Standard notations for Deep Learning

This document has the purpose of discussing a new standard for deep learning mathematical notations.

1 Neural Networks Notations.

General comments:

· superscript (i) will denote the $i^{th}$ training example while superscript $[l]$ will denote the $l^{th}$ layer

Sizes:

· $m$ : number of examples in the dataset
· $n_x$ : input size
· $n_y$ : output size (or number of classes)
· $n_h^l$ : number of hidden units of the $l^{th}$ layer

In a for loop, it is possible to denote $n_x = n_h^0$ and $n_y = n_h^{[\text{number of layers}+1]}$.

$L$ : number of layers in the network.

Objects:

· $X \in \mathbb{R}^{n_x \times m}$ is the input matrix
· $x^{(i)} \in \mathbb{R}^{n_x}$ is the $i^{th}$ example represented as a column vector
· $Y \in \mathbb{R}^{n_y \times m}$ is the label matrix
· $y^{(i)} \in \mathbb{R}^{n_y}$ is the output label for the $i^{th}$ example
· $W^{[l]} \in \mathbb{R}^{\text{number of units in next layer} \times \text{number of units in the previous layer}}$ is the weight matrix, superscript $[l]$ indicates the layer
· $b^{[l]} \in \mathbb{R}^{\text{number of units in next layer}}$ is the bias vector in the $l^{th}$ layer
· $\hat{y} \in \mathbb{R}^{n_y}$ is the predicted output vector. It can also be denoted $a^{[L]}$ where $L$ is the number of layers in the network.

Common forward propagation equation examples:

· $a = g^{[l]}(W_x x^{(i)} + b_1) = g^{[l]}(z_1)$ where $g^{[l]}$ denotes the $l^{th}$ layer activation function
· $\hat{y}^{(i)} = \text{softmax}(W_h h + b_2)$
· General Activation Formula: $a_j^{[l]} = g^{[l]}(\sum_k w_{jk} a_k^{[l-1]} + b_j^{[l]}) = g^{[l]}(z_j^{[l]})$
· $J(x, W, b, y)$ or $J(\hat{y}, y)$ denote the cost function.

Examples of cost function:

· $J_{CE}(\hat{y}, y) = -\sum_{i=0}^{m} y^{(i)} \log \hat{y}^{(i)}$
· $J_1(\hat{y}, y) = \sum_{i=0}^{m} | y^{(i)} - \hat{y}^{(i)} |$
2 Deep Learning representations

For representations:

- nodes represent inputs, activations or outputs
- edges represent weights or biases

Here are several examples of Standard deep learning representations

Figure 1: Comprehensive Network: representation commonly used for Neural Networks. For better aesthetic, we omitted the details on the parameters \( w_{ij} \) and \( b_i \) etc... that should appear on the edges

Figure 2: Simplified Network: a simpler representation of a two layer neural network, both are equivalent.