Name – Surname:

Number:

Q1(20)	Q2(10)	Q3(33)	Q4(20)	Q5(20)	T(100)

Q1.(20p) Assume you have trained a logistic regression over patients and found $h_\theta(x) = g(-5 + 2x)$ where x denotes the size of the tumor of the patient. Given size of the tumor, you want to predict whether the patient is cancer (y=1) or not (y=0). You collected the data for the size of the tumor for 3 people who is really cancer and 7 people who do not have cancer. The data is given below:

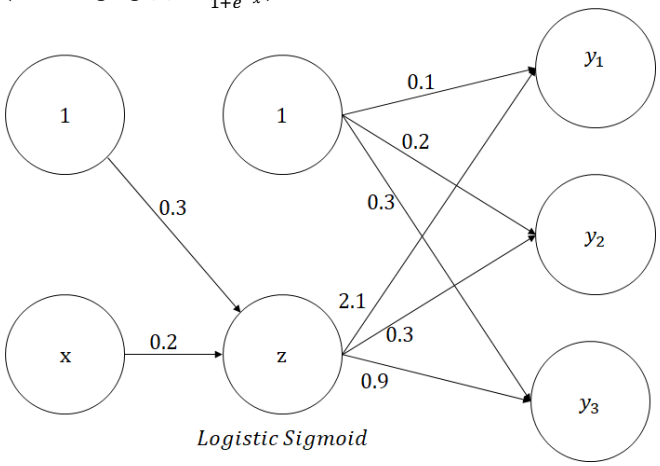
ID	Cancer ?	Tumor Size (x)	
1	Yes	3	
2	Yes	4	
3	Yes	2	
4	No	2	
5	No	1.5	
6	No	2.2	
7	No	3.5	
8	No	2.7	
9	No	1.8	
10	No	0.5	

a) (15) What is the precision and recall of this algorithm?

b) (5p) What is the F1 score?

Q2. (10p) Assume that you are dealing with a multiclass classification problem and there are three classes. You build an artificial neural network (ANN) with single hidden layer. Transfer function of the hidden layer is *logistic sigmoid*. Trained ANN is given below. Estimate the class of an entry $x = 3$.

(Hint: $logsig(x) = \frac{1}{1+e^{-x}}$ *)*



Q3.(a)(10p) Consider regularized logistic regression. Let

- $J(\theta) = \frac{1}{2m}\left[\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2 + \lambda\sum_{j=2}^{n}\theta_j^2\right]$
- $J_{train}(\theta) = \frac{1}{2m_{train}}\left[\sum_{i=1}^{m_{train}}(h_\theta(x_{train}^{(i)}) - y_{train}^{(i)})^2\right]$
- $J_{CV}(\theta) = \frac{1}{2m_{CV}}\left[\sum_{i=1}^{m_{CV}}(h_\theta(x_{CV}^{(i)}) - y_{CV}^{(i)})^2\right]$

For different values of regularization parameter λ both cost functions behave differently. Please plot J_{train} and J_{CV} as a function of λ (on the same plot) and comment on these plot.

(b)(8p) In logistic regression, the cost function for our hypothesis predicting $h_\theta(x)$ on a training example that has label $y \in \{0,1\}$ is:

$$\text{Cost}(h_\theta(x), y) = \begin{cases} -\log h_\theta(x) \text{ if } y = 1 \\ -\log(1 - h_\theta(x)) \; if \; y = 0 \end{cases}$$

Please check as true (T) or false (F). Each incorrect answer will cost you -2 point.

- ☐ If $h_\theta(x) = y$, then $\text{cost}(h_\theta(x), y) = 0$ (for $y = 0$ and $y = 1$).
- ☐ If $y = 0$ then $\text{cost}(h_\theta(x), y) \to \infty$ as $h_\theta(x) \to 1$
- ☐ If $y = 0$ then $\text{cost}(h_\theta(x), y) \to \infty$ as $h_\theta(x) \to \infty$
- ☐ Regardless of whether $y = 0$ or $y = 1$, if $h_\theta(x) = 0.5$, then $\text{cost}(h_\theta(x), y) > 0$

c. (15p). Why do we need train set, cross validation set and test set? Please explain.

Q4. (20p) We have seen several machine learning algorithms so far (Linear regression, logistic regression, neural networks etc). For the following problems, please indicate the type of the problem (regression, classification, clustering, etc) and propose solution method(s).

a) Find the price of the cars by looking at its features:

Problem	Regression
Method	Linear Regression

b) Predict whether the weather will be cloudy, rainy, sunny or snowy.

Problem	
Method	

c) Predict next day's weather temperature by looking at temperature ,cloudiness, sun, humidity, etc.

Problem	
Method	

d) Predict the letter grades by looking at the first midterm and the first 5 assignment grades.

Problem	
Method	

e) Group the IE students by looking at their grades of core courses (operations research, statistics, production planning, simulation and quality management) in their transcript.

Problem	
Method	

f) Predict whether a 1000x1000 resolution picture is an automobile or not. Do we need some preprocess?

Problem	
Method	

Q5(20) The K-means algorithm is given as follows:

Randomly initialize K cluster centroids $\mu_1, \mu_2, \dots, \mu_K \in R^N$

Repeat {

For i=1 to m

$c^{(i)}$:= index (from 1 to K) of cluster centroid closest to $x^{(i)}$

$\mu_k :=$ average of points assigned to cluster k

}

I have two features, say X1 and X2 and 4 sample points and I want to divide them into two clusters using K-means algorithm. The points are (0,0), (0,2) , (4,2) , 4,0). As the algorithm proposes, we will start the algorithm by choosing two random data points as the centroids of the clusters. Let the center of the first cluster be (0,0) and let the center of the second cluster be (4,2). Now please iterate the algorithm for one cycle, i.e., first find which cluster does each data point belong to, and then update the centroid of the two clusters.