TEST FOR DEV: Introduction to Data Science

WS 23/24, RWTH Aachen

November 29, 2023

1 Regression

In the next three chapters, we'll look at **error-based learning**. In general, this learning approach functions as follows: Error-based learning

- We have a **parameterized prediction model** which is initialized with random parameters.
- An **error function** is then used to evaluate the performance of this model when it makes predictions for instances in a training dataset. Error function
- Based on the results of the error function, the parameters are **iteratively adjusted** to create a more and more accurate model.

There are different approaches to realizing error-based learning:

- Regression (covered in this section)
- SVMs (covered in the next section)
- Neural networks (covered in a later section)
- Genetic algorithms, or other evolutionary approaches

We'll start with **regression** and the following basic idea. Regression

$$\text{Our model: } f : \text{descriptive features} \rightarrow \text{target features}$$
$$\text{Goal: find } f \text{ minimizing } error(\text{prediction}, \text{observed data})$$

When we compare this approach to decision trees, we see:

- Decision trees were initially developed for categorical features and then extended to continuous features.
- Regression followed the reverse path, which means it's most **suitable for continuous data**.
- Still, both are supervised learning techniques.

1.1 Simple linear regression

Consider the following simple example where we have:

- Rental price p_r as our target feature, and
- Size s as our descriptive (continuous) feature

We assume a linear dependency $p_r = b + a \cdot s$ and now want to base our prediction of the rental prize on the size. The example will guide us through this subchapter.

General problem regression

The **general problem** is given as follows:

- We have given n data rows in a set $\mathcal{D}$ with a target feature t and descriptive features $\mathbf{d} = (d[1], d[2], \cdots, d[m])$, and
- We want to find a regression function $\mathbb{M}_{\mathbf{w}}$ with a constant weight and a weight for each feature, where
- We predict $\text{pred}(t) = \mathbb{M}_{\mathbf{w}}(\mathbf{d}) = \mathbb{M}_{(w[0],w[1],\cdots,w[m])}(d[1], d[2], \cdots, d[m])$

In our example, we only have one descriptive feature $\mathbf{d} = (d[1])$, two weights $\mathbf{w} = (w[0], w[1])$, and the regression function is linear, so $\underbrace{\mathbb{M}_{\mathbf{w}}(d)}_{p_r} = \underbrace{w[0]}_{b} + \underbrace{w[1]}_{a} \cdot \underbrace{d[1]}_{s}$.

As one can see

1.2 Multiple descriptive features

1.3 Interpretation of results

1.4 Hanlding categorical features

1.5 Logistic regression

1.6 Extensions (non-linear and multinomial)