Unit 7 - Week 5

Thank you for taking the Assignment 5.

Assignment 5

Your last recorded submission was on 2020-10-20, 15:52 IST **Due date: 2020-10-21, 23:59 IST.**

1) For training a binary classification model with three independent variables, you choose to use neural networks. You apply one hidden layer with three neurons. What are the number of parameters to be estimated? (Consider the bias term as a parameter) ***2 points***

- ○ 16
- ○ 21
- ○ $3^4 = 81$
- ○ $4^3 = 81$
- ○ 12
- ○ 4
- ◉ None of these

2) Suppose the marks obtained by randomly sampled students follow a normal distribution with unknown μ. A random sample of 5 marks are 30, 50, 69, 2 and 99. Using the given samples find the maximum likelihood estimate for the mean. ***1 point***

- ○ 54.2
- ○ 67.75
- ◉ 50
- ○ Information not sufficient for estimation

Assessment submitted.
X

- Artificial Neural Networks IV - Training, Initialization and Validation (unit?unit=61&lesson=65)
- Parameter Estimation I - The Maximum Likelihood Estimate (unit?unit=61&lesson=66)
- Parameter Estimation II - Priors and the MAP estimate (unit?unit=61&lesson=67)
- Parameter Estimation III (unit?unit=61&lesson=68)
- Week 5 Feedback (unit?unit=61&lesson=69)
- Quiz : Practice Assignment 5 (assessment?name=138)
- **Quiz : Assignment 5 (assessment?name=155)**

Week 6

Text Transcripts

Download Videos

3) You are given the following neural networks which take two binary valued inputs $x_1, x_2 \in \{0, 1\}$ ***1 point***
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?

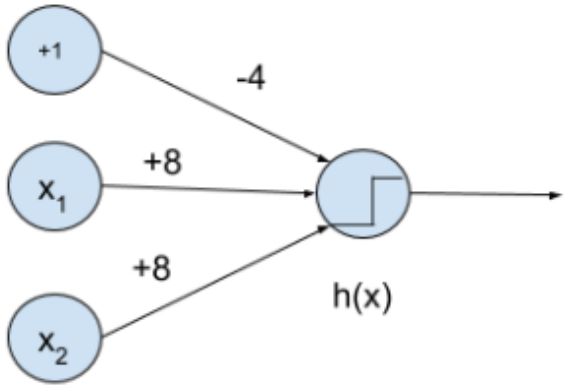

Figure 1: Q1

- (●) OR
- ( ) AND
- ( ) NAND
- ( ) None of the above.

4) Using the notations used in class, 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 ***1 point***

$$\alpha = \begin{bmatrix} 1 & 0.2 & 0.4 \\ -1 & 0.8 & 0.5 \end{bmatrix}$$

$$\beta = [0.3 \quad 0.4 \quad 0.5]$$

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

- ( ) 0.6710
- ( ) 0.9617
- ( ) 0.6948
- (●) 0.7052
- ( ) None of these

5) Which of the following statements is false: ***1 point***

- ( ) The chances of overfitting decrease with Increasing the number of hidden nodes and increasing the number of hidden layers.
- ( ) A neural network with one hidden layer can represent any Boolean function given sufficient number of hidden units and appropriate activation functions.

Assessment submitted.
X

◉ Two hidden layer neural networks can represent any continuous functions (within a tolerance) as long as the number of hidden units is sufficient and appropriate activation functions used.

6) Consider the function $f_1(x) = \frac{e^{\alpha_0+\alpha x}}{1+e^{\alpha_0+\alpha x}}$ and $f_2(x) = \frac{e^{\beta_0+\beta x}}{1+e^{\beta_0+\beta x}}$ shown in the figure below: ***1 point***

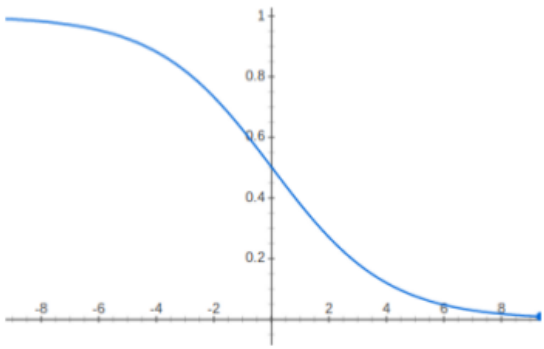

Figure 2: $f_2(x)$

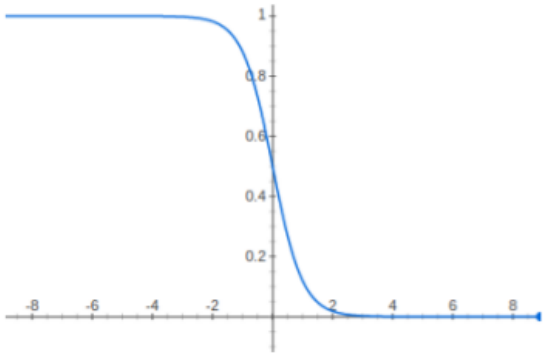

Figure 3: $f_1(x)$

Which of the following is correct?

○ $0 < \beta < \alpha$

◉ $0 < \alpha < \beta$

○ $\alpha < \beta < 0$

○ $\beta < \alpha < 0$

7) 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 $x_2 = 1, x_1 = 0, y = 5$. Given this information $\frac{\partial f}{\partial w_2}$? ***1 point***

○ 0.150

○ -0.25

○ 0.125

◉ 0.098

○ None of these

8) If the learning rate is 0.5, what will be the value of w_2 after one update using backpropagation algorithm? ***1 point***

◉ 0.4197

○ -0.4197

○ 0.5625

Assessment submitted.

X

- 0.5625

9) Which of the following are true when comparing ANNs and SVMs? ***1 point***

- [x] ANN error surface has multiple local minima while SVM error surface has only one minima
- [x] After training, an ANN might land on a different minimum each time, when initialized with random weights during each run.
- [ ] As shown for Perceptron, there are some classes of functions that cannot be learnt by an ANN. An SVM can learn a hyperplane for any kind of distribution.
- [x] In training, ANN's error surface is navigated using a gradient descent technique while SVM's error surface is navigated using convex optimization solvers.

You may submit any number of times before the due date. The final submission will be considered for grading.

Submit Answers