Logistic Regression and Back-Propagation

Exercise

1. Recall the sigmoid function, $\sigma(z) = \frac{1}{1+e^{-z}}$. The logistic regression classifier, that we've learned in class, is a binary classifier. The estimated probability $\hat{p}(x^{(i)};\theta)$ is defined as

$$\hat{p}(y^{(i)} = 1|x^{(i)};\theta, b) = h_{\theta,b}(x^{(i)}) = f(\theta^T x^{(i)} + b)$$

where $f(z)$ is usually the sigmoid function $\sigma(z) = \frac{1}{1+e^{-z}}$.

   (a) Assume that $\{(x^{(i)}, y^{(i)})\}_{i=1}^m$ is a set if i.i.d samples. Write the estimated probability of the entire set. i.e., write $p(y^{(1)}, .., .\hat{y^{(m)}}|x^{(1)}, .., x^{(m)})$ in terms of $\{(x^{(i)}, y^{(i)})\}_{i=1}^m$ and $h_{\theta,b}(\cdot)$.

   (b) Is the sigmoid function convex, concave oe none? Prove your claim.

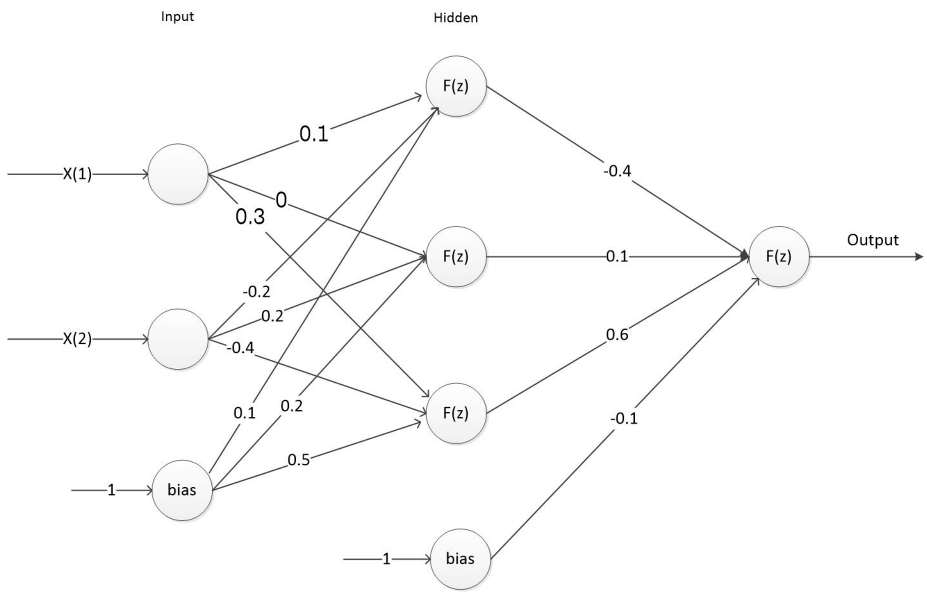
   (c) Assume that the sigmoid function is replaced with the following piecewise linear function

$$f(z) = \begin{cases} 0, & \text{if } z < -0.5 \\ 0.5 + z, & \text{if } -0.5 \le z \le 0.5 \\ 1, & \text{if } z > 0.5 \end{cases} \tag{1}$$

   Let $x = (x_1, x_2)$ be a binary vector, namely, $x_1 \in \{0, 1\}, x_2 \in \{0, 1\}$. can you find θ_1, θ_2 and b such that $f(\theta^T x+b)$ is the logical or between x_1 and x_2? If yes, do it. If no, prove it doesn't exist.

   (d) Can you find θ, b such that $f(\theta^T x + b)$ is the logical exclusive or (XOR) between x_1, x_2? If yes, do it. If no, prove it doesn't exist.

2. Let F be the Sigmoid function and the cost function be MSE. Let $X = [0.6, 0.1]^T$ with label 1. Do the forward propagation of X, then perform back-propagation of the error.



(a) What will be the weights after a single step with learning rate of 0.1?

(b) Do the same as above, but with cross-entropy cost instead of MSE.