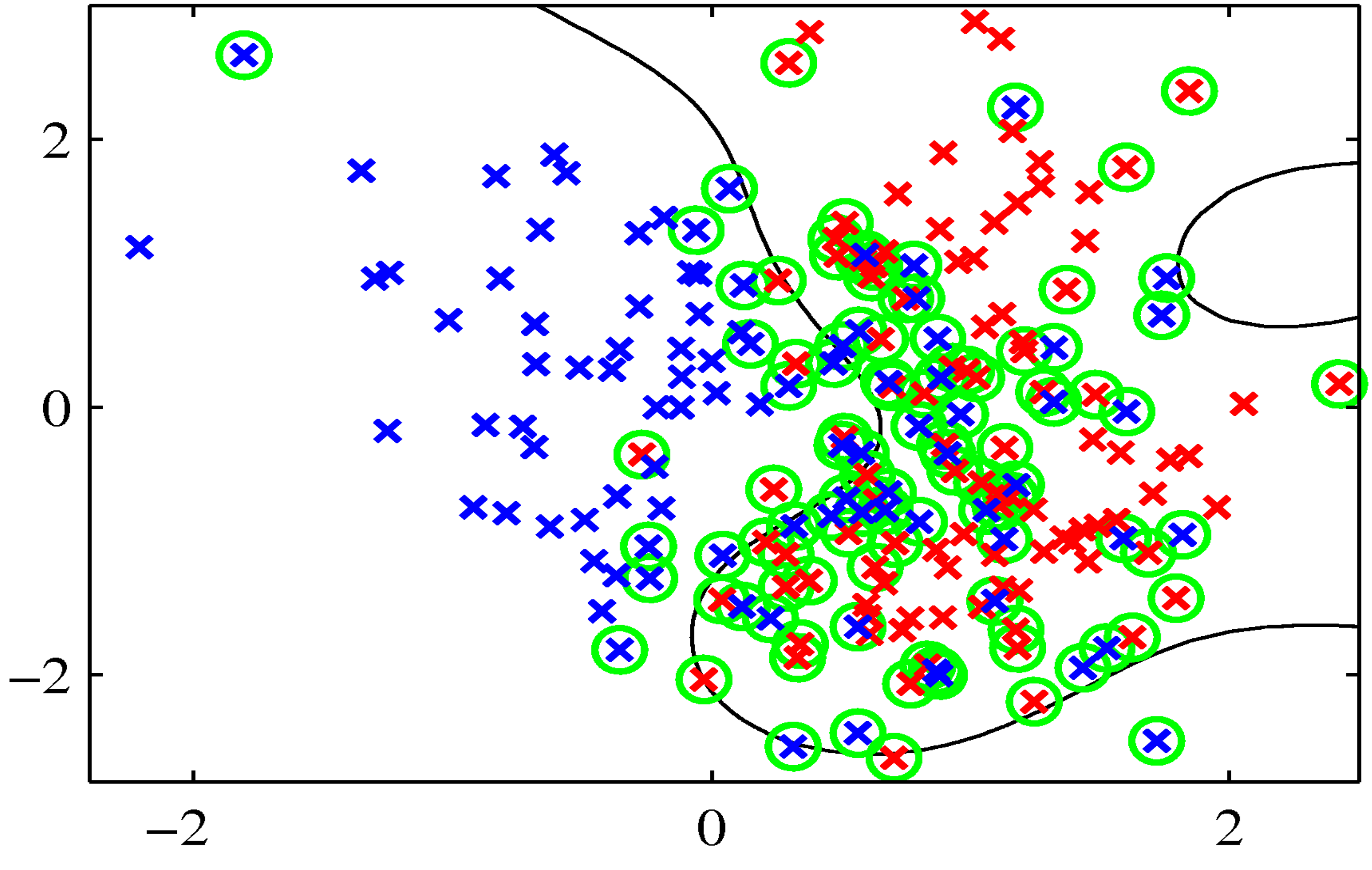
- Each unit ("neuron") - linear combination followed by non-linear transformation
- Gradient descent (so called "back propagation")

Support Vector Machines

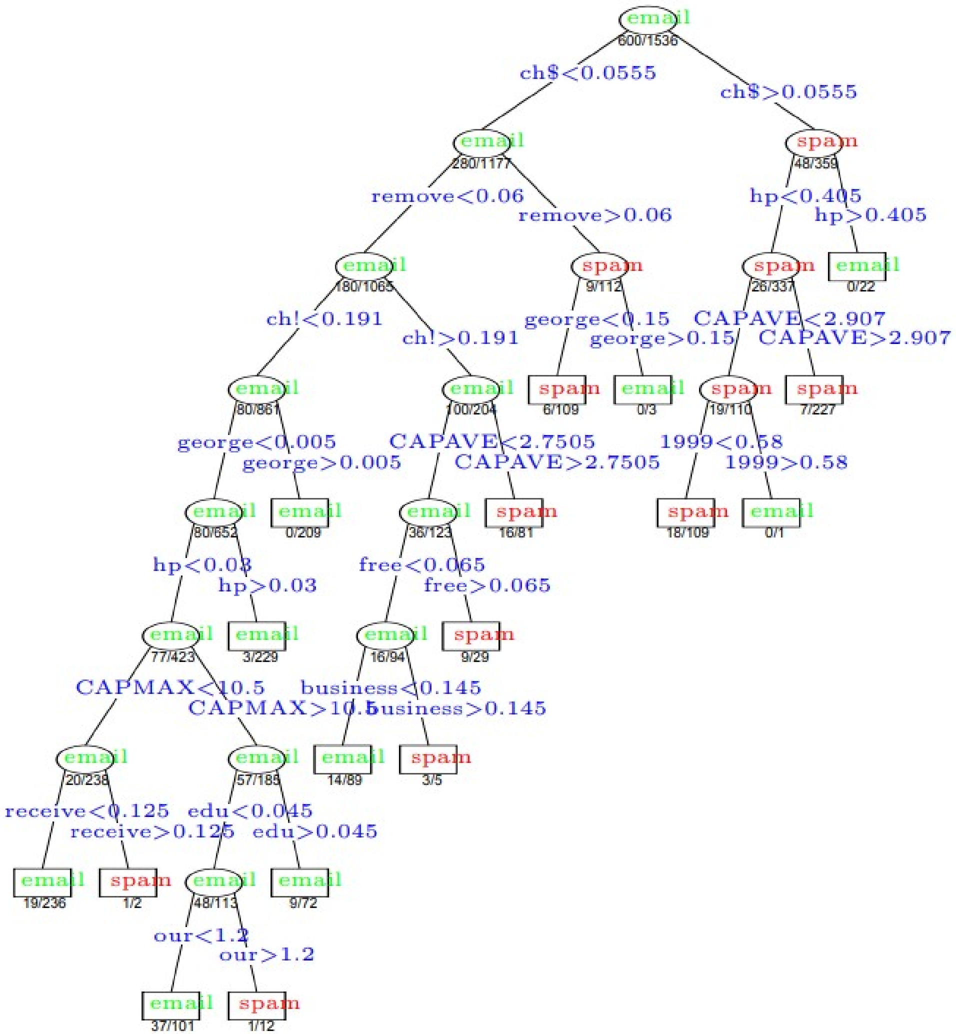


- Linear classifier maximizing margin
- Quadratic programming, dual programs
- Kernel trick: expansion into infinite dimensional vector space
  K(x,y) – dot product in the expanded space (intuition: similarity measure)

Support Vector Machines

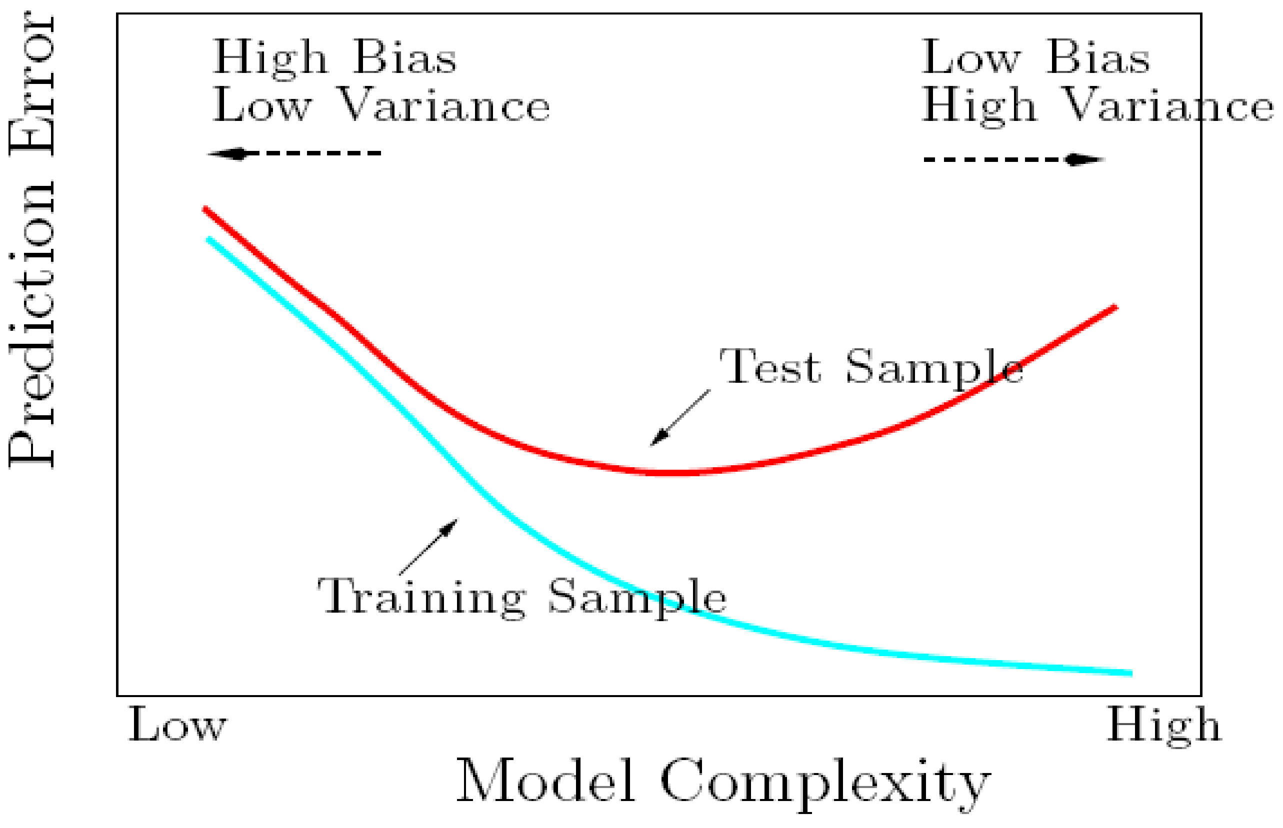


Decision Trees and Random Forests



- ID3 algorithm for building trees (based on entropy measure)
- Stopping criteria
- Bagging – ensemble of complex classifiers
- Boosting – ensemble of simple classifiers

Bias and Variance



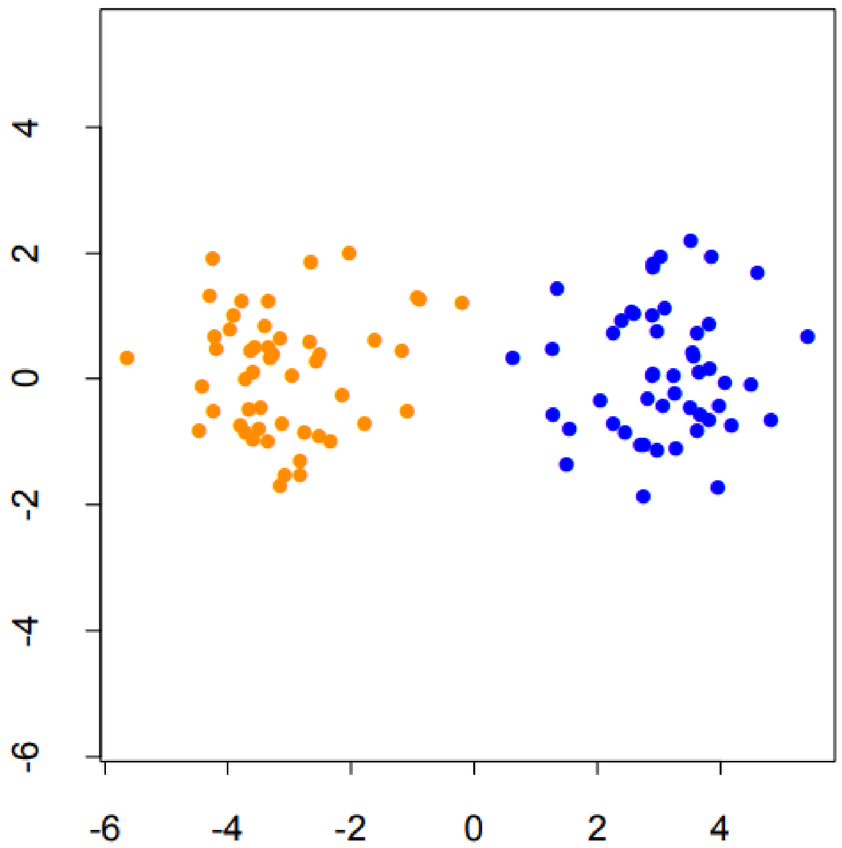
PAC Learning
(Probably Approximately Correct)

- How many training data points do we need to train a classifier?
- For large enough t,
  training and testing error
  with high probability (>1-δ)
  will not differ much (<ε)
- PAC learning theory provides bounds on t for specific H, ε and δ

PAC learning - bounds

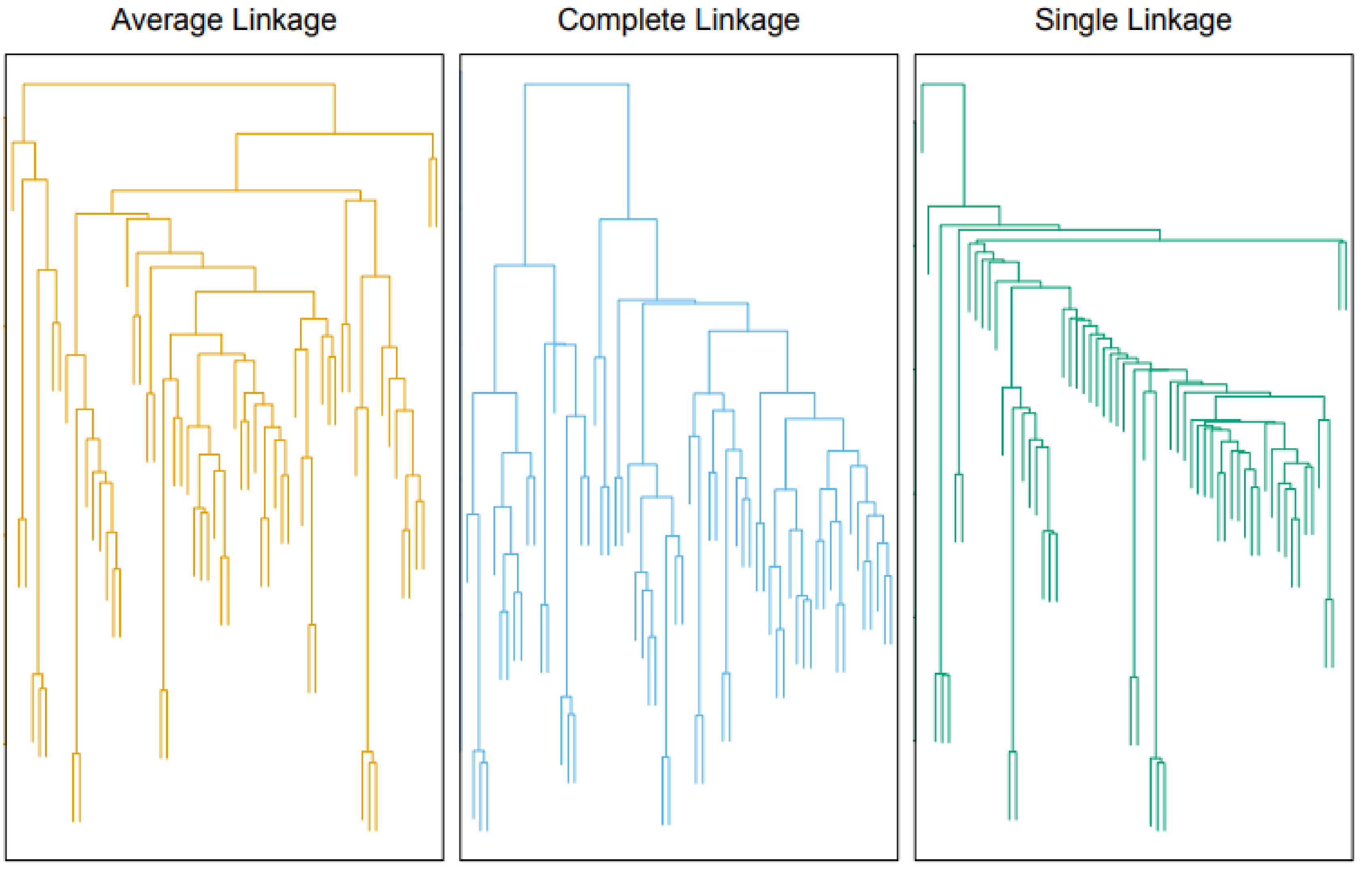
- **Finite hypothesis space:**
  t=O(log |H|)
- **Infinite hypothesis space:**
  - Vapnik-Červonenkis (VC) dimension d
    (t grow linearly with d)
    Neural networks: d=Θ(W.log n)
    (W – # weights, n – # sigmoids)
  - SVM: t=O(1/r^2 log^2 1/r)
    (r – margin size)

Clustering

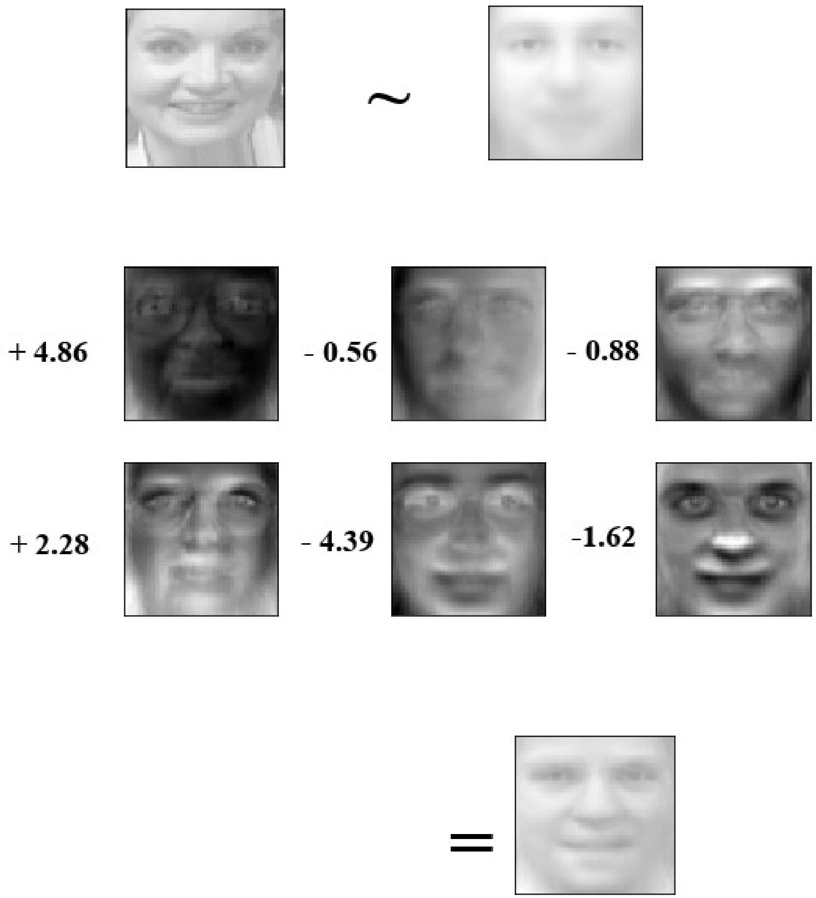


- K-means and k-medoids clustering
- Iterative methods to find a good solution
- Beware: slow!

Hierarchical clustering



Dimensionality reduction



- Principal Component Analysis (PCA)
- Kernel trick (again)
- Multi-dimensional scaling (i.e. t-SNE)

Mathematical Methods

- Matrix algebra, solving systems of linear equations
- Eigenvectors and eigenvalues
- Partial derivatives, Lagrange multipliers
- Numeral mathematics: Gradient descent
- Optimization: linear and quadratic programming, duality
- Analytical geometry
- Vector spaces

(Un)related Classes

Spring 2020:

- 2-INF-188: Current Topics in Machine Learning (Boža)
- 2-AIN-132: Neural Networks (Farkaš)
- 2-AIN-235: AI Algorithms in Robotics (Petrovič)
- (2021) 2-AIN-288: Speech Recognition

Fall 2020:

- 1-BIN-301: Methods in Bioinformatics (Vinař, Brejová)
- 2-AIN-268: Deep Learning in Computer Vision (Černeková)
- 2-PMS-129: Stochastic Optimization Methods (Harman)