YILDIZ TECHNICAL UNIVERSITY
FACULTY OF MECHANICAL ENGINEERING
DEP OF INDUSTRIAL ENGINEERING
2022/2023 FALL SEMESTER
END3971 – Artificial Intelligence and Expert Systems
EXAM: Midterm 1

MUDEK Criteria					Total
Question(s)	1	2	3	4	
Grade					

Name-Surname : KEY
Student ID :
Signature :
Res. Assist :

Duration: 60 minutes
Date: 28.11.2021-12:00
Exam Place: B401 B402 B302

PLEASE WRITE YOUR RESULTS TO THE BOXES!

Q1. **(40p)** Recall the example that we have solved in the lecture to find the price of a house. We will do the same thing here, in the exam. I have a dataset of 21 houses given in the following

No	Size	Price		No	Size	Price
1	88	280		12	160	546
2	128	438		13	216	686
3	200	690		14	232	814
4	120	414		*15*	*156*	*496*
5	172	558		*16*	*132*	*420*
6	148	466		*17*	*208*	*690*
7	264	916		*18*	*200*	*652*
8	216	732		*19*	*152*	*526*
9	280	960		*20*	*160*	*520*
10	240	816		**21**	**216**	**760**
11	252	842				

Each observation has size as its feature and the output is the price. For example the first house's size is 44m2 and its price is 140 TL. I ordered the data randomly and then I picked first 14 houses as my training set, the next 6 houses (italic ones) as my CV (cross validation) set and the last one for the test set (bold one). Recall that the associated error for any model (both for train, CV and test) is given by the following:

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

Model 1 : $\theta_0 = -13$ and $\theta_1 = 3.5$

Model 2 : $\theta_0 = -3$ and $\theta_1 = 3.3$

For the training and CV errors, I have this following table →

	Training	CV
Model 1	260.50	403.17
Model 2	400.26	104.39

For Model1, I have calculated the following value for the first 13 data points in order to calculate the ***training*** error:

Model 1	$\sum_{i=1}^{13}\left(h_\theta(x^{(i)}) - y^{(i)}\right)^2 = 7069$

For Model2 I have calculated the following value for the first 5 data points of the CV set (15 to 19) to calculate the ***CV* error**:

Model 2	$\sum_{i=15}^{19}\left(h_\theta(x^{(i)}) - y^{(i)}\right)^2 = 1228$

a. (10p) Calculate the train error for Model 1.

Train Error 260.50

b. (10p) Calculate the CV error for Model 2.

CV Error 104.39

c. (5p) Which model will you pick? Why? Briefly please!

Model 2.

Students are NOT permitted to bring mobile phones and/or any other unauthorised electronic devices into the examination room

Good Luck...

d. (10p) What is the real world performance of the model that you picked in (c)?

144.5 / 1260.2

e. (5) What is the predicted price for a 100 m² house by Model1 and Model2?

Model 1	Model 2
337	327

Q2. (20p) Let's continue with the first question. I want to use a different metric for the error function. In particular, I will use the following:

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

That is, instead of taking the square, I take the absolute value and calculate the errors accordingly. Well, I calculated the train errors of both models. They are **8.29** and **12.56**, for Model1 and Model2, respectively. For CV errors I calculated the following values:

Model 1	$\sum_{i=15}^{19}\left\|h_\theta(x^{(i)}) - y^{(i)}\right\| = 133.00$
Model 2	$\sum_{i=15}^{19}\left\|h_\theta(x^{(i)}) - y^{(i)}\right\| = 67.40$

a) (10p) What is the CV error of Model 1 and Model 2?

Model 1	Model 2
13.33	6.03

b) (10p) I have calculated test errors as 8.5 and 25.1 for Model1 and Model2, respectively. Using all these results, which model would you choose under the new error function? Briefly explain.

Model 1.

Q3. (30p) We have a dataset of patients that has a specific tumor type. Each data point has the size of the tumor as its single feature and an output that shows whether a tumor is really a cancer (y=1) or not (y=0). The calculated logistic regression equation is given as:

$h_\theta(x) = g(2x - 5)$ where $g(z) = {}^{1}/_{1 + e^{-z}}$

Here *x* is the size of the tumor and *g(z)* is the sigmoid function.

We have the following **test set** that has five patients.

Patient	1	2	3	4	5
Size	3	4	1.5	1	2
Cancer?	1	1	0	0	1
Prediction	1	1	0	0	0

0.731 0.953 0.119 0.047 0.269

Finally recall the error metric that is given in the lecture:

$$\frac{1}{m}\sum_{i=1}^{m} err\left(h_\theta\left(x^{(i)}\right), y^{(i)}\right)$$

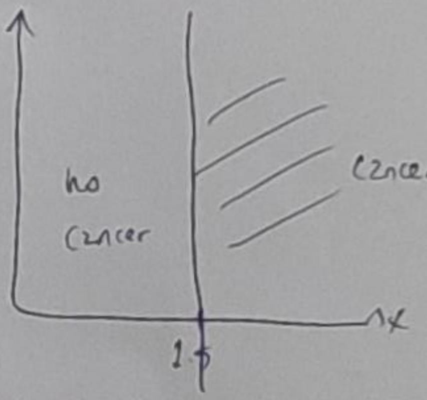
Where $err\left(h_\theta\left(x^{(i)}\right), y^{(i)}\right) = 0$ if your prediction is correct and 0 otherwise. Use natural logarithm (LN) whenever necessary.

a) (10) Find the prediction of the model for all test set (5 patients) given in the table. Please write the results to the last line of the table given in the question above.

b) (5p) Consider a patient with a tumor of size 1.7 cm. What is the predicted probability of this patient being cancer?

0.168

c) (10p) Please plot the decision boundary.



d) (5) What is the performance of your model in terms of the given error metric?

80%

Q4. (10p) Briefly explain why do we use test and cross validation sets. You can write at the back of the paper if you need.