SML project

Linus Falk

November 23, 2022

1 Logistic regression

The Logistic regression is constructed by modifying the linear regression model so it can be used for classification problem. We start with linear regression and build from that:

$$z = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_p x_p \tag{1}$$

To fit our model to a prediction of probability $\mathrm{p}(y = 1|x)$ we use the logistic function also know as the Sigmoid function, defined as:

$$f(\mathrm{z}) = \frac{e^z}{1 + e^z} \in \begin{bmatrix} 0 & 1 \end{bmatrix} \tag{2}$$

This implies in the classification problem with 2 classes that we have the probability or the model for the second class $\mathrm{p}(y = -1|x) : 1 - f(\mathrm{z})$. Choosing 1 and -1 as labels simplify our expressions such that we are left with:

$$f(\mathrm{z}) = \frac{e^z}{1 + e^z} = 1 - f(\mathrm{z}) = \frac{e^{\theta^T x}}{1 + e^{\theta^T x}} \tag{3}$$

We want to find the parameters θ with the use of our training data. With the use the maximum likelihood approach we have:

$$\hat{\theta} = \arg\max \mathrm{p}(\mathbf{y} \mid \mathbf{X};\theta) = \arg\max \sum_{i=1}^{n} \ln \mathrm{p}(\mathrm{y}_i|\mathbf{x}_i;\theta) \tag{4}$$

turning this into minimization problem by using the negative log likelihood as cost function and since we have chosen our labels in a clever way we end up with the cost function:

$$J(\theta) = \frac{1}{n} \sum_{i=1}^{n} \ln(1 + e^{-y_i \theta^T x_i}) \tag{5}$$

This modification is not "perfect" since we must use numerical methods for finding the parameters since there is no closed expression for the cost.