Chapter 01 – Supervised Learning

Author: Gianmarco Scarano

gianmarcoscarano@gmail.com

1. Supervised Learning

Supervised Learning is a Machine Learning technique in which we have labels and we want to train some kind of data (Tabular / Text / Images, which are mainly referred as Tensors) on generalizing on these labels. In other words, we aim at finding return the appropriate output y.

1.1 Regression vs. Classification

In Regression, what we would like to do is to fit a line between multiple points, while in classification, we want to classify (for example) two classes: 0 and 1.

Most of the times Regression is used for predicting a continuous number/variable, while in Classification, our model outputs classes, not real numbers.

1.2 Model Parameters

In order to fit this data in a proper manner, we have to find a function which approximates best our values. This function is a sum of various parameters (nonetheless the parameters of the model) which can be defined as follows:

$$f(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_n x_n$$

Where the vector $\theta = [\theta_0, \theta_1, \theta_2, \ldots, \theta_n]^T$ are the $(n+1)$ parameters of our model, while x_n are the samples given in input to our model.

In the end, we can rewrite our output in the following way -> $y = f(x;\ \theta)$.

1.3 Loss functions

Measuring the performances of the parameters can be quite difficult, that's why Loss functions come handy.

A loss function $L(\theta)$ is a function of the parameters and for each sample we compute the following measurement: $L(\theta) = \mathcal{L}(f(x_i;\ \theta), y_i)$ which substantially measures the loss on that specific x_i sample.

We have different kinds of losses, such as MSE (Mean Squared Error – Sum of the squares of these deviations for all training pairs) for regression problems or Binary Cross-Entropy (BCE) for classification problems. BCE is computed as follows:

$$L[\phi] = \sum_{i=1}^{I} -(1-y_i)\log\left[1 - \boxed{\mathrm{sig}}[\mathrm{f}[\mathbf{x}_i, \phi]]\right] - y_i \log\left[\boxed{\mathrm{sig}}[\mathrm{f}[\mathbf{x}_i, \phi]]\right]$$

Sigmoid Function $\dfrac{1}{1+e^{-\mathbf{h}}}$