CS412 - Machine Learning - 2021/02

Assignment #2 Report

Name-Surname: Muhammed Orhun Gale
Student No : 26754
ipynb Link:
https://colab.research.google.com/drive/1tBwkOlMN6ZSuoSiQDQoTx-Eun0qOJLsi?usp=sharing

Q1)

8)

Problem Definition: House price prediction problem is a regression problem where as inputs there exist area of the houses, number of rooms, view of the house and crime rates in the neighborhood that the house is located and output is the price of the given house according to the above stated parameters. In order to solve this problem, a regression model can be built. Therefore, I build Neural Network regression models with 1-Hidden Layers and 2-Hidden Layers with different activation functions and layer size parameters.

Train/val/test sets, size and how split: As the training set, I had a data set with 4800 examples with 5 attributes such as number of rooms, square meters, view, crime rate and price of the house. I used 10% of the training set as the validation set and trained my models with the remaining 4320 (90%) examples. For the testing set, I had a data set with 1200 examples with 4 attributes which are the same as the training set except the price since it is the parameter that I would predict.

Feature extraction and preprocessing: Attributes view and crime_rates were categorical variables and in order to efficiently process them I used get_dummies function of the pandas library and turned them into numerical values. Also, in order to employ the data efficiently, I used minmax scaler and scaled all variables between 0 and 1.

Observations

- Try a few learning rates for N=25 hidden neurons, train for the indicated amount of epochs. Comment on what happens when learning rate is large or small? What is a good number/range for the learning rate?

1-Hidden Layer w/ 25 nodes:

```
sigmoid 0.001 25 MSE Score:  0.00036183500196784735
sigmoid 0.01  25 MSE Score:  0.0006680361111648381
sigmoid 0.1   25 MSE Score:  0.0001257983822142705 3
```

```
Relu    0.001 25 MSE Score:  0.0005061730043962598
Relu    0.01  25 MSE Score:  0.00011399292998248711
Relu    0.1   25 MSE Score:  0.0006180982454679906
```

I observed that the optimal score (min loss for the cost function) occurs with the learning rate around 0.01 for a NN which has 1-hidden layer and ReLU activation function for models that have 25 nodes. Also, I tried this learning rate along with other given rates on models which have 1 or 2 hidden layers with 'RELU' or 'Sigmoid' activation functions and node sizes vary around 25, 50, 100 nodes along with learning rates 0.001 and 0.1. Consequently, I found that the best performing model has two layers with 0.001 learning rate, RELU activation function and 25 nodes on first layer and 100 nodes in second layer along with the best performing 1-hidden layered model also have 0.001 learning rate.

- Use that learning rate and vary the number of hidden neurons for the given values and try the indicated number of epochs. Give the validation mean squared errors for different approach and meta-parameters tried **in a table** and state which one you selected as your model. How many hidden neurons give the best model?

Num hid layers	#Layer 1	#Layer 2	Learning Rate	Activation Func.	MSE Score
2	25	100	0.001	ReLU	5.8343037380836904e-05
2	25	100	0.001	Sigmoid	0.00033768106368370354
1	100	-	0.001	ReLU	6.558789755217731e-05
1	100	-	0.001	Sigmoid	0.00035190890775993466
...					
1	25	-	0.001	ReLU	0.0005061730043962598
1	50	-	0.001	ReLU	6.825794844189659e-05
2	25	50	0.001	ReLU	7.015091978246346e-05
2	25	25	0.001	ReLU	8.615724800620228e-05
2	50	100	0.001	ReLU	6.011507866787724e-05
2	100	100	0.001	ReLU	7.17694201739505e-05
...					

Best result for 1-hidden layered model is 6.558789755217731e-05 with 100 nodes however, best result overall is 5.8343037380836904e-05 with a 2-hidden layered model with 25 nodes in hidden layer 1 and 100 nodes with hidden layer 2. Since it is the best performing model (gives the minimum loss), I selected the 2-hidden layered (25,100) with ReLU function as my model to evaluate the test set.

- State what your test results are with the chosen approach and meta-parameters:

We have obtained the best results on the validation set with the two-hidden layered NN approach using a value of 0.001 for learning_rate parameter, 25 nodes on the first hidden layer, 100 nodes on second hidden layer and RELU for activation function. The result of this model on the test data gave MSE 0.0004797478031832725 as the score.

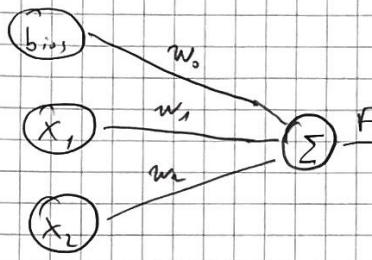
- How slow is learning? Any other problems?

Training 1-hidden layer models took ~3 minutes while training 2-hidden layer models took ~8-9 minutes but in return for long training time 2-hidden layer models performed better on validations. Also, models with lower learning rates took more time to train. Additionally, validation and testing took under 20 seconds. I did not encounter any problems.

Q2 and Q3 are on the following pages 4 and 5 ======>

Q2)

(Q 2)



- If the neural network would replicate an AND function the truth table table should be like this where $F(Y)$ is the activation function

X_1	X_2	$F(\Sigma)$
0	0	0
0	1	0
1	0	0
1	1	1

⇒ Then the function can be stated as follows:

$$\Sigma = w_1 \cdot X_1 + w_2 \cdot X_2 + w_0$$

⇒ Then the truth table should be like this:

X_1	X_2	Σ
0	0	w_0
0	1	$w_2 \cdot X_2 + w_0$
1	0	$w_1 \cdot X_1 + w_0$
1	1	$w_1 \cdot X_1 + w_2 \cdot X_2 + w_0$

⇒ Assume that

$w_0 = -5$
$w_1 = 3$
$w_2 = 3$

⇒

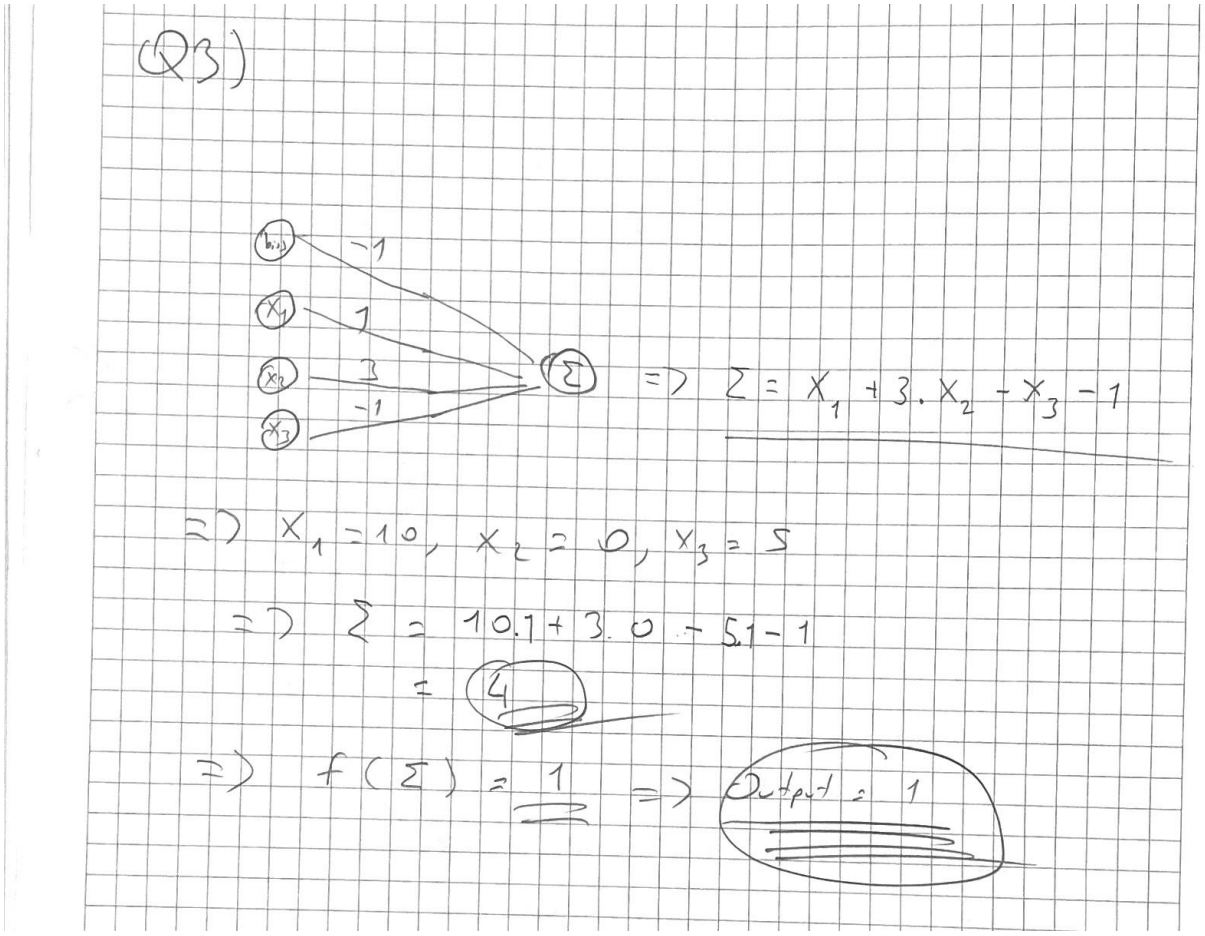
X_1	X_2	Σ	$F(\Sigma)$ (Activation)
0	0	-5	0
0	1	-2	0
1	0	-2	0
1	1	1	1

⇒ Bias can be $= -5$

$w_1 = 2$

$w_2 = 2$

Q3)


Q3)
b=1
x1
x2
x3
-1
1
3
-1
Σ
=> Σ = X1 + 3.X2 - X3 - 1
=> X1 = 10, X2 = 0, X3 = 5
=> Σ = 10.1 + 3.0 - 5.1 - 1
= 4
=> f(Σ) = 1 => Output = 1