

# Machine Learning - Assignment 1

Francesco Saverio Zuppichini

October 2, 2017

## 1 The Perceptron

### 1.1 Vectorized equations

The vectorized equation for a single perceptron

$$output = \sum_{i=1}^N XW + b \quad (1)$$

Where  $X = x_1, \dots, x_n$ ,  $W = w_1, \dots, w_n$ . We denote  $y$  as the output of the perceptron

### 1.2 Mean Squared Error

The mean squared error function for our single perceptron

$$E(w) = \frac{1}{N} \sum_{i=1}^N (\underbrace{y(x_i, w_i)}_{\text{predicted}} - \underbrace{t_i}_{\text{actual}})^2 \quad (2)$$

One trick that is usual done is to multiply equation (2) by  $\frac{1}{2}$ , so when we take the derivative the 2 goes away. This is called One Half Mean Squared Error.

### 1.3 Derivate of the error with respect to the weights

In order to reduce our error function and adjust each weight we need to compute the first derivati with respect to the weights. We will use the modified version of equation (2) called One Half Mean Squared Error.

$$\frac{\delta E}{\delta w_i} = \frac{1}{N} \sum_{i=1}^N (y(x_i, w_i) - t_i)x_i \quad (3)$$

## 1.4 Gradient Descend

The gradient descend is an iterative optimisation algorithm that follow the direction of the negative descent in order to find the minimum of an objective function. It is can be used as Learning Algorithm since it allows to reduce our error function, equation (2), and adjust the weights properly. It's equation

$$w_{k+1} = w_k - \eta \nabla E(w_k) \quad (4)$$

Where  $\eta$  is the step size, also called **learning rate** in Machine Learning. This parameter influence the behavior of Gradient Descent, a small number can lead to local minimum, while a bigger learning rate could "over-shoot" and decreaasing the converge ration.

For this reasons, numerous improvements have been proposed to avoid local minima and increase its convergence ration. Some of them are: Conjugate Gradient and Momentum.