Supervised Machine Learning Methods

Konrad U. Förstner

ZB MED – Information Centre for Life Science & TH Köln

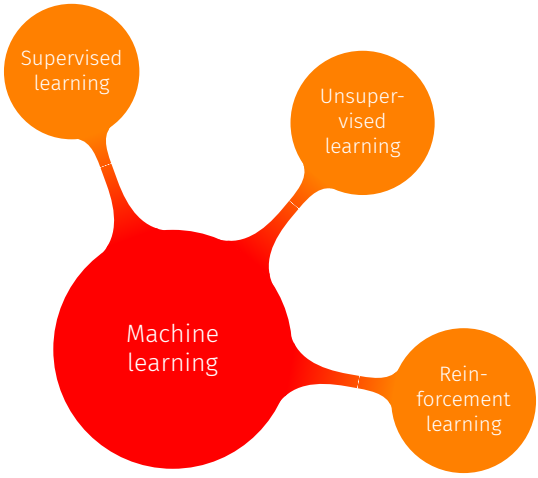
Workshop *Systems Biology: From large datasets to biological insight*
at the European Bioinformatics Institute
2019-07-11

ZB MED Information Centre for Life Sciences
Technology
Arts Sciences
TH Köln

Small survey before we start

Who is already familiar with ...

- ... supervised machine learning - classification and regression?
- ... Python?
- ... Jupyter Notebooks?
- ... scikit-learn?


Supervised learning
Unsuper-vised learning
Machine learning
Rein-forcement learning

Basic concept of supervised machine learning

Supervised learning means to generalize from given examples.

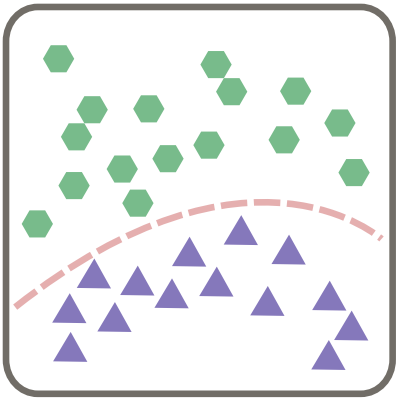
Mapping from input X (two-dimensional matrix) to output vectore y (label for classification, values for regression).

$$X_1 \rightarrow y_1$$
$$X_2 \rightarrow y_2$$
$$X_3 \rightarrow y_3$$

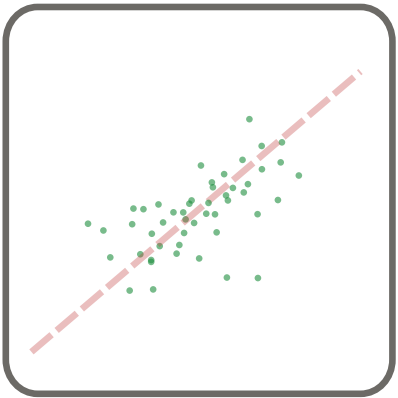
Generating a function to map input variables (X) to an output variable y.

$$y = f(X)$$

Two tasks that can be solved with supervised learning

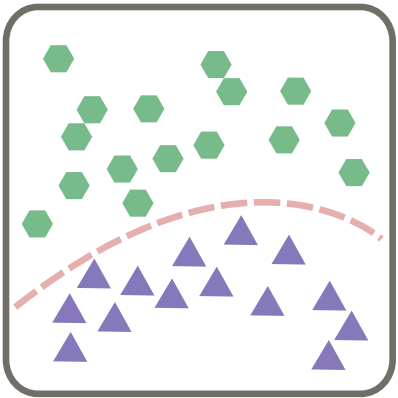


Classification

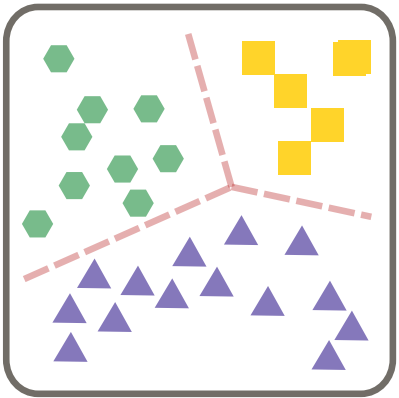


Regression

Classification types



Binary classification



Multi-class classification

Think-Pair-Share

Where in your research or your daily life have you experienced supervised machine learning?

Entities and their features

- Entities (aka. samples or data points) are described by features (aka. covariates or attributes).
- Feature can be
  - categorical
    - Nominal (e.g. eye color, gender)
    - Ordinal (e.g. very bad, bad, good, very good)
  - numerical
    - Discrete (e.g. gene length in nucleotide)
    - Continous (e.g. body length)

Features

Feature selection: Choosing the most descriptive features.

Feature encoding: Translating categorial values into numerical values.

Feature scaling: Normalize feature values (e.g. scale by minimal and maximum values)

How well does the model fit?

Overfitting: Good performance on the training data, poor generalization to other data.

Underfitting: Poor performance on the training data and poor generalization to other data

Regularization: Different methods to prevent overfitting

Overview of different methods

- Nearest Neighbor
- Naive Bayes
- Linear regression
- Logistic regression
- Decision Trees
- Artificial neural network (multilayer perceptron)
- Genetic Programming

1 Introduction

2 Foundations and terms

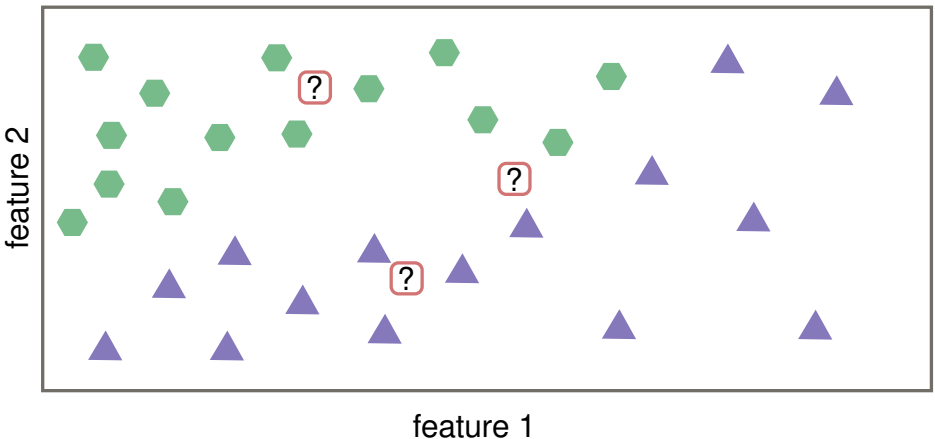
3 Selected supervised learning methods
- k-Nearest Neighbors
- Linear models
- Support Vector Machines (SVMs)
- Decision trees and Random forest
- Artificial Neural Networks

4 Reading recommendations

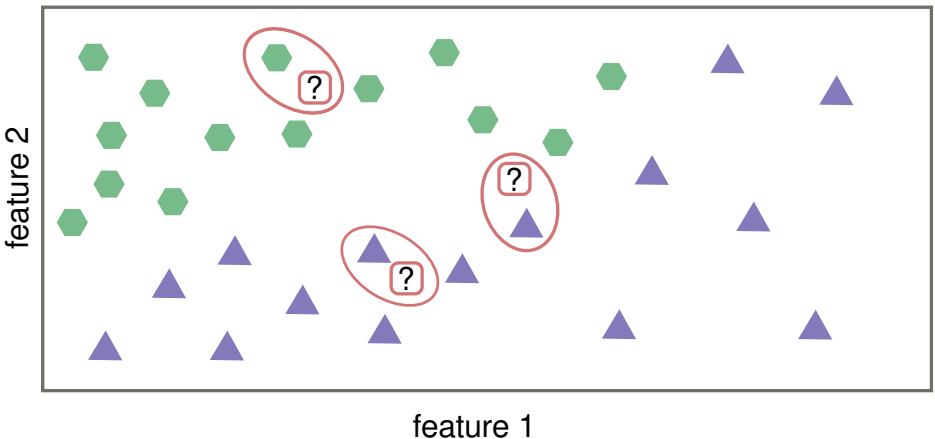
k-Nearest Neighbors

- For classification and regression
- Simplest case of supervised machine learning
- Can be easily applied to multi-class classification

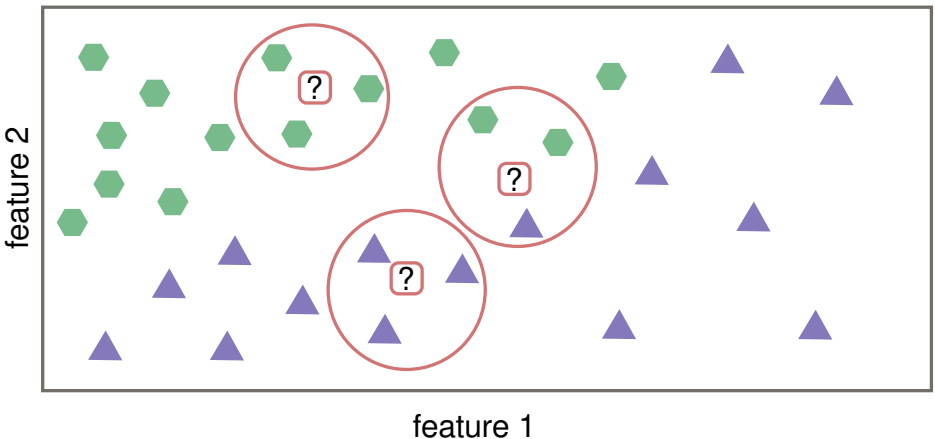
k-Nearest Neighbors



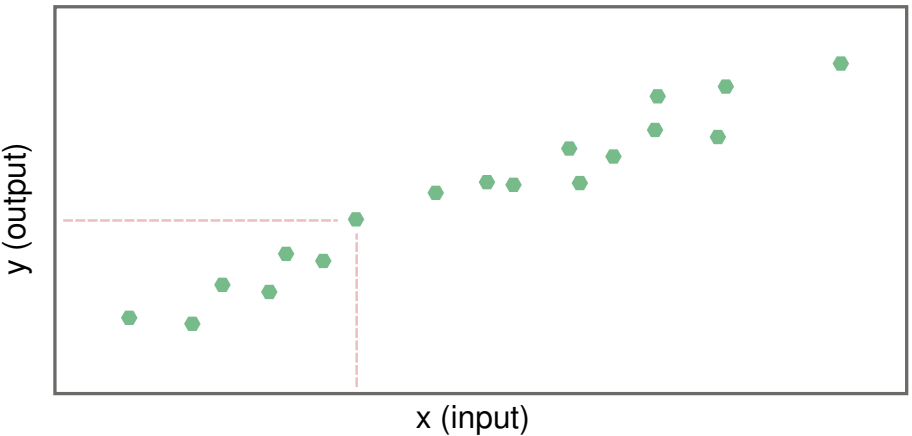
k-Nearest Neighbors



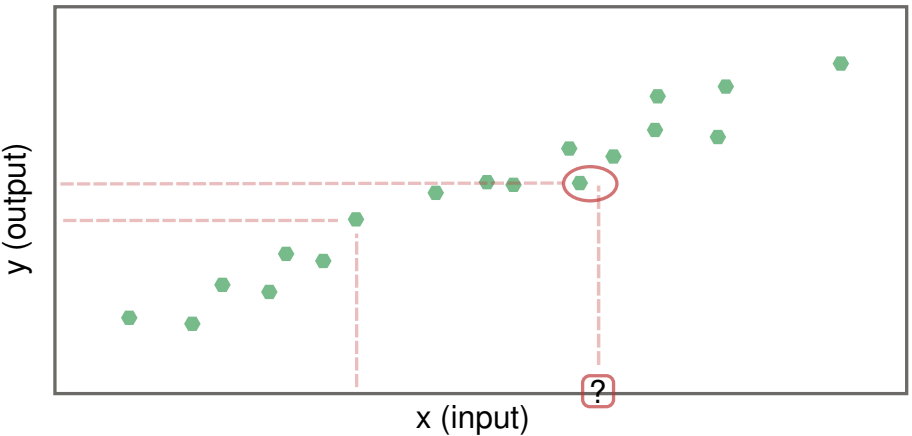
k-Nearest Neighbors



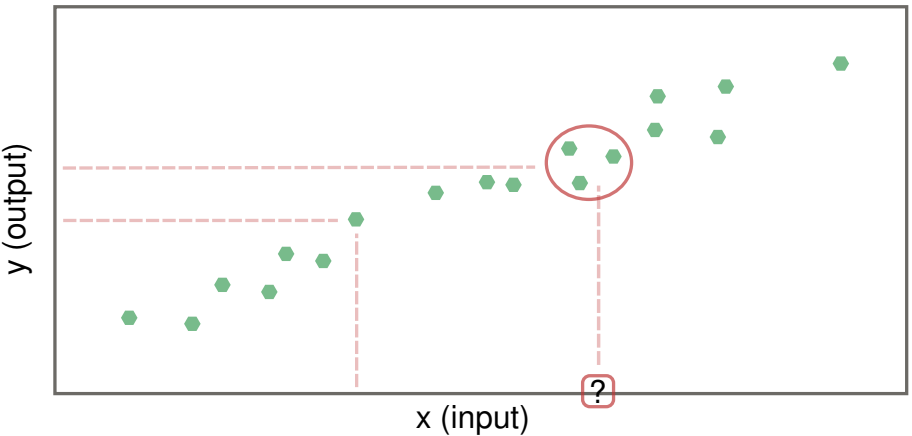
k-Nearest Neighbors



k-Nearest Neighbors



k-Nearest Neighbors



1 Introduction

2 Foundations and terms

3 Selected supervised learning methods
- k-Nearest Neighbors
- Linear models
- Support Vector Machines (SVMs)
- Decision trees and Random forest
- Artificial Neural Networks

4 Reading recommendations

Linear models

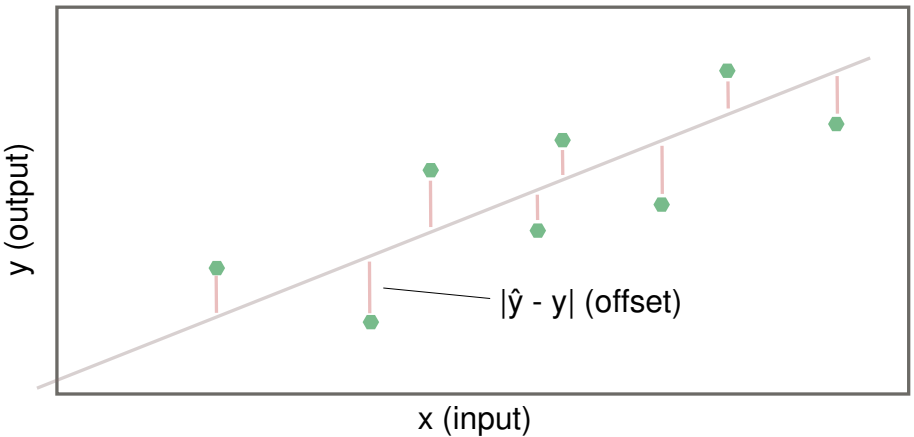
$$\hat{y} = w_1 x_1 + w_2 x_2 + w_3 x_3 + ... + w_n x_n + b$$

with n as the number of features
w are the different weights/coefficients
b the intercept

Different ways to estimate the parameters

- Ordinary Least Squares
  - no parameters - easy to use but no possibility to adapt
- Ridge
  - coefficients should be close to zero
  - more resistant against overfitting
- Least Absolute Shrinkage and Selection Operator (LASSO)

Ordinary least squares (OLS)



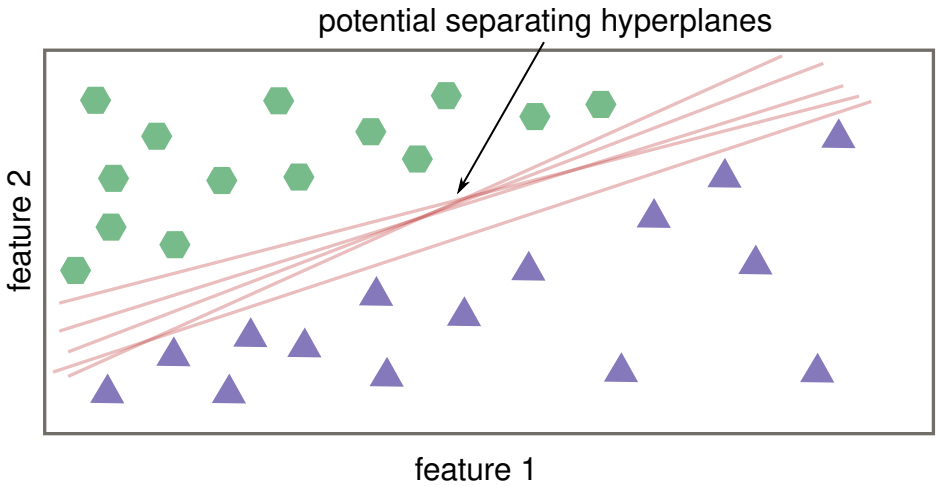
Minimize the offset between $\hat{y}$ and y the mean squared error (MSE) or sum of squared errors (SSE).

Once the parameters (b and the weights w) of

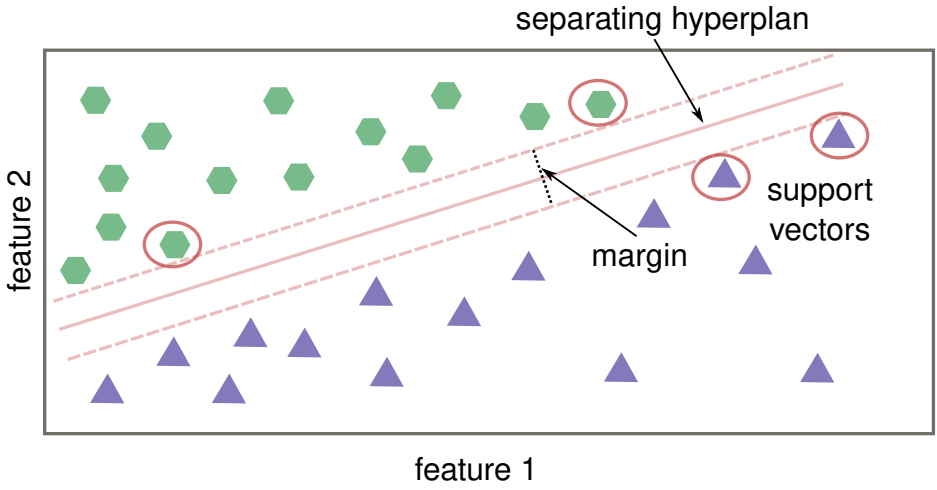
$$\hat{y} = w_1 x_1 + w_2 x_2 + w_3 x_3 + ... + w_n x_n + b$$

are estimated the prediction can be performed by putting the x values of the data points into the equation to predict the y value.

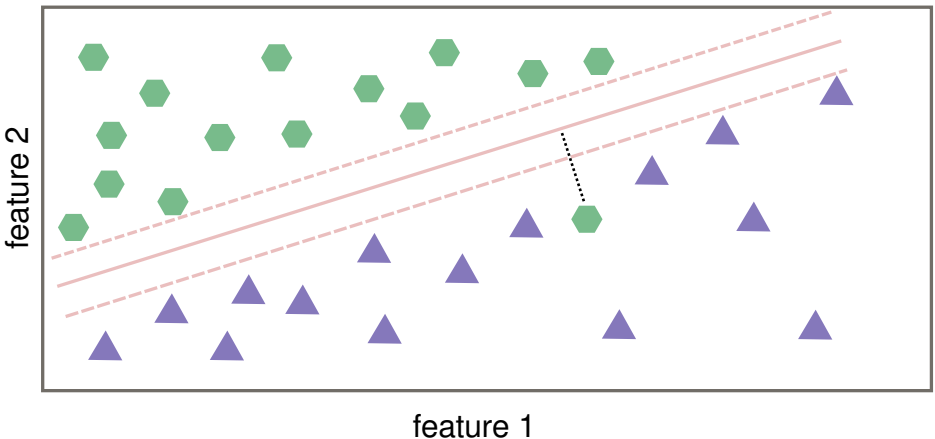
Support Vector Machines (SVMs) – Separeting hyperplan



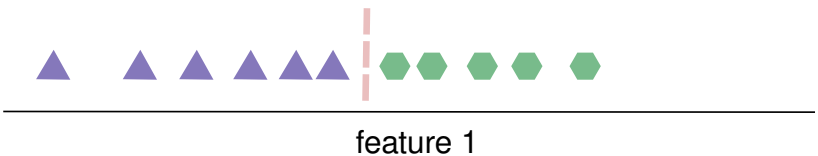
Support Vector Machines (SVMs) – Margin



Support Vector Machines (SVMs) – Soft Margin

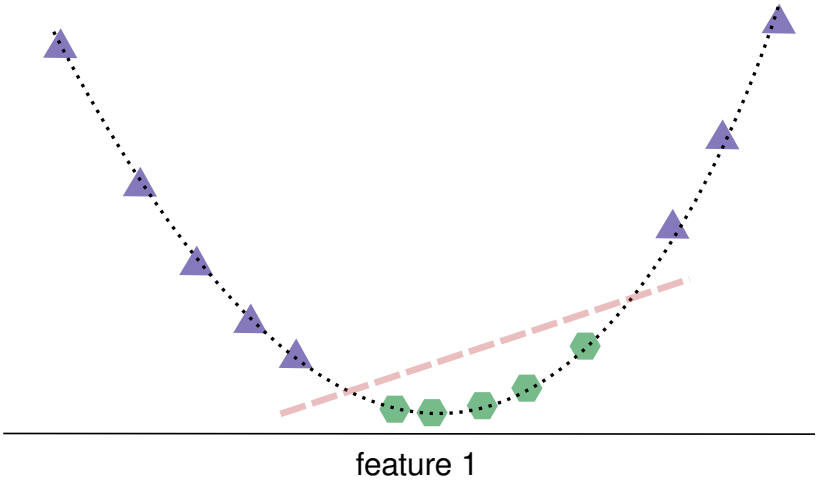


Support Vector Machines (SVMs) – Kernel trick

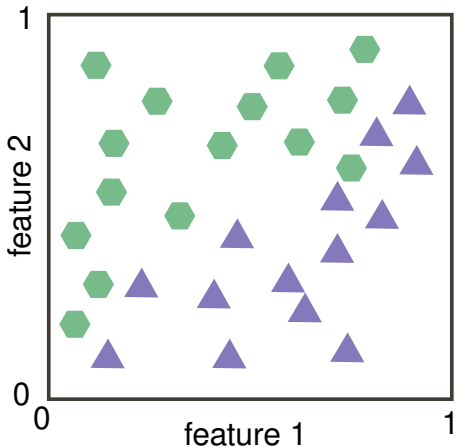


SVM – Kernel trick

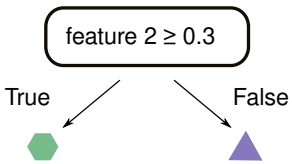
Support Vector Machines (SVMs) – Kernel trick

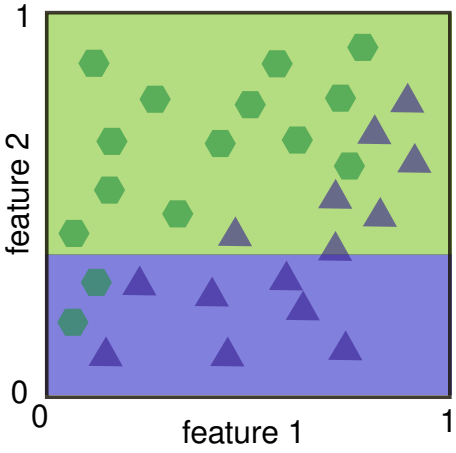


Decision Trees

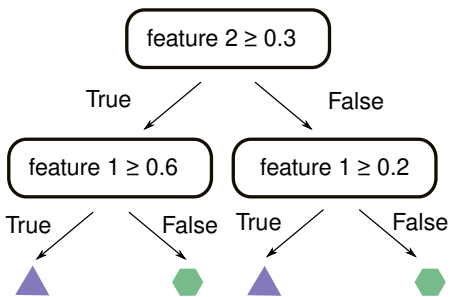


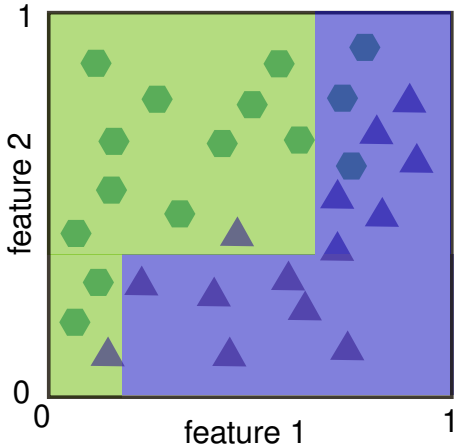
Decision Trees





Decision Trees





Random forest

- In the random forests approach, many different decision trees are grown by a randomized tree-building algorithm.
- The training set is sampled with replacement to produce a modified training set of equal size to the original but with some training items included more than once.
- In addition, when choosing the question at each node, only a small, random subset of the features is considered.
- Random decision trees from subsets of features or data point

1 Introduction

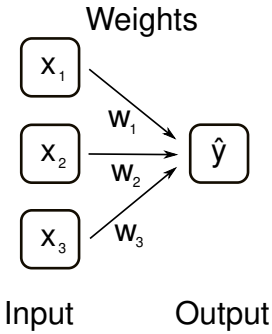
2 Foundations and terms

3 Selected supervised learning methods
- k-Nearest Neighbors
- Linear models
- Support Vector Machines (SVMs)
- Decision trees and Random forest
- Artificial Neural Networks

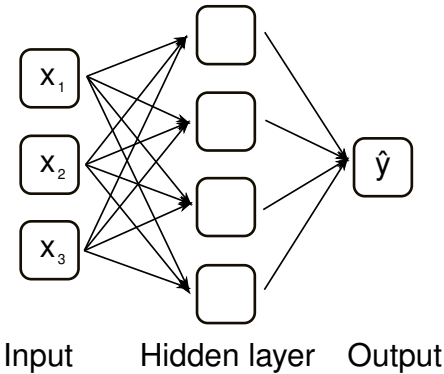
4 Reading recommendations

Artificial Neural Networks

- aka. Multilayer perceptrons or Feed-forward neural networks
- Inspired by natural Neural Networks
- For classification or regression


Weights
x1
x2
x3
w1
w2
w3
ŷ
Input
Output

Artificial Neural Networks



Artificial Neural Networks

CNN (Convolutional Neural Networks), CNN (Convolutional Neural Networks),
RNN(Recurrent Neural Networks),

... tomorrow a full session about that

Reading recommendations

- Machine Learning: A Probabilistic Perspective by Sarah Guido, Andreas Müller, 2016, ISBN-13: 978-1449369415 https://openlibrary.org/isbn/9781449369415
- Machine Learning: A Probabilistic Perspective by by Kevin P. Murphy, 2012, ISBN-13: https://openlibrary.org/isbn/9780262018029