Machine Learning - Session 1

Supervised Learning and Model evaluation

Content

1. Introduction
2. Supervised Learning
    - Classification
    - Regression
3. Model evaluation
4. Hands - On

Introduction

Some Artificial Intelligence paradigms:

- Searching and Planning (Traditional problem solving algorithms)
- Knowledge-Based (Inference and Logic)
- Machine Learning (Data-driven hypothesis)

Introduction

Some Artificial Intelligence paradigms:

- **Searching and Planning**
- **Knowledge-Based**
- **Machine Learning**
  - Learning from examples
    - **Supervised Learning**
    - **Unsupervised Learning**
    - Reinforcement Learning
    - Others (Self-supervised, Semi-supervised, etc.)
  - Deep Learning (Complex models with tons of data and computation)
  - Statistical Learning (Uncertain and evidence)
  - Knowledge in Learning (ML with prior knowledge and logic)

Introduction

Learning from examples

- **Supervised learning**
  - Classification
  - Regression

- **Unsupervised learning**
  - Association Rules
  - Cluster Analysis
  - Decomposition

Supervised Learning

Typical algorithms for classification and regression

- Linear Regression
- Logistic Regression
- NN (Neural Networks)
- SVM (Support Vector Machine)
- Decision Tree
- Random Forest and related (Boosting)

Supervised Learning

- **Others**
  - KNN (K nearest neighbours)
  - GP (Gaussian Processes)
  - NB (Naive Bayes)
  - QDA/LDA (Quadratic/Linear Discriminant Analysis)

Supervised Learning

Linear regression.

$$y(\mathbf{x}, \mathbf{w}) = \mathbf{w}^T\mathbf{x}$$

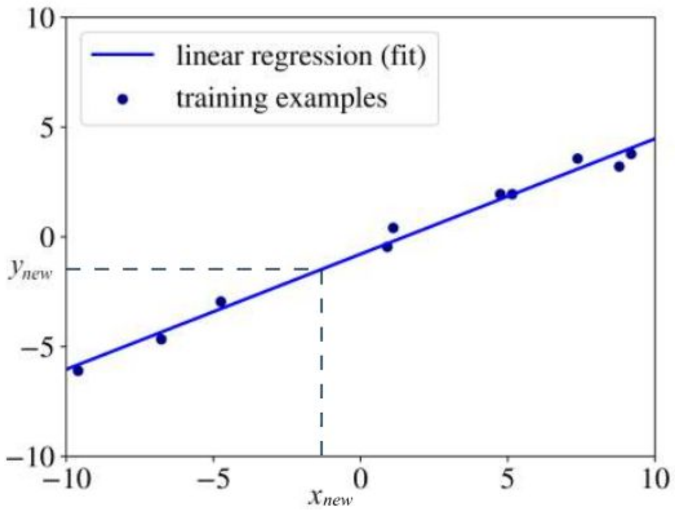
The parameters are learned minimizing the cost function:

$$_{min}|y - y_{real}|^2$$

- Gradient Descent for optimization
- Regularization techniques for modifying parameter values:
  - Lasso
  - Ridge

Supervised Learning

Linear regression.



Supervised Learning

Logistic regression. Adapts the linear regression for classification problems using the sigmoid function:

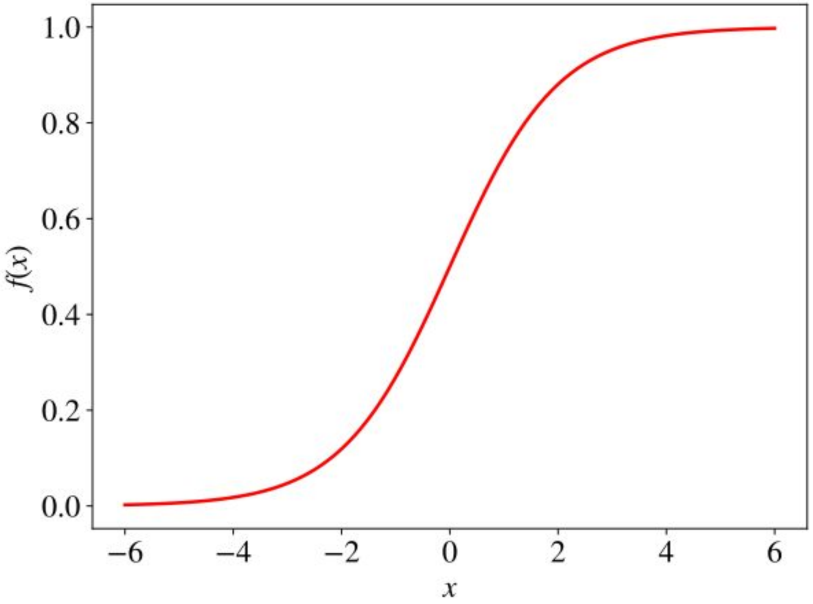
$\sigma(\mathbf{x}, \mathbf{w}) = 1 / (1 + \exp(-\mathbf{w}^T\mathbf{x}))$, thus the prediction of the target variable is:

$y(\mathbf{x}, \mathbf{w}) = \sigma(\mathbf{x}, \mathbf{w})$

- Maximizes the log-likelihood function

Supervised Learning

Logistic regression.



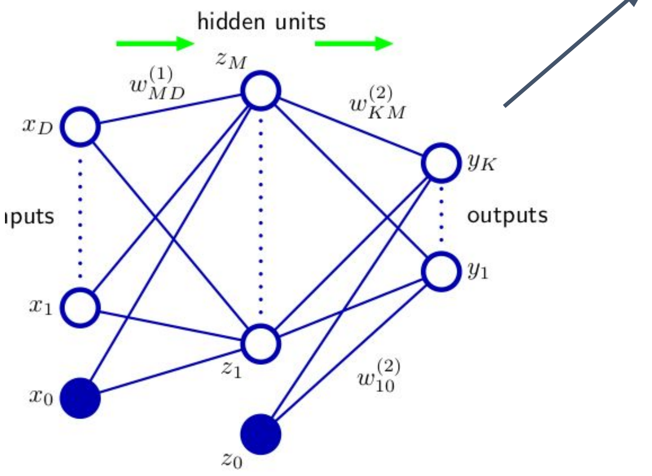
Supervised Learning

Neural Networks

- Linear combination of linear and nonlinear functions
- Applies to both regression and classification
- Backpropagation algorithm for evaluating the error function gradient
- Typical activation functions:
  - Sigmoid (Softmax for multiclass)
  - ReLU (Rectified Linear Unit) and related, PReLU, ELU, etc.
  - Tanh

Supervised Learning

Neural Networks. Brief representation of mathematical functions:

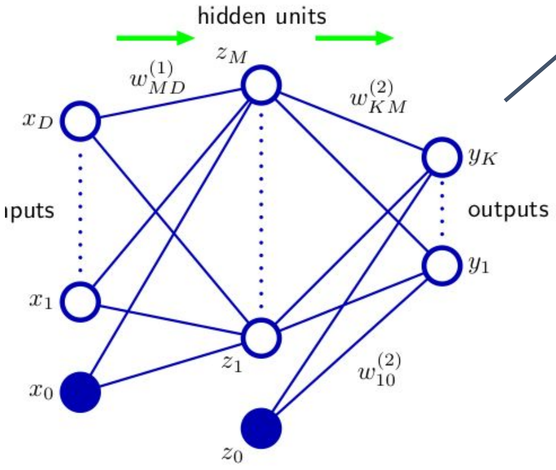


$$y_k = \sigma(a_k)$$

$$\sigma(a) = \frac{1}{1 + \exp(-a)}.$$

Supervised Learning

Neural Networks. Brief representation of mathematical functions:



$$y_k = \sigma(a_k)$$

$$\sigma(a) = \frac{1}{1+\exp(-a)}.$$

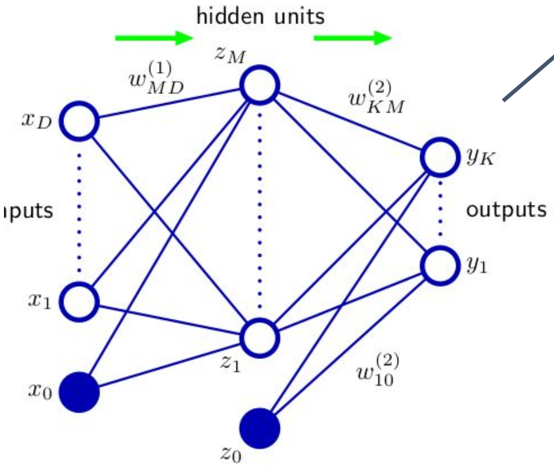
$$a_k = \sum_{j=1}^{M} w^{(2)}_{kj} z_j + w^{(2)}_{k0}$$

Bias parameters

Supervised Learning

Neural Networks. Brief representation of mathematical functions:

$$z_j = h(a_j).$$



$$y_k = \sigma(a_k)$$

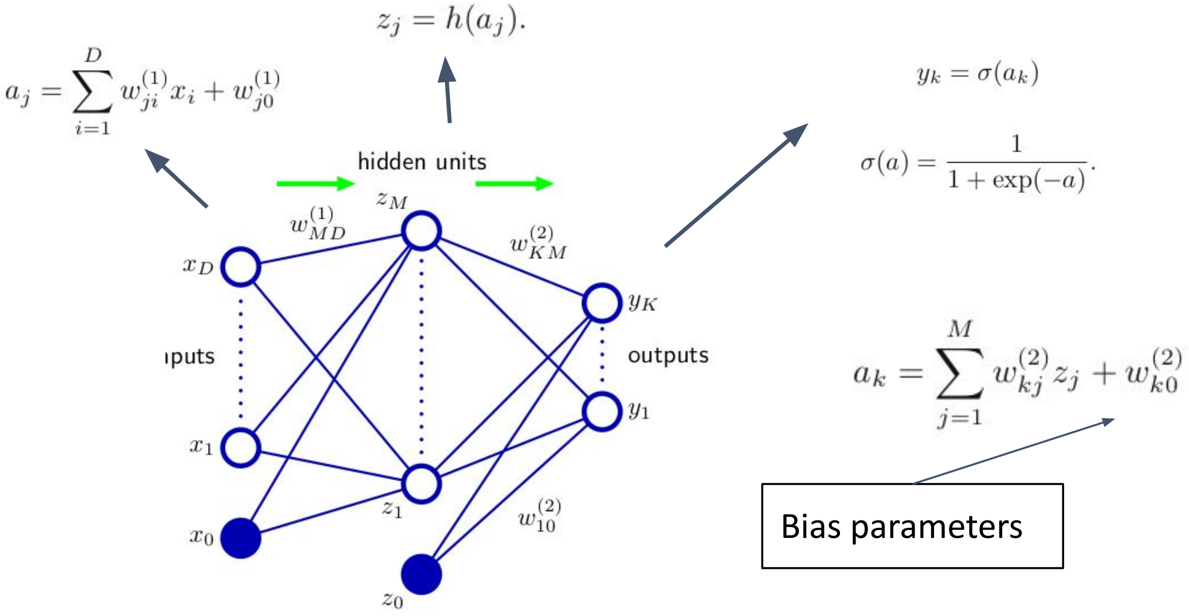
$$\sigma(a) = \frac{1}{1+\exp(-a)}.$$

$$a_k = \sum_{j=1}^{M} w^{(2)}_{kj} z_j + w^{(2)}_{k0}$$

Bias parameters

Supervised Learning

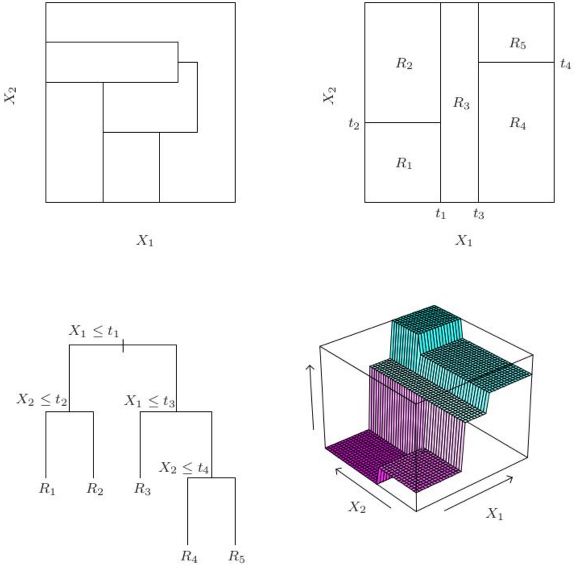
Neural Networks. Brief representation of mathematical functions:



Supervised Learning

Decision Trees

Tree-based methods partition the feature space into a set of rectangles, and then fit a simple model n each one.



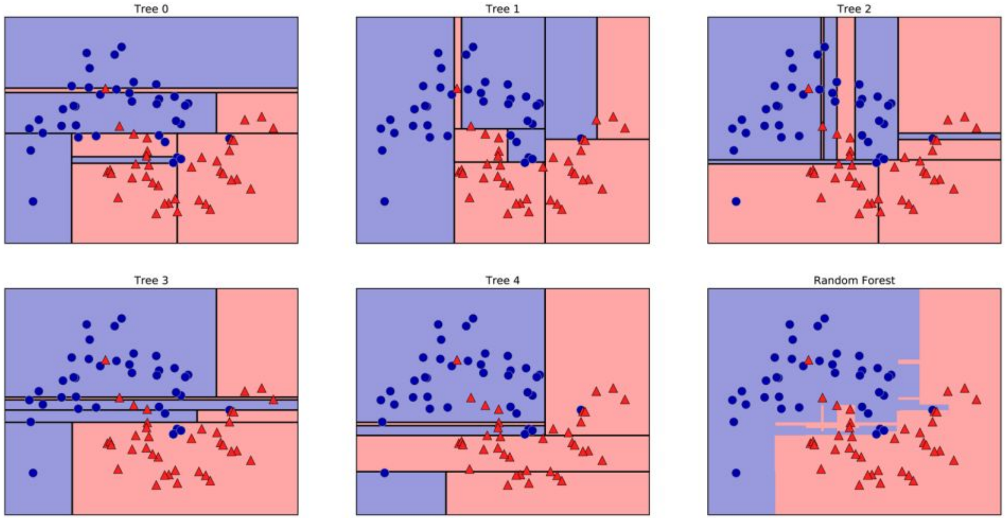
Supervised Learning

Random Forest

- Ensemble of models
- Bagging (Bootstrap aggregating)
- Decision Trees as base estimators
- Reduce Variance
- Easy explainable
- Robust to outliers and non-scaled data

Supervised Learning

Random Forest



Supervised Learning

AdaBoost and Gradient Boosting Machines (GBM)

Boosting is a technique where the model fitting is made sequentially modifying the weights of erroneously predicted samples and establishing a committee of predictors

- AdaBoost reduces bias
- AdaBoost uses an exponential loss function
- GBM generalize the loss function

Supervised Learning

XGBoost

- Extension of GBM
- Highly scalable
- Sophisticated regularization techniques
- Variants LightGBM, CatBoost

Model selection and evaluation

For model selection one must choose the best hyperparameters, for example: number of trees, hidden units, regularization coefficients, etc.

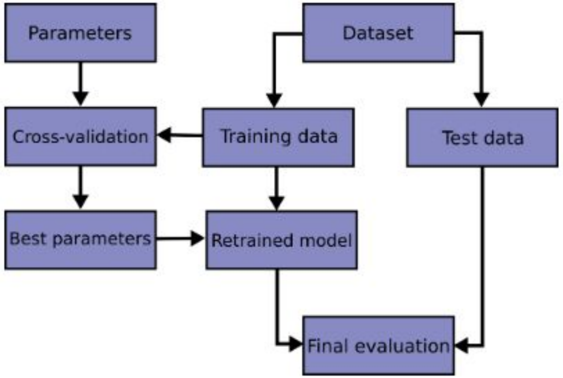
Several search algorithms are employed for hyperparameter search:

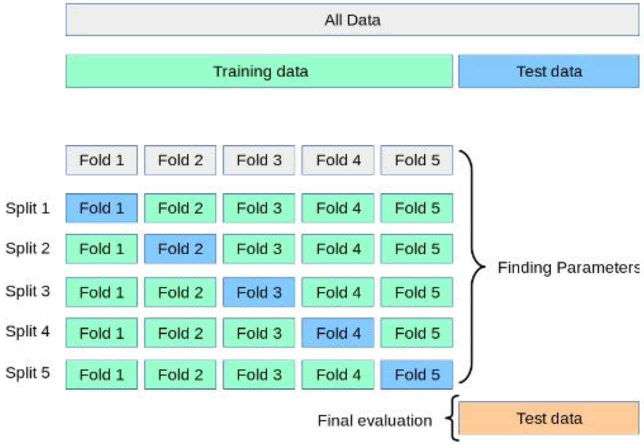
- Grid Search
- Random Search
- Bayes Optimization

It can be considered an optimization problem in a n-dimensional search space.

Model selection and evaluation

In order to select the best hyperparameters the model is subjected to Cross-Validation. The training data set is divided in k folds where each fold is used to test the training in the remaining data. Then the model is evaluated in a separate test (or validation) set:

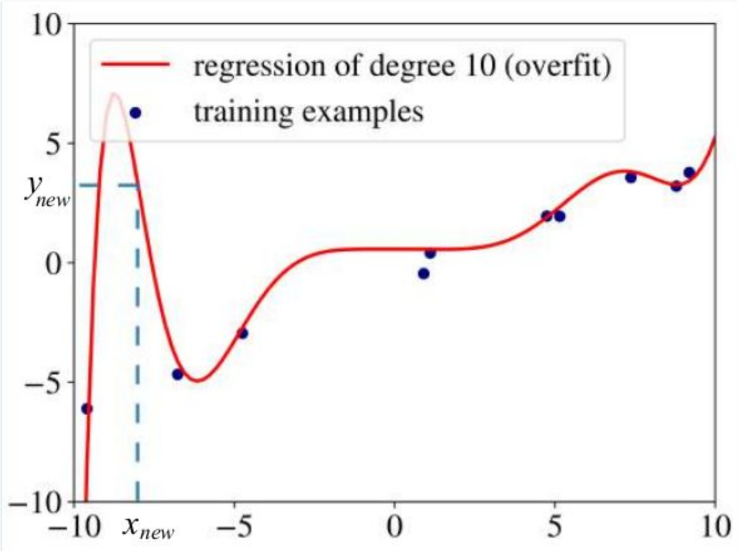


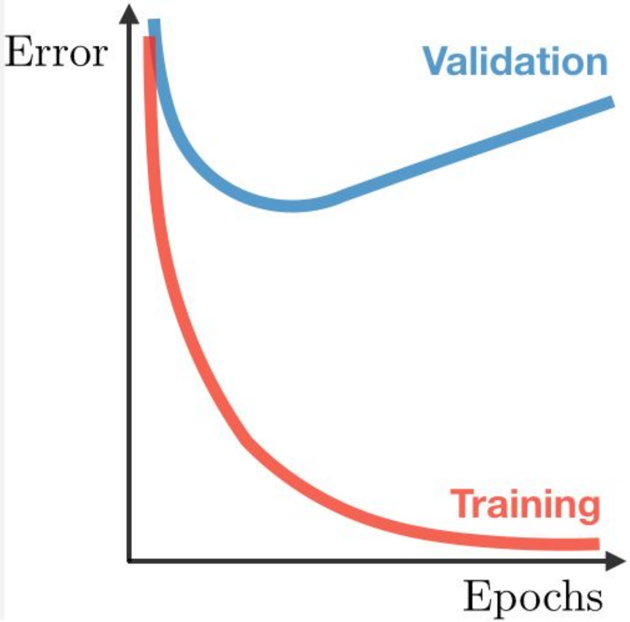


The objective is to avoid overfitting, the problem when the model performs well in a train set but performs inefficiently in the test/validation set

Model selection and evaluation

Overfitting example in a polynomial regression (left) and NN (right)





Model selection and evaluation

Performance metrics. When the model is selected, metrics are employed for validating the model in a real world scenario. For example a regressor can be evaluated with:

- RMSE (Root Mean Squared Error) (Interpretable units)
- MAPE (Mean Absolute Percentage Error) (Percentage error)

Model selection and evaluation

Performance metrics. In the classification case, the real cost of predicting wrong is taken into account. Thus, most classification metrics are functions of False Positives, False Negatives, True Positives and true Negatives. Typical metrics are:

- Precision
- Recall
- F1-score
- ROC (Area under the Receiver Operating Characteristic)

Hands-On