APML-
Applied Machine Learning

ASSIGNMENT 2 PRESENTATION

20-FEB-2020

Presented By:
Lim Yuan Her (P7359221)

Agenda

- **Part A: Unsupervised Learning**
    - ❑ K-Selection
    - ❑ Model Building
    - ❑ Model Evaluation

- **Part B: Deep Learning**
    - ❑ Data Exploration
    - ❑ Data Preprocessing
    - ❑ Model Training and Evaluation
    - ❑ Hyperparameter Tuning
    - ❑ Tuned Model Evaluation
    - ❑ Conclusion

Unsupervised Learning – K-Selection

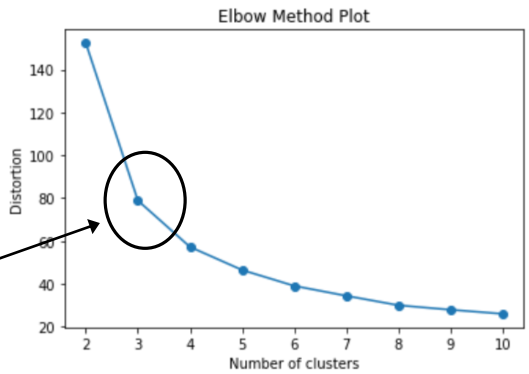
IRIS Data

- Consists of 150 records each with 4 features (sepal length, sepal width, petal length, petal width)

- Available as part of sklearn.datasets library

Selection of K

- Using Elbow Method



“elbow” point – point after which distortion/inertia start decreasing in a linear fashion

Based on the result, the optimal K value is **3**

Data Pre-processing

- Data standardized to reduce bias due to varying scale range of different features

SINGAPORE POLYTECHNIC SP

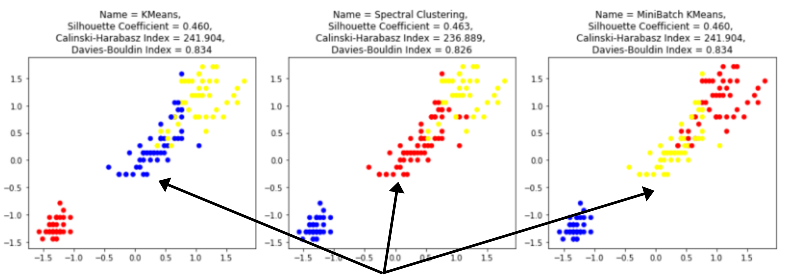
Unsupervised Learning – Model Building and Evaluation (1)

❑ Model Building

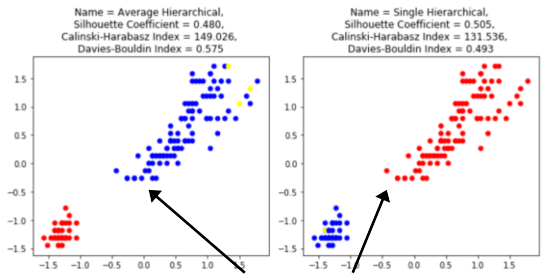
- 5 algorithms were evaluated, namely K-Means, MiniBatch K-Means, Spectral Clustering, Birch and Hierarchical Clustering (Ward/Complete/Average/Single)

- Model performance evaluated based on composite score derived from Silhouette Coefficient, Calinski-Harabasz Index and Davies-Bouldin Index

❑ Model Performance (based on standardized data)

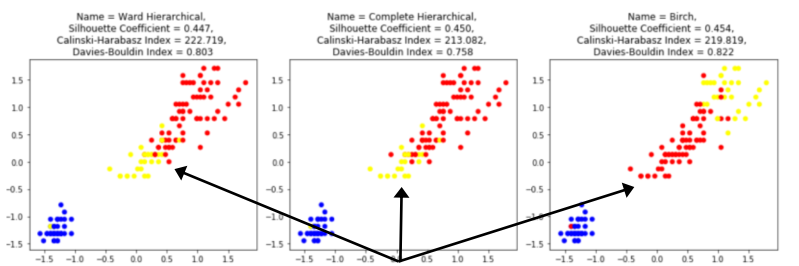


Evenly clustered data



Poorly clustered data

<u>K-Means</u> and **<u>Spectral Clustering</u>** provides optimal results with clear delineation of the different clusters identified



Uneven clustered data

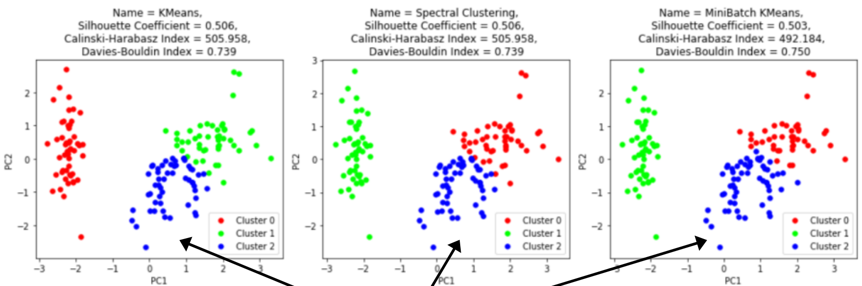
	Name	NumClusters	SilhouetteScore	CalinskiHarabaszScore	DaviesBouldinScore	Score
0	KMeans	3	0.459948	241.904402	0.833595	92.748694
2	MiniBatch KMeans	3	0.459948	241.904402	0.833595	92.748694
1	Spectral Clustering	3	0.462976	236.888875	0.825716	90.559533
5	Birch	3	0.453550	219.818769	0.821830	81.935524
3	Ward Hierarchical	3	0.446689	222.719164	0.803467	79.933840
4	Complete Hierarchical	3	0.449618	213.081710	0.758358	72.654880
6	Average Hierarchical	3	0.480267	149.025799	0.575269	41.173226
7	Single Hierarchical	3	0.504646	131.535896	0.492925	32.719890

Unsupervised Learning – Model Building and Evaluation (2)

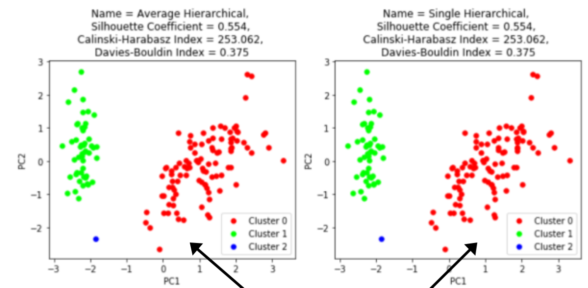
❑ Model Performance (based on 2-component PCA data)

- Principal Component Analysis (PCA) was used to reduce data to a 2-dimensional representation to facilitate visualization

- Clustering performed on PCA data using same algorithms as that for standardized data

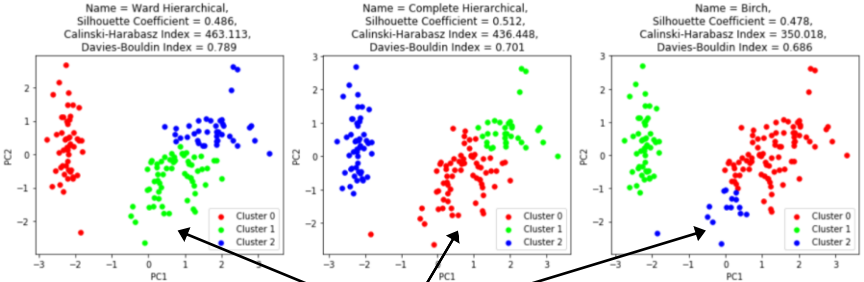


Evenly clustered data



Poorly clustered data

	Name	NumClusters	SilhouetteScore	CalinskiHarabaszScore	DaviesBouldinScore	Score
0	KMeans_PCA	3	0.506153	505.957631	0.738682	189.170470
1	Spectral Clustering_PCA	3	0.506153	505.957631	0.738682	189.170470
2	MiniBatch KMeans_PCA	3	0.503489	492.184439	0.750328	185.938119
3	Ward Hierarchical_PCA	3	0.485768	463.112725	0.788649	177.418550
4	Complete Hierarchical_PCA	3	0.511822	436.447626	0.701087	156.611211
5	Birch_PCA	3	0.478296	350.017579	0.685556	114.770305
6	Average Hierarchical_PCA	3	0.553808	253.061764	0.374795	52.526551
7	Single Hierarchical_PCA	3	0.553808	253.061764	0.374795	52.526551



Uneven clustered data

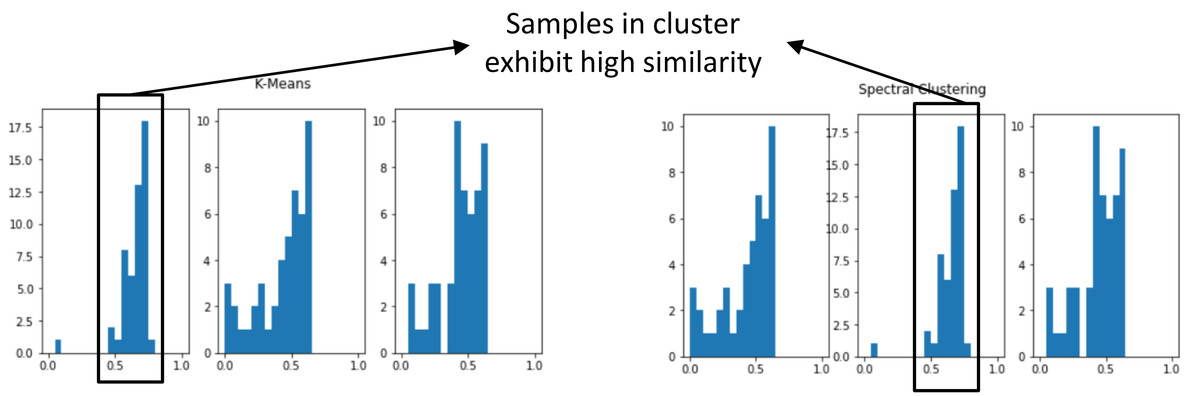
<u>K-Means</u> and **<u>Spectral Clustering</u>** provides optimal results with clear delineation of the different clusters identified

Unsupervised Learning – Model Building and Evaluation (3)

- Based on the results, <u>**K-Means**</u> and <u>**Spectral Clustering**</u> with 2-component PCA data provided the optimal results with clear delineation of the different clusters identified

	Name	NumClusters	SilhouetteScore	CalinskiHarabaszScore	DaviesBouldinScore	Score
0	KMeans_PCA	3	0.506153	505.957631	0.738682	189.170470
1	Spectral Clustering_PCA	3	0.506153	505.957631	0.738682	189.170470
2	MiniBatch KMeans_PCA	3	0.503489	492.184439	0.750328	185.938119

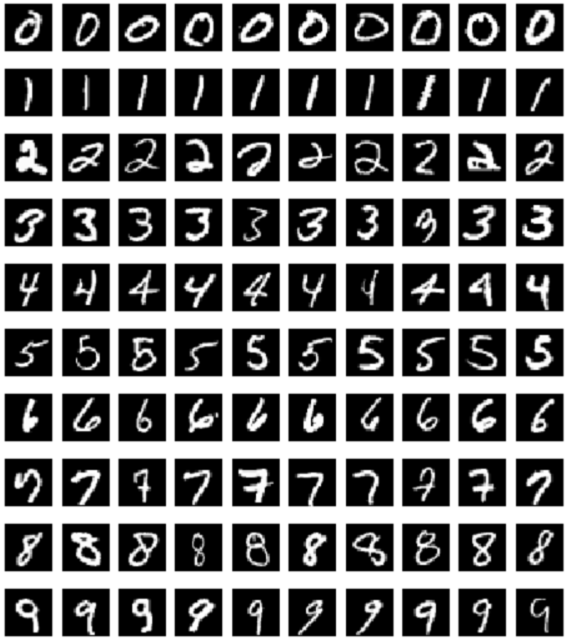
- From the silhouette profiles for <u>**K-Means**</u> and <u>**Spectral Clustering**</u> models, only one of the clustered class data exhibit the best performance.



Deep Learning – Data Exploration

- **MNist Data**
  - Consists of 60,000 images of handwritten digits from 0 to 9

- **Data Exploration**
  - 10 classes (0 to 9)



Deep Learning – Data Preprocessing

- **Image Data Format**
  - Channel ordering is Channels Last as Tensorflow default

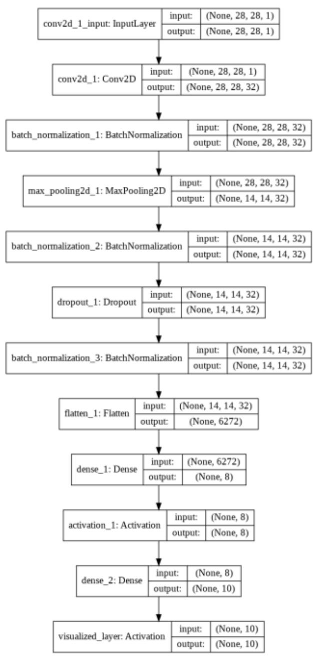
- **Data Transformation**
  - Pixel Value Normalization
    - Pixels are represented in the range 0 to 255, but for faster convergence, data values are normalized to the range 0 to 1.

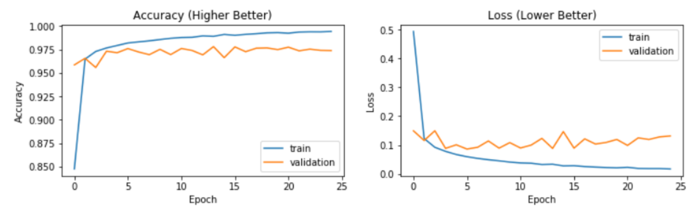
  - Convert class vector to binary class matrices

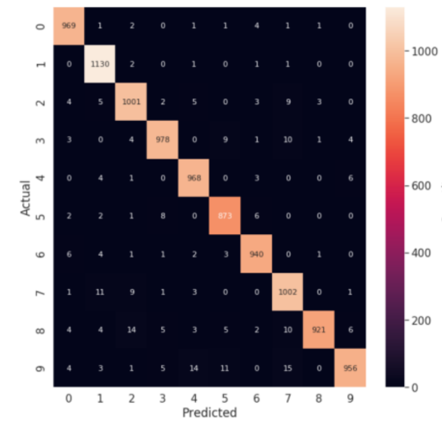
Deep Learning– Model Building and Evaluation

❑ Model Training

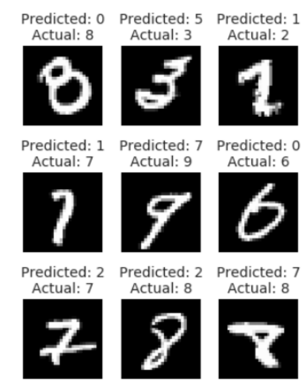
- A base CNN was evaluated with the following architecture:







	precision	recall	f1-score	support
0	0.98	0.99	0.98	980
1	0.97	1.00	0.98	1135
2	0.97	0.97	0.97	1032
3	0.98	0.97	0.97	1010
4	0.97	0.99	0.98	982
5	0.97	0.98	0.97	892
6	0.98	0.98	0.98	958
7	0.96	0.97	0.97	1028
8	0.99	0.95	0.97	974
9	0.98	0.95	0.96	1009
accuracy			0.97	10000
macro avg	0.97	0.97	0.97	10000
weighted avg	0.97	0.97	0.97	10000



❑ Model Evaluation

- Model was evaluated based on loss and validation accuracy metrics

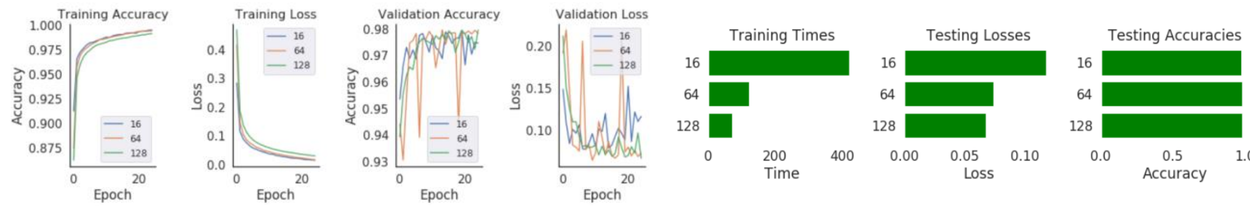
```
Test loss: 13.16
Test accuracy: 97.38
```

Deep Learning – Hyperparameter Tuning (1)

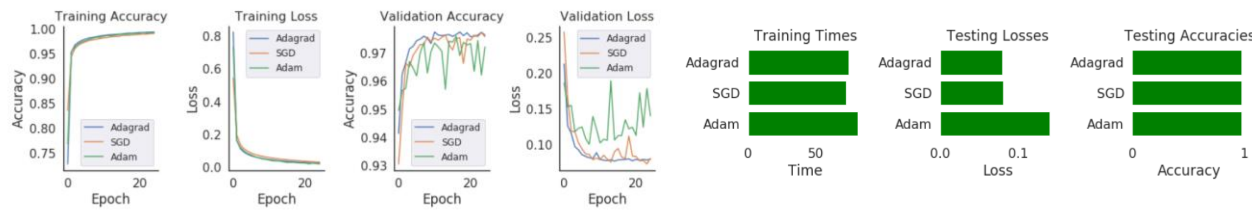
❑ Model Tuning

- Hyperparameter tuning was performed for batch size, optimizer, activation function, # of neurons, dropout parameters
- # of epochs kept constant at 25

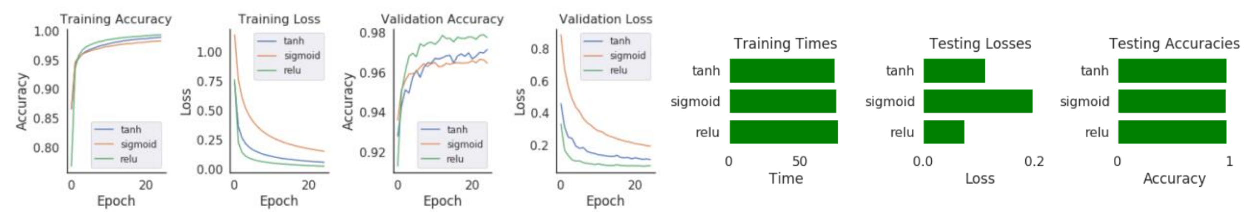
Batch Size
128



Optimizer
Adagrad



Activation Function
Relu

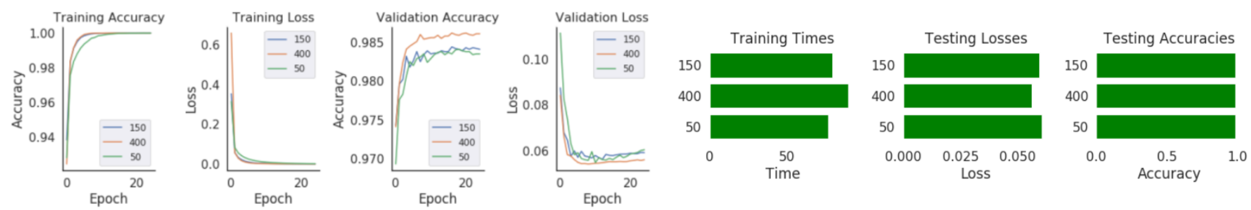


Deep Learning – Hyperparameter Tuning (2)

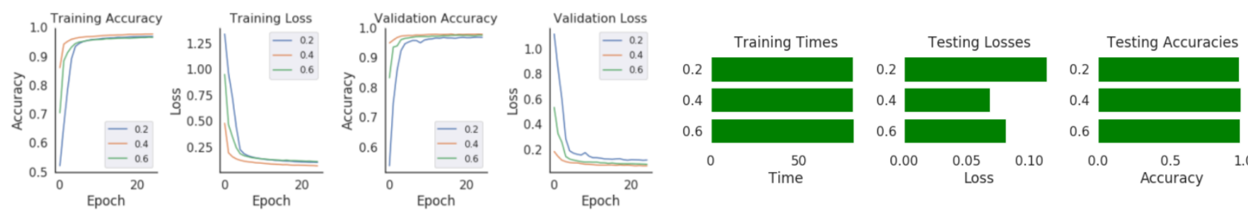
❑ Model Tuning

- Evaluation criteria based on testing loss/accuracy

# of Neurons
400



Dropout
0.4



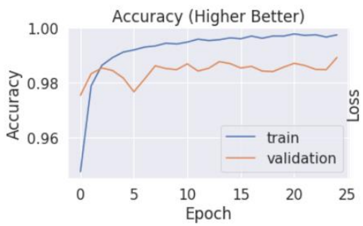
Based on the hyperparameter tuning results optimal parameters are:

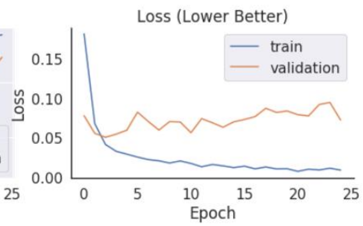
1. Batch Size = **128**
2. Optimizer = **Adagrad**
3. Activation Function = **Relu**
4. # of Neurons = **400**
5. Dropout = **0.4**

Deep Learning – Tuned Model Evaluation (1)

❑ Tuned Model Performance

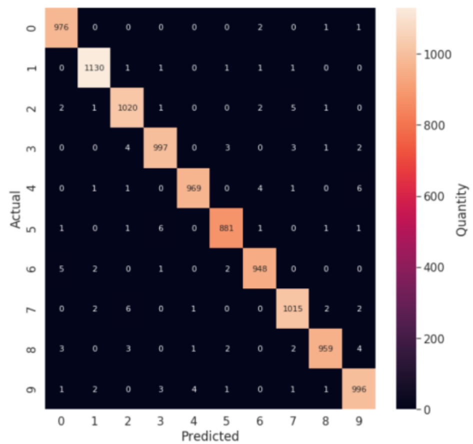
- Tuned model evaluated based on classification metrics and test loss/accuracy
- Tuned model achieved improved accuracy and loss metrics as compared to base model

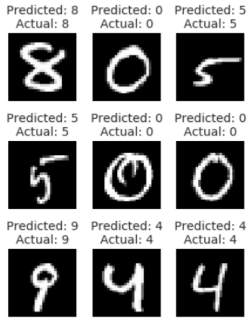




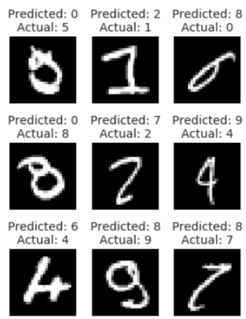
	precision	recall	f1-score	support
0	0.99	1.00	0.99	980
1	0.99	1.00	0.99	1135
2	0.98	0.99	0.99	1032
3	0.99	0.99	0.99	1010
4	0.99	0.99	0.99	982
5	0.99	0.99	0.99	892
6	0.99	0.99	0.99	958
7	0.99	0.99	0.99	1028
8	0.99	0.98	0.99	974
9	0.98	0.99	0.99	1009
accuracy			0.99	10000
macro avg	0.99	0.99	0.99	10000
weighted avg	0.99	0.99	0.99	10000

```
Test loss: 7.21
Test accuracy: 98.91
```





Correctly Classified

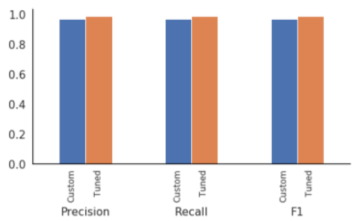


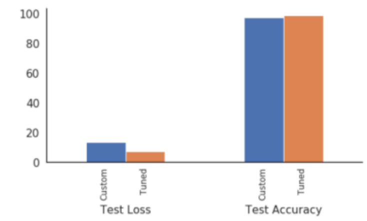
Mis-Classified

Deep Learning – Tuned Model Evaluation (2)

. Tuned Model Performance

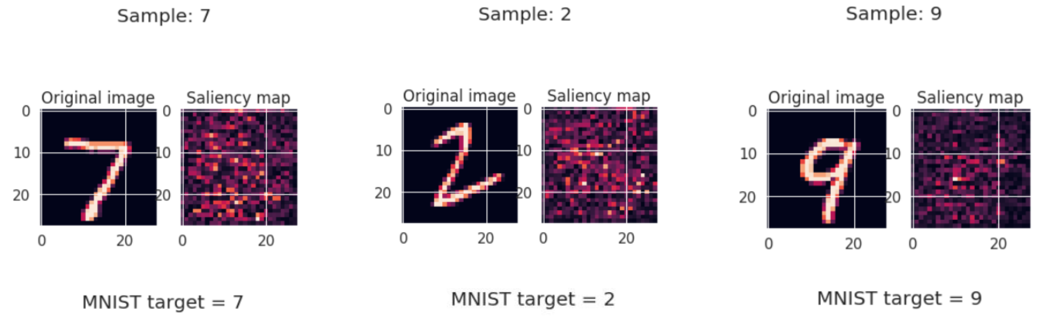
- Tuned model achieved improved classification metrics (Precision/ Recall/ F1) and validation loss/accuracy as compared to base model





- Tuned model Visualization

Saliency maps



Grad-CAM Class Activation Maps