Machine Learning - Theoretical exercise 5

Téo Bouvard

March 5, 2020

Problem 1

Let φ be a linear function used as the activation function for the two-layer network. As φ is linear, we have $\varphi(u) = ku$, with $k \in \mathbb{R}$. We first compute the outputs at the first layer.

$$\begin{aligned} \boldsymbol{y}^{(1)} &= \varphi(\boldsymbol{\Theta}^{(1)}\boldsymbol{x}) \\ &= k\boldsymbol{\Theta}^{(1)}\boldsymbol{x} \end{aligned}$$

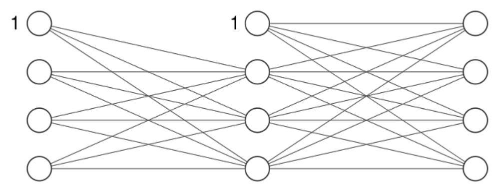
We can now use this result for computing the outputs at the second layer.

$$\begin{aligned} \boldsymbol{y}^{(2)} &= \varphi(\boldsymbol{\Theta}^{(2)}\boldsymbol{y}^{(1)}) \\ &= \varphi(\boldsymbol{\Theta}^{(2)}k\boldsymbol{\Theta}^{(1)}\boldsymbol{x}) \\ &= k^2\boldsymbol{\Theta}^{(2)}\boldsymbol{\Theta}^{(1)}\boldsymbol{x} \end{aligned}$$

We see that this result is equivalent to a single layer network having $k^2\boldsymbol{\Theta}^{(2)}\boldsymbol{\Theta}^{(1)}$ as its weight matrix and constant activation function.

Problem 2

a) The network has the following structure.



b) Normalizing the training vectors gives us the following normalized dataset.

$$x_1 = \begin{bmatrix} 1 \\ \frac{1}{4} \\ \frac{1}{4} \end{bmatrix} x_2 = \begin{bmatrix} 1 \\ \frac{1}{4} \\ 0 \end{bmatrix} x_3 = \begin{bmatrix} \frac{1}{2} \\ 0 \\ \frac{1}{4} \end{bmatrix} x_4 = \begin{bmatrix} \frac{1}{2} \\ 0 \\ 0 \end{bmatrix}$$

c) Let σ be the sigmoid function used as activation function for the network. We could compute the consecutive outputs by performing a matrix multiplication and then a matrix addition.

$$\boldsymbol{y} = \sigma(\boldsymbol{\Theta}\boldsymbol{x} + b)$$

but it is easier to incorporate the bias to our weight matrix, and add a unit component to each of our training vector, as this allows us to perform a single matrix multiplication. In the following, we this augmented weight matrix will be denoted $\boldsymbol{\Theta}$.

We first compute the output at the first hidden layer.

$$\boldsymbol{y}^{(1)} = \sigma(\boldsymbol{\Theta}^{(1)}\boldsymbol{x})$$