Mid Semester Examination Spring 2020

Name: Mahnaz Rafia Islam

Registration No: 17101007

Course Code: CSE-427

Course Title: Machine Learning.

Semester: 1st

Year: 4th

Date: 27.08.20

Answer to the question no: 1(a)

My ID - 17101007

$\theta_0 = 0 \qquad \theta_1 = 7$

Given, $\alpha = 0.1$

(i) **Dataset:**

x	y
4	10
6	16
8	19

(ii) **Hypothesis function:**

$$h_\theta(x) = \theta_0 + \theta_1 * x$$

(iii) **Parameter Initialization:**

$\theta_0 = 0 \qquad \theta_1 = 7$

$$h_\theta^{(1)}(x) = \theta_0 + \theta_1 * x^{(1)}$$

$$= 0 + 7 * 4 = 28$$

$$h_\theta^{(2)}(x) = \theta_0 + \theta_1 * x^{(2)}$$

$$= 0 + 7 * 6 = 42$$

$$h_\theta^{(3)}(x) = \theta_0 + \theta_1 * x^{(3)}$$

$$= 0 + 7 * 8 = 56$$

(iv) Cost function:

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta^{(i)}(x) - y^{(i)}\right)^2$$

$$= \frac{1}{2*3} \left\{(28-10)^2 + (42-16)^2 + (56-19)^2\right\}$$

$$= \frac{1}{6}(324 + 676 + 1369)$$

$$= 394.83 \text{ error/cost.}$$

(v) Gradient Descent:

repeat untill convergence

$$\{ \quad \theta_j = \theta_j - \frac{\alpha}{m} \frac{d}{d\theta_j} \{J(\theta)\} \quad \}$$

$$\theta_0 := \theta_0 - \frac{\alpha}{m} \sum_{i=1}^{m} \left(h_\theta^{(i)} - y^{(i)}\right)$$

$$= 0 - \frac{0.1}{3} \{(28-10) + (42-16) + (56-19)\}$$

$$= -2.7$$

$$\theta_1 := \theta_1 - \frac{\alpha}{m} \sum_{i=1}^{m} \left(h_\theta^{(i)} - y^{(i)}\right) . x^{(i)}$$

$$= 7 - \frac{0.1}{3} \left\{ (28-10).4 + (42-16).6 + (56-19).8 \right\}$$

$$= 7 - 17.46$$

$$= -10.46$$

$\therefore \theta_0 = -2.7$ and $\theta_1 = -10.46$.

Answer to the question no: 1(b)

I do ~~not~~ agree. because logistic regression is a classification model. Logistic Regression is one type of binary classification, where the output maybe only two types. such as 0/1, Yes/No, True/False etc, which indicates a classification model.

In Regression, output value is a continious value, such as - age of different people, temperature of today etc. But in logistic regression output value is/can be only two types, that's why it's not a regression algorithm. Logistic regression ~~has~~ has regression in it's name because the underlying techinique is quiete the same as linear regression. The term Logistic is taken from the Logit function. but it's not a regression problem, it's a binary classification model.

Answer to the question no: 2(a)

$a = 17101007 \% 5 = 2$

$b = 17101007 \% 7 = 0$

$c = 17101007 \% 9 = 8$

In the given dataset,

$$M_i(x_1) = \frac{2+3+6}{3} = 3.67$$

$$\cancel{S_i(x_1) = 6-2} \quad S_i(x_1) = 6-2 = 4$$

$$x_1^{(1)} = \frac{x_1^{(1)} - M_i(x_1)}{S_i(x_1)} = \frac{2-3.67}{4} = -0.4175$$

$$x_1^{(2)} = \frac{3-3.67}{4} = -0.1675$$

$$x_1^{(3)} = \frac{6-3.67}{4} = 0.5825$$

Now,

$$\mu_i(x_2) = \frac{0+10+50}{3} = 20$$

$$s_i(x_2) = 50-0 = 50$$

$$x_2^{(1)} = \frac{x_2^{(1)} - \mu_i(x_2)}{s_i(x_2)} = \frac{0-20}{50} = -0.4$$

$$x_2^{(2)} = \frac{10-20}{50} = -0.2$$

$$x_2^{(3)} = \frac{50-20}{50} = 0.6$$

Again,

$$\mu_i(x_3) = \frac{8+150+540}{3} = 232.67$$

$$s_i(x_3) = 540-8 = 532$$

$$x_3^{(1)} = \frac{x_3^{(1)} - \mu_i(x_3)}{s_i(x_3)}$$

$$= \frac{8-232.67}{532} = -0.42$$

$$x_3^{(2)} = \frac{150-232.67}{532} = -0.155$$

$$x_3^{(3)} = \frac{540-232.67}{532} = 0.577$$

The dataset is now normalized.

Answer to the question no: 2(b)

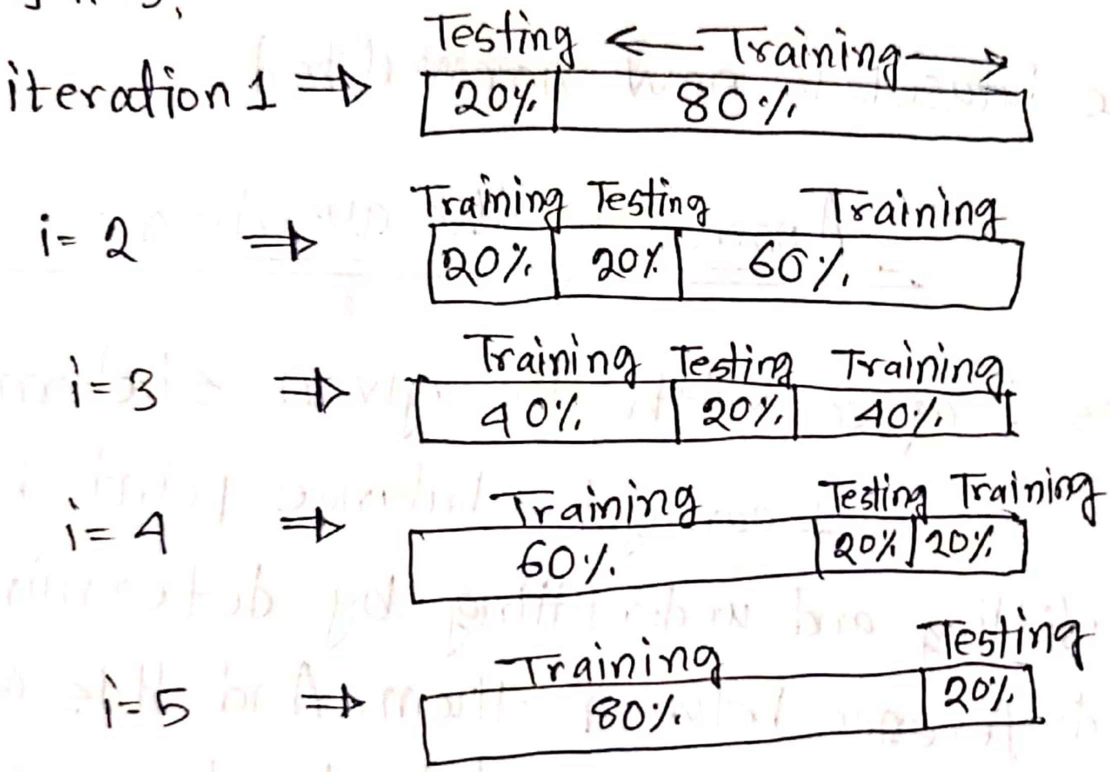
Yes I agree with the given statement.

I can get a balance point between overfitting and underfitting by determining the difference between them. And this can be done by any fitted models to training data and testing data. K-fold cross validation is one of the ~~bet~~ best to do this task. Here, k is the number of groups that a given data sample is to be split into.

Over fitting can be prevented by getting more training example and trying smaller set of features. Underfitting can be

prevented by adding more features and by adding polynomial feature.
In K-fold the procedure is like following:

if K=5,



So this is an example of K-fold cross validation where the dataset is randomly partitioned into K-equal size sub samples. This significantly reduces overfitting and underfitting as well as High Bias and High Variance.

Answer to the question no: 4(a)

Frequency Table:

	Yes	No
S(Small)	1	2
M(Medium)	2	3
L(Large)	3	0
	6	5

Likelihood Table:

	Yes	No
S(Small)	1/6	2/5
M(Medium)	2/6	3/5
L(Large)	3/6	0/5

Given,

Mutation Rate 'medium' $\rightarrow B$

Probability of new species emergence is 'Yes' $\rightarrow A$

$$P(B) = \frac{5}{11}, \quad P(A) = \frac{6}{11}$$

$$P(B|A) = \frac{2}{6}$$

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

$$= \frac{\frac{2}{\cancel{6}} \times \frac{\cancel{6}}{11}}{\frac{5}{11}}$$

$$= \frac{2}{\cancel{11}} \times \frac{\cancel{11}}{5}$$

$$= \frac{2}{5}$$

(Ans)

Answer to the question no: 4 (b)

No of positive labels = 17101007 % 4 = 3

No of negative labels = 17101007 % 7 = 0

Here, $P_+ = \frac{3}{3}$ (Fraction of positive label)

$P_- = \frac{0}{3}$ (Fraction of negative label)

Entropy,

$$H(S) = -P_{(+)}\log_2(P_+) - P_{(-)}\log_2(P_-)$$

$$= -\left(\frac{3}{3}\right)\log_2\left(\frac{3}{3}\right) - \left(\frac{0}{3}\right)\log_2\left(\frac{0}{3}\right)$$

$$= 0.$$

This is the entropy of the given dataset.