Assignment 5

Introduction to Machine Learning
Prof. B. Ravindran

1. You are given the following neural networks which take two binary valued inputs $x_1, x_2 \in \{0, 1\}$ and the activation function is the threshold function($h(x) = 1$ if $x > 0$; 0 otherwise). Which of the following logical functions does it compute?

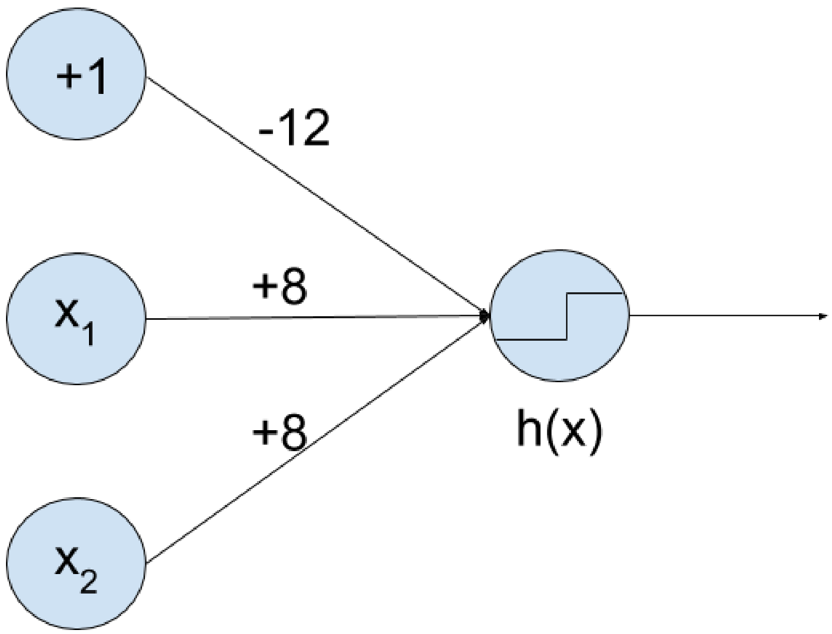

Figure 1: Q1

   (a) OR
   (b) AND
   (c) NAND
   (d) None of the above.

2. We have a function which takes a two-dimensional input $x = (x_1, x_2)$ and has two parameters $w = (w_1, w_2)$ given by $f(x, w) = \sigma(\sigma(x_1 w_1) w_2 + x_2)$ where $\sigma(x) = \frac{1}{1+e^{-x}}$. We use backpropagation to estimate the right parameter values. We start by setting both the parameters to 0. Assume that we are given a training point $x_1 = 1, x_2 = 0, y = 5$. Given this information answer the next two questions. What is the value of $\frac{\partial f}{\partial w_2}$?

   (a) 0.5
   (b) -0.25
   (c) 0.125

(d) -0.5

3. If the learning rate is 0.5, what will be the value of w_2 after one update using backpropagation algorithm?

   (a) 0.0625
   (b) -0.0625
   (c) 0.5625
   (d) - 0.5625

4. Given N samples $x_1, x_2, \ldots, x_N$ drawn independently from a Gaussian distribution with variance σ_2 and unknown mean μ, find the MLE of the mean.

   (a) $\mu_{MLE} = \frac{\sum_{i=1}^{N} x_i}{\sigma^2}$
   (b) $\mu_{MLE} = \frac{\sum_{i=1}^{N} x_i}{2\sigma^2 N}$
   (c) $\mu_{MLE} = \frac{\sum_{i=1}^{N} x_i}{N}$
   (d) $\mu_{MLE} = \frac{\sum_{i=1}^{N} x_i}{N-1}$

5. Continuing with the above question, assume that the prior distribution of the mean is also a Gaussian distribution, but with parameters mean μ_p and variance σ_p^2. Find the MAP estimate of the mean.

   (a) $\mu_{MAP} = \frac{\sigma^2 \mu_p + \sigma_p^2 \sum_{i=1}^{N} x_i}{\sigma^2 + N\sigma_p^2}$
   (b) $\mu_{MAP} = \frac{\sigma^2 + \sigma_p^2 \sum_{i=1}^{N} x_i}{\sigma^2 + \sigma_p^2}$
   (c) $\mu_{MAP} = \frac{\sigma^2 + \sigma_p^2 \sum_{i=1}^{N} x_i}{\sigma^2 + N\sigma_p^2}$
   (d) $\mu_{MAP} = \frac{\sigma^2 \mu_p + \sigma_p^2 \sum_{i=1}^{N} x_i}{N(\sigma^2 + \sigma_p^2)}$

6. Which among the following statements is (are) true?

   (a) MAP estimates suffer more from overfitting than maximum likelihood estimates.
   (b) MAP estimates are equivalent to the ML estimates when the prior used in the MAP is a uniform prior over the parameter space.
   (c) One drawback of maximum likelihood estimation is that in some scenarios (hint: multinomial distribution), it may return probability estimates of zero.
   (d) The parameters which minimize the expected Bayesian L1 Loss is the median of the posterior distribution.

7. Using the notations used in class and the tutorial document, evaluate the value of the neural network with a 3-3-1 architecture (2-dimensional input with 1 node for the bias term in both the layers). The parameters are as follows

$$\alpha = \begin{bmatrix} 1 & 0.2 & 0.4 \\ -1 & 0.3 & 0.5 \end{bmatrix}$$

$$\beta = \begin{bmatrix} 0.3 & 0.4 & 0.5 \end{bmatrix}$$

Using sigmoid function as the activation functions at both the layers, the output of the network for an input of (0.8, 0.7) will be

(a) 0.6710

(b) 0.6617

(c) 0.6948

(d) 0.3369

8. Which of the following statements is/are true about Neural Networks?

(a) Neural Networks can model arbitrarily complex decision boundaries.

(b) Neural Networks can be used to emulate a Gaussian kernel SVM

(c) Training of a neural network is very sensitive to the initial weights.

(d) Ideal initialization for weights would be setting all of them to zeros