Machine Learning Review

Deep Learning

Stijn Lievens & Sabine Devreese

2023-2024

Machine Learning Review

What is Machine Learning?

Machine Learning Definition(s)

Machine learning is the science (and art) of programming computers so they can learn from data.

Machine Learning Definition(s)

Machine learning is the field of study that gives computers the ability to learn without being explicitly programmed. (Arthur Samuel, 1959)

Types of Machine Learning

- Supervised learning
- Unsupervised learning
- Reinforcement learning

Supervised Learning

- The training data you feed to the algorithm includes the desired solutions, called *labels*.
- The goal is to learn to predict the labels from the features.
- Examples:
  - Classification
  - Regression

Unsupervised Learning

- The training data is unlabeled.
- The goal is to learn the underlying structure of the data.
- Example:
  - Clustering

Reinforcement learning

- The learning system, called an *agent*, can observe the environment, select and perform actions, and get *rewards* or *penalties* in return.
- The goal is to learn the best strategy, called a *policy*, to get the most reward over time.

Example of ML-Algorithms

Linear Regression

Linear regression is a supervised learning algorithm that can be used to predict a *continuous* value.

$$\hat{y} = b + w_1 x_1 + w_2 x_2 + \cdots + w_n x_n$$

Here,

- $\hat{y}$ is the predicted value,
- b is the bias,
- $w_1, \ldots, w_n$ are the weights and
- $x_1, \ldots, x_n$ are the features.

Goal of Linear Regression

Find the values of b and $w_1, \ldots, w_n$ that minimize the *mean squared error* (MSE) on the training set.

$$\mathrm{MSE} = \frac{1}{m} \sum_{i=1}^{m} (\hat{y}^{(i)} - y^{(i)})^2$$

Solution

The MSE can be minimized analytically by solving the *normal equation*:

$$(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

Here, $\mathbf{X}$ is the feature matrix with one row per example and one column per feature, $\mathbf{y}$ is the vector of target values.
Note: in this case $\mathbf{X}$ will have a column of ones for the bias.

Logistic Regression

Logistic regression is a supervised learning algorithm that is used for binary classification.
The probability of belonging to the positive class is modeled as:

$$\hat{y} = \sigma(b + w_1 x_1 + w_2 x_2 + \cdots + w_n x_n)$$

with

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

Finding the Parameters

The parameters b and $w_1, \ldots, w_n$ are found by minimizing the *binary cross entropy loss* on the training set:

$$\mathrm{BCE} = -\frac{1}{m} \sum_{i=1}^{m} \Big(y^{(i)} \log(\hat{y}^{(i)}) + (1 - y^{(i)}) \log(1 - \hat{y}^{(i)}) \Big)$$

where $y^{(i)}$ is the true label of the i-th example, while $\hat{y}^{(i)}$ is the predicted probability of the i-th example belonging to the positive class.

No Closed Form Solution

- There is no closed form solution for the (optimal) parameters of logistic regression, i.e. the parameters that minimize the binary cross entropy loss.
- They are found by using **gradient descent**.
  - See courses “Mathematics for Machine Learning” and “Machine Learning” for more details.

Machine vs. Deep Learning

Machine Learning

- Typically uses structured data, i.e. tabular data.
- The algorithms themselves are more or less “fixed” with a limited number of hyperparameters.
  - E.g. linear regression, logistic regression, decision trees, random forest, SVMs,
- Algorithms are typically not very data hungry.

Deep Learning

- Typically uses unstructured data, e.g. images, text, audio, video, ...
- Consists of many different components that can be **combined** in many different ways.
- Typically requires a lot of data.