LVC 1: Glossary of Notations

x = Input vector for the Neural Network

 $x_i = i^{th}$ component of the input vector **x**

 \mathbf{w} = The vector containing the weights for all the components of the input vector \mathbf{x}

 $w_i = i^{th}$ component of the weight vector **w**

 \mathbf{w}^{T} = Transpose of the weight vector \mathbf{w}

d = Total number of input features

b =The bias Term

 $\sum_{j=1}^{d} w_{j} x_{j} = \text{Weighted sum of all input features starting from } j = 1 \text{ to } j = d, \text{ where } j \text{ is the iterator}$

z = Output of each neuron, i.e., the sum of "weighted sum of inputs" and "the bias"

f(z) = Activation function applied on the vector z

 θ = The weight array **w** concatenated with the bias term

 $F(\mathbf{x}; \Theta)$ = The output of the neural network when the input given is \mathbf{x} and the weights are given by the vector Θ

Sigmoid = This is an activation function which takes an input and gives an output lying withing the range of 0 to 1

tanh = This is an activation function which takes an input and gives an output lying withing the range of -1 to 1

ReLu = This activation function also called Rectified Linear Unit takes an input and returns it as it is, if it's positive, and returns zero if the input is negative

 a_i = Softmax Activation Function

 $L(data; \theta)$ = Loss term, i.e., the difference between the output of the Neural Network and the actual target variable, when the Neural Networks received the particular data and had the particular set of weights, i.e., θ

 $\Delta\omega_{ij}$ = Small change in weights ω_{ij}

 $\frac{\partial L}{\partial \omega_{ij}}$ = Partial derivative of the loss with respect to the particular weight given by ω_{ij}

 y^{i} = The actual value of the target variable for the i^{th} data point

 $F(\mathbf{x}^{i}; \theta)$ = The predicted output for the i^{th} data point

 ω_{ij}^{t} = The value of weight ω_{ij} at time = t

 ω_{ij}^{t+1} = The value of weight ω_{ij} at time = t + 1

 η = Learning rate for updating the weights

 λ = Regularization coefficient

 $\frac{\lambda}{2} ||\theta||^2$ = Regularization term for L2 regularization