Question 1 – (20 marks)

1. **(1 mark)** Machine learning can be split up into two predominant categories of problems. What are they?
2. **(1 mark)** Suppose you are working on weather prediction and you would like to predict weather it will be raining at 5 p.m. the next day. What category of supervised learning would this belong to?
3. **(2 mark)** What is Arthur Samuel's definition of Machine Learning?
4. **(2 marks)** Why is it necessary to normalize the features prior to running gradient descent for regression or classification?
5. **(2 marks)** Define linear regression and classification for one variable? Explain with examples.
6. **(1 mark)** What is the purpose of the learning rate? What happens if this value is too small? What happens if this value is too large?
7. **(1 marks)** What means "Batch gradient" descent? What means "Stochastic gradient" or "on-line" gradient descent?
8. **(1 mark)** What is the hypothesis $h_\theta(x)$ used in logistic regression?
9. **(2 mark)** What is the difference between classification and logistic regression?
10. **(1 mark)** What is the hypothesis/Model in machine learning?
11. **(1 mark)** Adding new features and finding the optimal parameters using the training set will allow for high accuracy though there is a chance for over fitting. (True or False)?
12. **(1 marks)** Explain what the problem of over fitting is and explain how you would reduce over fitting in machine learning problems.?
13. **(2 marks)** What is Tom Mitchell's definition of Machine Learning?
14. **(1 mark)** What is the problem of overfitting?
15. **(1 marks)** What is the goal of classification?

****_____**_____**_____**_____**_____**_____**_____**_____**_____**_____****

Question 2 – (10 marks)

Recall Suppose we are given $m = 5$ training examples:

$\theta_0 = \theta_1 = 0 \quad \alpha = 1 \quad N = \quad m = 5$

$\theta_0 \leftarrow \theta_0 -$

x	y
-2	-1
-1	0
1	2
2	3
4	6

Parts to solve

1. **(6 marks)** Given the initial values of $\theta_0, = \theta_1 = 0$ and a learning rate of $\alpha = 0.1$, compute $N = 3$ iterations of gradient descent for finding these parameters for linear regression. Compute both the parameters θ_0, θ_1 as well as the cost of iteration.

2. **(2 marks)** After the third iteration in this question, predict what the outputs would be when $x = -0.5, 0.5, and\ 1.5$.

Question 3 – (25 marks)

Suppose we have the following table that illustrates our training data set concerning thermal resistors where our goal is to create a prediction model where given the current (in milliamps) flowing through the resistor, we want to determine what the voltage (in millivolts) across the resistor would be.

Example	Current x (mA)	Voltage y (mV)
1	1	2.2
2	2.5	3.76
3	4	5.4
4	6	7.9

Because of the highly non-linear relationship between the current and voltage, we will be using new features to allow for higher accuracy. Specifically, we will use two features for our prediction model: $x_1 = x^2, x_2 = \sqrt{x}$.

Parts to solve

1. **(2 marks)** Using the measurements above, complete the following table in order to create these two features. You may complete the table here or do this in your response booklets:

Example	x_1	x_2
1		
2		
3		
4		

2. (7 marks) Because of the wide dynamic range, we will need to **feature normalize** the above table before we create our prediction model. Normalize the features above so that they exhibit **zero mean and unit variance**. You can complete the table below here or do this in your response booklets:

Example	x_1	x_2
1		
2		
3		
4		

3. (12 marks) With the **normalized features**, using a learning rate $\alpha = 0.75$, the **regularization parameter** $\lambda = 0.5$ and with the initial parameters $\theta_0 = \theta_1 = \theta_2 = 1$, compute $N = 1$ iteration and state the parameters $\theta_0, \theta_1, \theta_2$ as well as the cost $J(\theta)$.

4. (4 marks) Using the learned parameters found in the previous step after the first iteration, predict what the output voltage would be if the input current is 2 mA and 3 mA.

Question 4 – (5 marks)

Suppose we have the following training example data set. There are three classes in total with two features per training example. A table that shows the features per training example x_1, x_2 and corresponding expected output class y are shown below:

x_1	x_2	y
-2	3	1
-2	4	1
-1	3	1
3	3	2
4	4	2
4	3	2
-1	-1	3
-2	0	3
-3	-2	3

(5 marks) Plot these points in cartesian grid. Which type of classification do you suggest for this question (explain)?