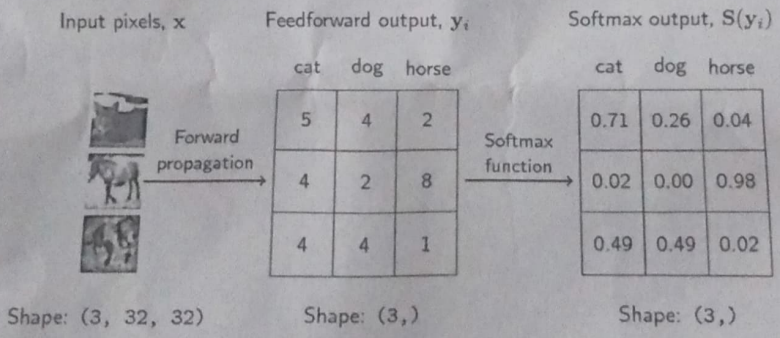
Mid-Semester Examination (March 2024)
Department of Computer Science
(Faculty of Mathematical Sciences)
University of Delhi, Delhi-110007
MCAC 204: Machine Learning
Semester II
Year of admission: 2023

Time: One Hour **Max. Marks: 30**

1. Consider a multiple linear regression problem. Assume the learning rate α is a small positive number. For $1 \leq j \leq n$, if $\frac{\partial J(\vec{w},b)}{\partial w_j}$ is a negative number, what occurs to w_j after one update step of the gradient descent algorithm? Justify your answer with the help of an example. **(2 marks)**

2. You are given a multi-class classification problem with five classes. The dataset comprises six attributes two of which takes Boolean values, and the remaining four attributes can take any of the four possible values. How many distinct hypotheses can be defined? **(3 marks)**

3. Consider the following neural network architecture to classify three input instances to one of the classes – cat, dog, and horse. **(3 marks)**

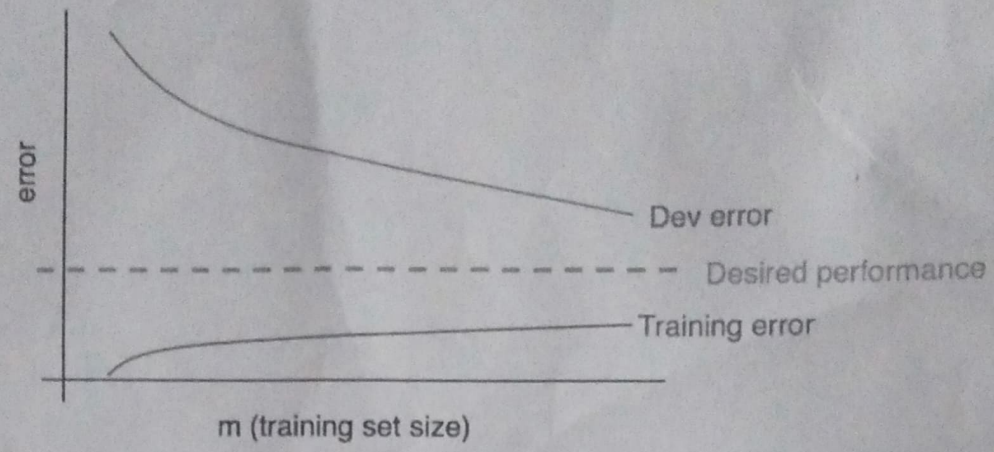


How will you interpret the output of softmax function in the above scenario?

4. You design a fully connected neural network architecture where all activations are sigmoid. You initialize the weights with large positive numbers. Is this a good idea? Justify your answer. **(3 marks)**

5. Suppose you have a regularized multiple linear regression model. How the increase or decrease in the value regularization parameter λ affects the parameter values $w_1, w_2, \ldots, w_n$? **(3 marks)**

6. Consider the following graphs illustrating the change in error for the training and development set w.r.t. increase in size of the training set. Provide the interpretation, providing the suitable reason for the behaviour of the model. Justify w.r.t. bias and variance. **(3 Marks)**



7. Assume that before training your neural network, the setting is: **(3 marks)**
   - The data is zero centered.
   - All weights are initialized independently with mean 0 and variance 0.001.
   - The biases are all initialized to 0.
   - Learning rate is small and cannot be tuned.

   Justify which activation function between tanh and sigmoid is likely to lead to a higher gradient during the first update.

8. Consider the following dataset: **(5 marks)**

Example	Citations	Size	InLibrary	Price	Editions	Buy
1	Some	Small	No	Affordable	One	No
2	Many	Big	No	Expensive	Many	Yes
3	Many	Medium	No	Expensive	Few	Yes
4	Many	Small	No	Affordable	Many	Yes

Use the candidate elimination learning algorithm to determine the most general and the most specific hypothesis for the given training data.

9. Derive a gradient descent training rule for a one-layer network with a single output unit employing Tanh activation function. The network is supplied with m training instances. The network uses the cost function for logistic regression given by: **(5 marks)**

$$E(t,y) = -(y\log(t) + (1-y)\log(1-t))$$

where t is the target output and y is the neural network output.