Chiranjeev
P21CS007
chiranjeev.1@iitj.ac.in

Assignment - 6
Neural Network

27-11-2021

Colab file link -
https://colab.research.google.com/drive/15l93KvFhohuaWnEpxEFdy7Dsfqobh3mh?usp=sharing

Classification Task

I. Dataset Description:

a. **Name of Dataset :** Breast Cancer Wisconsin (Diagnostic) Dataset (link - Dataset link)

b. **About Dataset :** Features in the data are computed from a digitalized image of a fine needle aspirate (FNA) of breast mass that describe characteristics of the cell nuclei present in the image in the 3-dimensional space.

c. **Dataset features :**

id
diagnosis
radius_mean
texture_mean
perimeter_mean
area_mean
smoothness_mean
compactness_mean
concavity_mean
concave points_mean
symmetry_mean
fractal_dimension_mean
radius_se
texture_se
perimeter_se
area_se
smoothness_se
compactness_se
concavity_se
concave points_se
symmetry_se
fractal_dimension_se
radius_worst
texture_worst
perimeter_worst
area_worst
smoothness_worst
compactness_worst
concavity_worst
concave points_worst
symmetry_worst
fractal_dimension_worst

d. **Features Dropped :**
Feature ‘id’ is dropped as for classification task id is not an attribute of breast. And also found that there are not any Not Available (NA) values in the dataset.

e. **Features Selected :**

All features mentioned above are chosen except 'id' for training and testing as all features describes the attributes of breast for cancer detection.

II. Splitting the Dataset:

Used train_test_split of sklearn to split the dataset into train and test.
Split the Dataset into: 60% - train set, 20% validation set, 20% test set

- **Learning Rate:** 0.001
- **Optimizer:** Adam
- **Number of Epochs**: 5
- **Batch size**: 16

- **Observations with different Architectures:**

1. **On trying various Activation Functions:**

Neural Network	**Network 1**	**Network 2**
Activation function **(Softmax used on last layer for classification with confidence)**	Relu	Tanh
Layers	• layer1(input layer) = 31 • layer2(hidden layer) = 31 • layer3(Output layer) = 2	• layer1(input layer) = 31 • layer2(hidden layer) = 31 • layer3(Output layer) = 2
Train loss (Categorical Cross Entropy)	0.3833	0.3744
Test loss (Categorical Cross Entropy)	0.3852	0.3835
Train Accuracy	94.7802%	96.9780%
Test Accuracy	**95.6043%***	95.6043%
Confusion Matrix for Test set	confusion_matrix: [[33 3] [1 54]]	confusion_matrix: [[33 3] [1 54]]

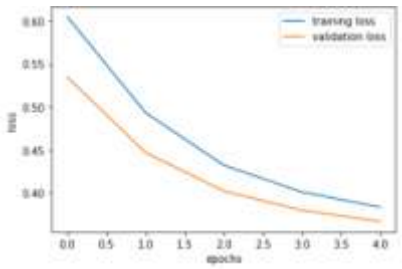
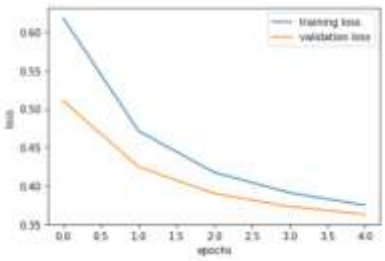
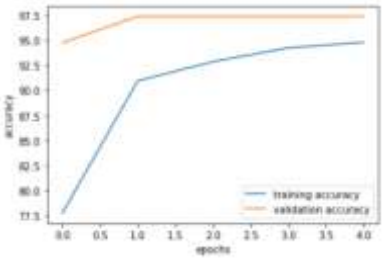
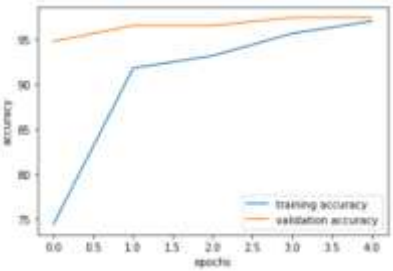
Epochs Vs Train/Validation Loss graph		
Epochs Vs Train/Validation Accuracy graph		

Table 1

Observation from Table 1: On analyzing and visualizing Table 1. We can say that Network 1 perform better than Network 2 as the Network 1 is performing better on test set (i.e., test accuracy is 95.6043%). While Network 2 has lesser accuracy on test set as compared to Network 1.
Therefore, Network 1 is a better model.

2. **On increasing number of layers to 4 and now comparing with different Activation functions:**

Neural Network	**Network 3**	**Network 4**
Activation function **(Softmax used on last layer for classification with confidence)**	Relu	Tanh
Layers	• layer1(input layer) = 31 • layer2(hidden layer) = 31 • layer3(hidden layer) = 31 • layer4(Output layer) = 2	• layer1(input layer) = 31 • layer2(hidden layer) = 31 • layer3(hidden layer) = 31 • layer4(Output layer) = 2
Train loss (Categorical Cross Entropy)	0.3549	0.3521
Test loss (Categorical Cross Entropy)	0.3639	0.3618
Train Accuracy	97.8022%	97.5275%
Test Accuracy	95.6043%	96.7032%

Confusion Matrix for Test set	`confusion_matrix:`
	`[[33  3]`
	`[ 1 54]]`
		`confusion_matrix:`
		`[[34  2]`
		`[ 1 54]]`

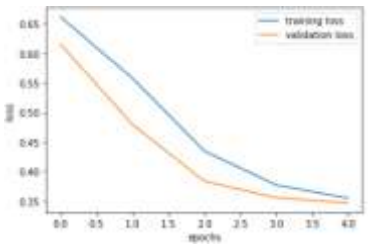
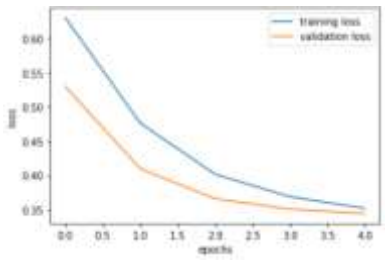
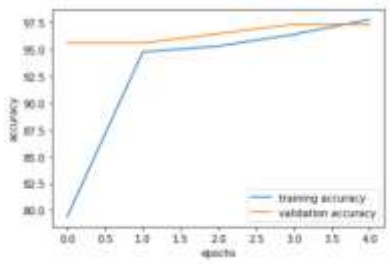
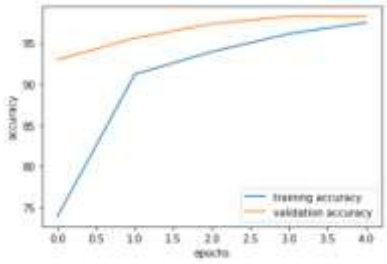
Epochs Vs Train/Validation Loss graph		
Epochs Vs Train/Validation Accuracy graph		

Table 2

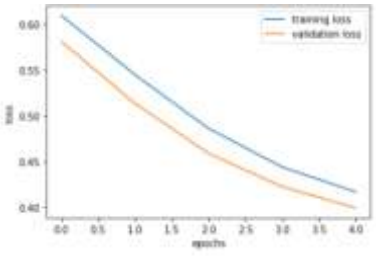
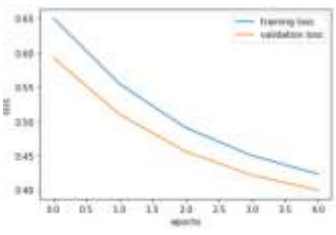
Observation from Table 2: On analyzing and visualizing Table 2. We can say that Network 2 perform better than Network 1 on the basis of Test Accuracy. However both the models are not performing well on test set as compared to the train set. As both models have less performance than train set.

Therefore, Comparatively Network 2 is a better model than Network 1. However, individually both are not good models.

3. On changing number of neurons in layers and now comparing with different Activation functions:

Neural Network	**Network 5**	**Network 6**
Activation function **(Softmax used on last layer for classification with confidence)**	Relu	Tanh
Layers	• layer1(input layer) = 31 • layer2(hidden layer) = 10 • layer3(Output layer) = 2	• layer1(input layer) = 31 • layer2(hidden layer) = 31 • layer3(Output layer) = 2
Train loss (Categorical Cross Entropy)	0.4166	0.4232
Test loss (Categorical Cross Entropy)	0.4147	0.4242
Train Accuracy	93.9560%	92.5824%
Test Accuracy	**94.5054%**	93.4065%

Confusion Matrix for Test set	`confusion_matrix:`
	`[[32  4]`
	`[ 1 54]]`
		`confusion_matrix:`
		`[[32  4]`
		`[ 2 53]]`

Epochs Vs Train/Validation Loss graph		

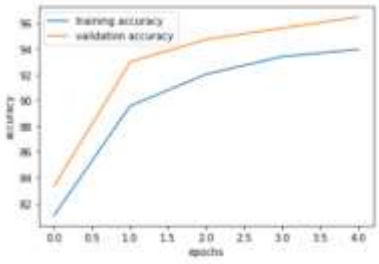
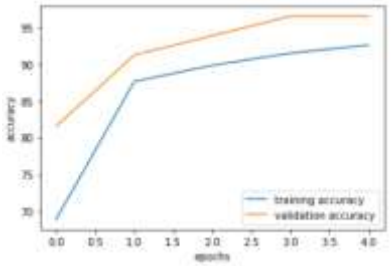
Epochs Vs Train/Validation Accuracy graph		

Table 3

Observation from Table 3: On analyzing and visualizing Table 3. We can say that Network 5 perform better than Network 6 as the Network 5 is performing better on test set (i.e., test accuracy is 94.5054%). While Network 6 has lesser accuracy on test set as compared to Network 5.
Therefore, Network 5 is a better model.

- **<u>OVERALL RESULT (for Regression Task)</u> :**

  From Table 1, Table 2, Table 3. We can analyse on seeing on the basis of the convergence of loss (train as well as validation) and performance of each model on test set on the basis of test Accuracy.

  Note:
  - **Test Accuracy of Network 1(95.6043%) model is best as compared to other models, is marked with * in Table 1.**
  - Have not considered Network 3 for comparison with this above Network 1 result for accuracy because on individual basis Network 2 is overfitting, thus it has less test accuracy(95.6043%) as compared to its own train accuracy(97.8022%).
  - Also, Observed that it's taking very less epochs to train a good model, may be because the dataset is not much complex dataset.

Regression Task

I. Dataset Description:

a. Name of Dataset : Boston Housing Dataset (link - Dataset link)

b. About Dataset: The Boston Housing Dataset is a derived from information collected by the U.S. Census Service concerning housing in the area of Boston MA.

c. Dataset features :

- CRIM - per capita crime rate by town
- ZN - proportion of residential land zoned for lots over 25,000 sq.ft.
- INDUS - proportion of non-retail business acres per town.
- CHAS - Charles River dummy variable (1 if tract bounds river; 0 otherwise)
- NOX - nitric oxides concentration (parts per 10 million)
- RM - average number of rooms per dwelling
- AGE - proportion of owner-occupied units built prior to 1940
- DIS - weighted distances to five Boston employment centres
- RAD - index of accessibility to radial highways
- TAX - full-value property-tax rate per $10,000
- PTRATIO - pupil-teacher ratio by town
- B - 1000(Bk - 0.63)^2 where Bk is the proportion of blacks by town
- LSTAT - % lower status of the population
- MEDV - Median value of owner-occupied homes in $1000's

d. Data Pre-processing :
Found that there are not any Not Available (NA) values in the dataset.

e. Features Selected :
All features mentioned above are chosen for training and testing.

f. Target Value to be Predicted :
MEDV (Median Value of owner-occupied homes in $1000's)

II. Splitting the Dataset:

Used train_test_split of sklearn to split the dataset into train and test.
Split the Dataset into: 75% - train set, 10% validation set, 15% test set

- **Learning Rate:** 0.005
- **Optimizer:** Adam
- **Number of Epochs**: 100
- **Batch size**: 64

- **Observations with different Architectures:**

1. On trying various Activation Functions:

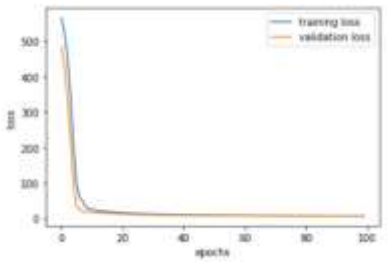
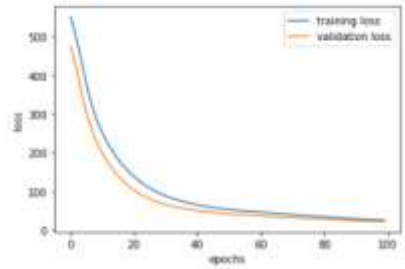
Neural Network	**Network 1**	**Network 2**
Activation function	Relu	Tanh
Layers	• layer1(input layer) = 13 • layer2(hidden layer) = 32 • layer3(hidden layer) = 16 • layer4(Output layer) = 1	• layer1(input layer) = 13 • layer2(hidden layer) = 32 • layer3(hidden layer) = 16 • layer4(Output layer) = 1
Train loss (MSE)	7.7426	25.5968
Test loss (MSE)	**7.6429***	28.5481
R2 score on train set	0.89964	0.66797
R2 score on test set	**0.89970***	0.62569
Epochs Vs Train/Validation Loss graph		

Table 1

Observation from Table1: On analyzing and visualizing Table 1. We can say that Network 1 perform better than Network 2 as the Network 1 is converging to lesser loss as compared to Network N1 and also the R2-score on test for Network 1 is better that Network 2. **Therefore, Network 1 is a better model.**

2. **On reducing number of layers to 3 and now comparing with different Activation functions:**

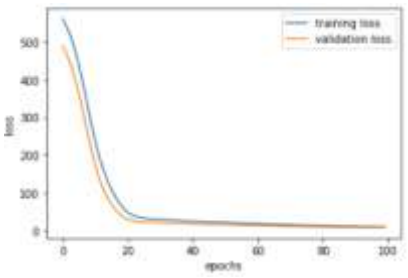
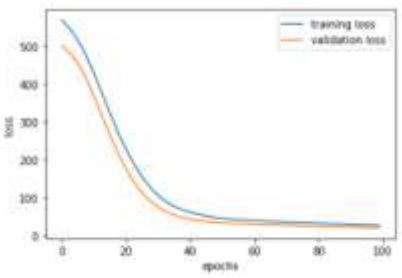
Neural Network	**Network 3**	**Network 4**
Activation function	Relu	Tanh
Layers	• layer1(input layer) = 13 • layer2(hidden layer) = 16 • layer3(Output layer) = 1	• layer1(input layer) = 13 • layer2(hidden layer) = 16 • layer3(Output layer) = 1
Train loss (MSE)	10.4129	26.0302
Test loss (MSE)	**9.4624**	27.4410
R2 score on train set	0.86439	0.67021
R2 score on test set	**0.87593**	0.64021
Epochs Vs Train/Validation Loss graph		

Table 2

Observation from Table2: On analyzing and visualizing Table 2. We can say that Network 3 perform better than Network 4 as the Network 3 is converging to lesser loss as compared to Network 4 and also the R2-score on test for Network 3 is better that Network 4.
Therefore, Network 3 is a better model.

3. **On changing number of neurons in layers and now comparing with different Activation functions:**

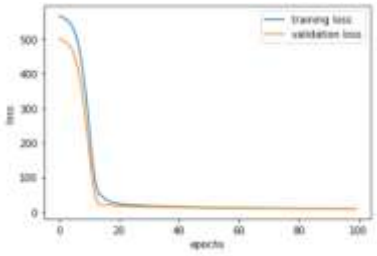
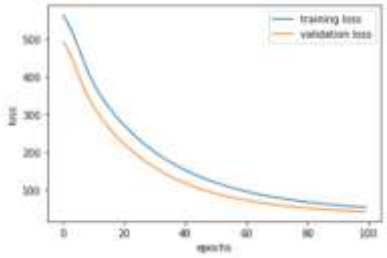
Neural Network	**Network 3**	**Network 4**
Activation function	Relu	Tanh
Layers	• layer1(input layer) = 13 • layer2(hidden layer) = 16 • layer3(hidden layer) = 8 • layer4(Output layer) = 1	• layer1(input layer) = 13 • layer2(hidden layer) = 16 • layer3(hidden layer) = 8 • layer4(Output layer) = 1
Train loss (MSE)	9.1660	54.1624
Test loss (MSE)	**7.8227**	64.0218
R2 score on train set	0.88063	0.29445
R2 score on test set	**0.89743**	0.16059
Epochs Vs Train/Validation Loss graph		

Table 3

Observation from Table3: On analyzing and visualizing Table 3. We can say that Network 5 perform better than Network 6 as the Network 5 is converging to lesser loss as compared to Network 6 and also the R2-score on test for Network 5 is better that Network 6.

Therefore, Network 5 is a better model.

- **<u>OVERALL RESULT (for Regression Task)</u> :**

  From Table 1, Table 2, Table 3. We can analyse on seeing on the basis of the convergence of loss (train as well as validation) and performance of each model on test set on the basis of R2 score.

  Note:
  - **R2 Score of Network 1 model is best as compared to other models, is marked with * in Table 1.**
  - **Also, we can visualize that Relu performs better as compared to tanh for this dataset.**