

CS 584: Machine Learning

Face Recognition

Phase 1:

- I implemented the KNeighbors classifier and SVC classifier for the given dataset of Bush and Williams.
- I was able to obtain the following results for Bush and Williams dataset as per the given settings for both KNeighbors classifier and SVC classifier.

Phase 1 Analysis:

Bush Dataset:

KNeighbors classifier-

- We can see that the KNeighbors classifier gives us best result i.e. F1 score of 0.1394 for 1 neighbor. The reason being is that as the number of 'k' neighbors are increasing so does the precision of the classifier. The increase in 'k' neighbors also result in gradual decrease in recall of the classifier.
- Ideally, we are looking for high precision and high recall which gives us correct results. However, neighbors >1 returns very few results, but most of its predicted labels are correct when compared to trained labels because of high precision and low recall.

KNeighbors Classifier	Parameters	Mean F1
	n_neighbors=1	0.139423483
	n_neighbors=3	0.094405671
	n_neighbors=5	0.035327385
Best Result	n_neighbors=1	0.139423483

KNeighbors Classifier – Bush (Phase 1)

SVC classifier-

- SVC classifier for the Bush dataset returns best result (i.e. F1 score 0.6587) when we have penalty parameter C as 1000, Kernel type as 'rbf' and Kernel Coefficient Gamma as 0.0001. As provided in bush.pdf of phase 1, we have zero mean F1 score for linear kernel with C=0.001, poly kernel with degree 1, 2, 3 and 5 and C=1, rbf kernel with gamma 0.0001, 0.001 and C as 1.
- The one reason we got best result on 'rbf' kernel is that it depends on what seed value (random state) we specified in k-fold cross validation.

SVC Classifier	Parameters	Mean F1
Best Result	C=1000, Kernel=rbf, Gamma = 0.0001	0.658792842

SVC Classifier – Bush (Phase 1)

Williams Dataset:

KNeighbors classifier-

- We can see that the KNeighbors classifier gives us best result i.e. F1 score of 0.1267 for 1 neighbor. As the number of ‘k’ neighbors are increasing, precision and recall becomes zero. The reason behind the precision and recall becoming zero and in turn mean F1 score, is that we have very less target variables to predict for Williams dataset i.e. 52 out of 13233.
- Ideally, we are looking for high precision and high recall which gives us correct results. However, neighbors >1 returns zero mean F1, because of zero precision and recall.

KNeighbors Classifier	Parameters	Mean F1
	n_neighbors=1	0.126754083
	n_neighbors=3	0
	n_neighbors=5	0
Best Result	n_neighbors=1	0.126754083

KNeighbors Classifier – Williams (Phase 1)

SVC classifier-

- SVC classifier for the Williams dataset returns best result (i.e. F1 score 0.5430) when we have penalty parameter C as 150, Kernel type as ‘rbf’ and Kernel Coefficient Gamma as 0.0001. As provided in bush.pdf of phase 1, we have zero mean F1 score for linear kernel with C=0.001, poly kernel with degree 1, 2, 3 and 5 and C=1, rbf kernel with gamma 0.0001, 0.001 and C as 1.
- The one reason we got best result on ‘rbf’ kernel is that it depends on what seed value (random state) we specified in k-fold cross validation.

SVC Classifier	Parameters	Mean F1
Best Result	C=150, Kernel=rbf, Gamma = 0.0001	0.543022774

SVC Classifier – Williams (Phase 1)

Phase 2:

- Implementation of PCA (Principal Component Analysis) to speed up the machine learning algorithm such as KNeighbors classifier and SVC classifier for the given dataset of Bush and Williams.
- Algorithms implemented in phase 1 (i.e. KNeighbors classifier and SVC classifier) are too slow which we saw based on their performance in terms of their F1 score computed. Hence, we are using PCA to speed up of the fitting of these algorithms. PCA reduces the input dimensions which helped KNeighbors classifier and SVC classifier to be fast due to dimensions reduced from 4096 to some dimensions (n_components in PCA).
- With PCA, I was able to get the better results as compared to phase 1 for KNeighbors classifier and SVC classifier except for SVC model of Bush dataset.

Phase 2 Analysis:

Bush Dataset:

KNeighbors classifier-

- KNeighbors classifier with PCA gives us best result i.e. F1 score of 0.1593 for 1 neighbor and with 77 PCA components (dimensions instead of 4096). After few experiments with different components of PCA and neighbors > 1, I found the F1 score to be low as compared to the one which I found above with PCA parameters to be n_components=77, whiten=False, svd_solver = full.
- Ideally, we are looking for high precision and high recall which gives us correct results. However, neighbors >1 returns zero mean F1, because of zero precision and recall.

KNeighbors Classifier	KNeighbors Classifier Parameters	PCA Parameters	Mean F1
Best Result	n_neighbors=1	n_components=77, whiten=False, svd_solver = full	0.159391963

KNeighbors Classifier – Bush (Phase 2)

SVC Classifier-

- SVC classifier for the Bush dataset returns best result (i.e. F1 score 0.6510) when we have penalty parameter C as 1000, Kernel type as 'rbf' and Kernel Coefficient Gamma as 0.0001.
- The one reason we got best result on 'rbf' kernel is that it depends on what seed value (random state) we specified in k-fold cross validation.

SVC Classifier	SVC Classifier Parameters	PCA Parameters	Mean F1
Best Result	C=1000, Gamma=0.0001, Kernel= 'rbf'	n_components=1752	0.651077213

SVC Classifier – Bush (Phase 2)

Williams Dataset:

KNeighbors classifier-

- We can see that the KNeighbors classifier gives us best result i.e. F1 score of 0.2054 for 1 neighbor and with PCA components as 250, whiten=False and svd_solver='full'. After few experiments, as the number of 'k' neighbors are increasing, precision and recall becomes zero. The reason behind the precision and recall becoming zero and in turn mean F1 score, is that we have very less target variables to predict for Williams dataset i.e. 52 out of 13233.
- Ideally, we are looking for high precision and high recall which gives us correct results. However, neighbors >1 returns zero mean F1, because of zero precision and recall.

KNeighbors Classifier	KNeighbors Classifier Parameters	PCA Parameters	Mean F1
Best Result	n_neighbors=1	n_components=250, whiten=False, svd_solver = full	0.205467373

KNeighbors Classifier – Williams (Phase 2)

SVC classifier-

- SVC classifier for the Williams dataset returns best result (i.e. F1 score 0.5625) when we have penalty parameter C as 150, Kernel type as 'rbf' and Kernel Coefficient Gamma as 0.0001. SVC with PCA also considers dimensions to be 1752, svd_solver = 'auto' and random state as 5 to give the best result.
- The one reason we got best result on 'rbf' kernel is that it depends on what seed value (random state) we specified in k-fold cross validation as well.

SVC Classifier	SVC Classifier Parameters	PCA Parameters	Mean F1
Best Result	C=150, Gamma=0.0001, Kernel= 'rbf'	n_components=1752, whiten=False, svd_solver= 'auto', random_state= 5	0.562509953

SVC Classifier – Williams (Phase 2)

Comparison of Phase 1 and Phase 2-

We can see the improvement in F1 score in phase 2 (except SVC for Bush dataset) as compared to that of phase 1 F1 score. One reason I found which gave such improvement in F1 results was to have different random states specified in K-fold cross validation. Other significant improvement in phase 2 is due to dimensionality reduction which helped feature reduction and reduced time for training the model. We could also see that the Fit time and Score was significantly reduced for phase 2 as a result of dimensionality reduction.

Phase 3:

- Phase 3 of the face recognition develops deep learning model and trains it on Bush and Williams datasets.
- It differs completely than phase 1 and 2 as these phases (i.e. 1 and 2) deals with using KNeighbors and SVC Classifiers along with PCA.
- The dataset given is reshaped to 64*64 image format with grayscale coloring.
- I used 2D CNN and MaxPooling layer while designing deep learning model with single output layer of sigmoid activation function.
- F1 score for both Bush and Williams dataset is pickled and submitted in phase 3 along with models of format '.h5'

Phase 3 Analysis:

Bush Dataset:

- The BUSH model receives an input of grayscale image with input shape (64,64,1) i.e. width 64, height 64 and grayscale color.
- Convolution layer computes the output of neurons that are connected to the input, each computing a dot product between their weights (3,3) and filters resulted in output parameters of 320 (32 Filters*3*3 + 32).
- Pooling layer performs the down sampling by factor of 2. Fully connected layer is dense layer with 128 units which computes the class scores and is connected to output layer of sigmoid function.
- The higher F1 score is obtained using 10 epochs for BUSH dataset after carefully choosing it as too high or too low epoch would result in overfitting or underfitting.

Bush Dataset	Train F1 Score	Test F1 Score
	1.0	0.7248322147

Layers	Description
2D Convolution Layer	<ul style="list-style-type: none"> 32 Filters Kernel Size (3,3)- Height and Width of 2D convolution window Input Shape (64,64,1) Activation function- RELU (Rectified Linear Unit)
MaxPooling Layer	<ul style="list-style-type: none"> Pool Size (2,2) – Downscaling the input
Flattening Layer	<ul style="list-style-type: none"> Flattens the output of MaxPooling layer to 1D to be used by Fully Connected Layer
Fully Connected Layer (Dense)	<ul style="list-style-type: none"> 128 Units- dimensionality of output space Activation function- RELU (Rectified Linear Unit)
Output Layer	<ul style="list-style-type: none"> Output layer having sigmoid activation function

Bush Dataset Layer Description (Phase 3)

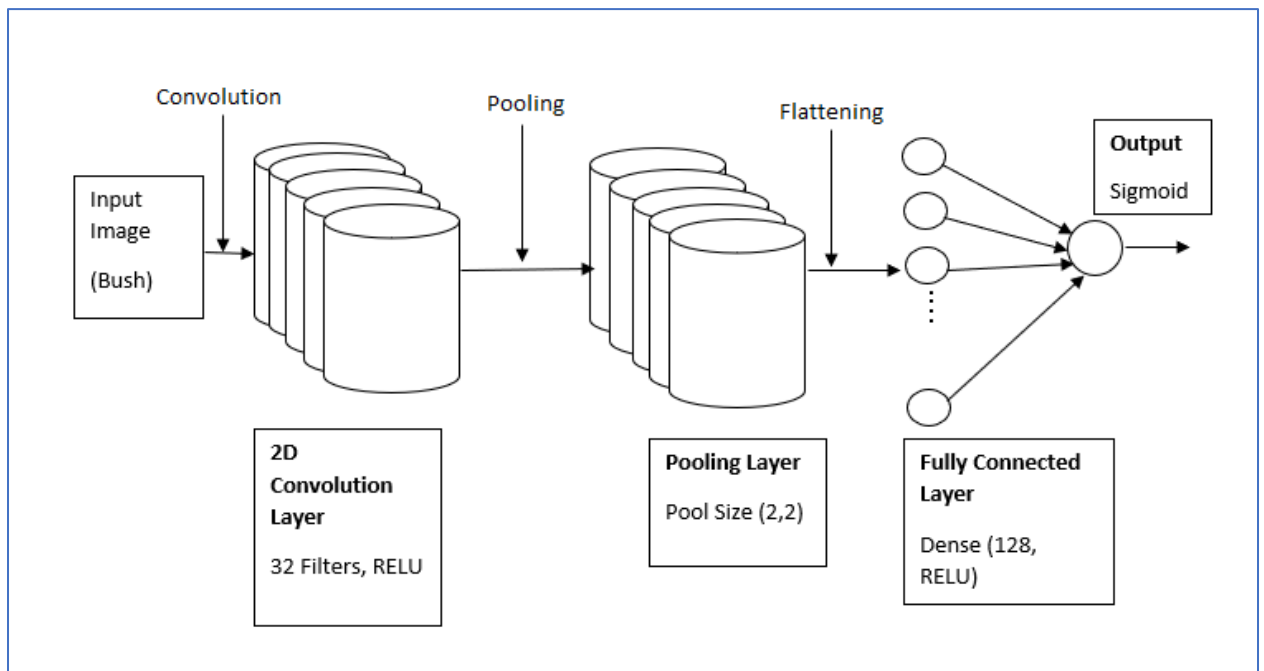


Figure.1 CNN Layer Architecture for Bush Model

Williams Dataset:

- The WILLIAMS model receives an input of grayscale image with input shape (64,64,1) i.e. width 64, height 64 and grayscale color.
- Convolution layer computes the output of neurons that are connected to the input, each computing a dot product between their weights (3,3) and filters resulted in output parameters of 80 (8 Filters*3*3 + 8). Pooling layer performs the down sampling by factor of 2.

- Fully connected layer is dense layer with 16 units which computes the class scores and is connected to output layer of sigmoid function.
- The higher F1 score is obtained using 10 epochs for WILLIAMS dataset after carefully choosing it as too high or too low epoch would result in overfitting or underfitting.

Williams Dataset	Train F1 Score	Test F1 Score
	0.98591549	0.521739130

Layers	Description
2D Convolution Layer	<ul style="list-style-type: none"> • 8 Filters • Kernel Size (3,3)- Height and Width of 2D convolution window • Input Shape (64,64,1) • Activation function- RELU (Rectified Linear Unit)
MaxPooling Layer	<ul style="list-style-type: none"> • Pool Size (2,2) – Downscaling the input
Flattening Layer	<ul style="list-style-type: none"> • Flattens the output of MaxPooling layer to 1D to be used by Fully Connected Layer
Fully Connected Layer (Dense)	<ul style="list-style-type: none"> • 16 Units- dimensionality of output space • Activation function- RELU (Rectified Linear Unit)
Output Layer	<ul style="list-style-type: none"> • Output layer having sigmoid activation function

Williams Dataset Layer Description (Phase 3)

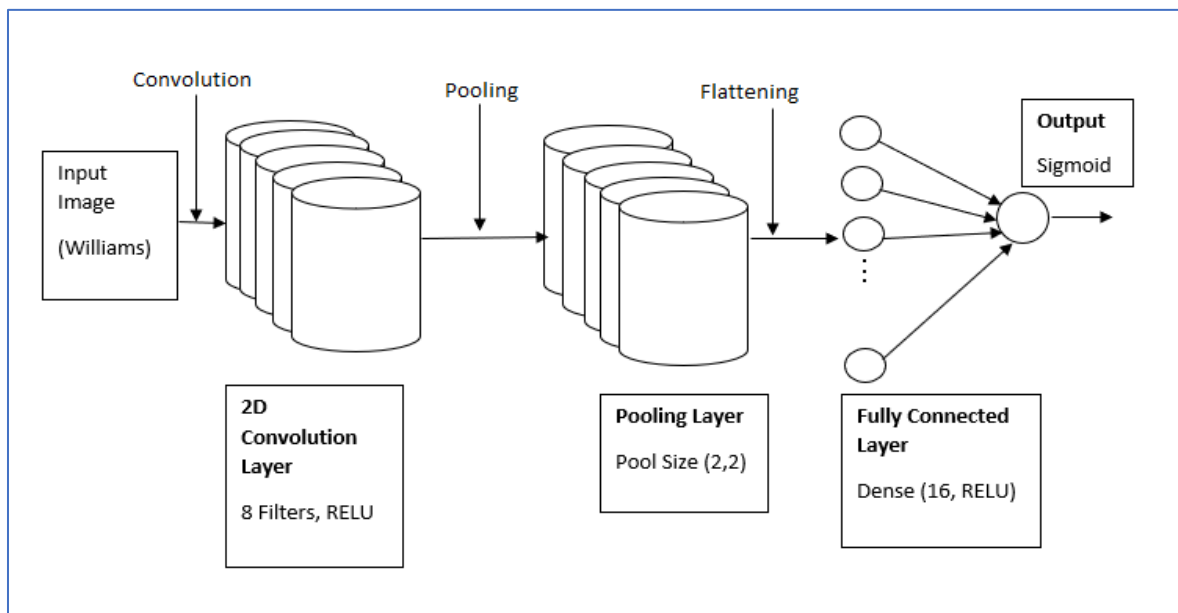


Figure.2 CNN Layer Architecture for Williams Model

Phase 4:

- Phase 4 of the face recognition deals with transfer learning which is improvement of learning in new task through transfer of knowledge from related task that has been already learned.
- I used the datasets of Cats and Dogs from following URL to pre-train the model.
https://drive.google.com/drive/folders/1XaFM8BJFligrqeQdE-5Id0V_SubJAZe
- Since all the images are in RGB format, scaling was performed on them to transform them to grayscale images with the shape of 64*64.
- I used 2D CNN and MaxPooling layer while designing deep learning model with single output layer of sigmoid activation function to pre-train it on Cats and Dogs dataset.
- The pretrained model i.e. initialized-model.keras was saved to drive. To train it onto the given dataset (Bush and Williams), I split the dataset into train and test splits.
- Loaded model is then trained on train split and recorded its F1 score predicting on test split.
- For Bush and Williams datasets, two different models were pre-trained. However, the best result for both of them came with the similar parameter settings used for pre-training.
- Phase 4 analysis is same for Bush and Williams dataset as two sets use two pre-trained models but with same parameter settings.

Phase 4 Analysis:

Bush Dataset:

- Cats and Dogs dataset is used for pre-training the model which is rescaled to be in grayscale format of shape 64*64 from RGB format.
- Convolution layer computes the output of neurons that are connected to the input, each computing a dot product between their weights (3,3) and filters resulted in output parameters of 80 (8 Filters*3*3 + 8).
- Pooling layer performs the down sampling by factor of 2. Fully connected layer is dense layer with 16 units which computes the class scores and is connected to output layer of sigmoid activation function of model.
- The model built is pretrained on training set of the Cats and Dogs with 2814/32 (no of images in training set / batch size using in scaling the training set) steps per epochs.
- The pretrained model is saved as initialized-model and loaded it to train it on training set of Bush dataset.
- The Bush dataset is split in 1/.3 and its train set is used by loaded model for training and F1 score is computed after predicted on test set.

Bush Dataset	Train F1 Score	Test F1 Score
	0.987377772791	0.7217125382

Williams Dataset:

- Follow all the steps of Bush dataset except the last step of splitting the Williams dataset into train and test split. And apply the pre-trained model on train split generated for Williams dataset.

Williams Dataset	Train F1 Score	Test F1 Score
	1.0	0.56

Layers	Description
2D Convolution Layer	<ul style="list-style-type: none">8 FiltersKernel Size (3,3)- Height and Width of 2D convolution windowInput Shape (64,64,1)Activation function- RELU (Rectified Linear Unit)
MaxPooling Layer	<ul style="list-style-type: none">Pool Size (2,2) – Downscaling the input
Flattening Layer	<ul style="list-style-type: none">Flattens the output of MaxPooling layer to 1D to be used by Fully Connected Layer
Fully Connected Layer (Dense)	<ul style="list-style-type: none">16 Units- dimensionality of output spaceActivation function- RELU (Rectified Linear Unit)
Output Layer	<ul style="list-style-type: none">Output layer having sigmoid activation function

Bush/Williams Dataset Layer Description (Phase 4)

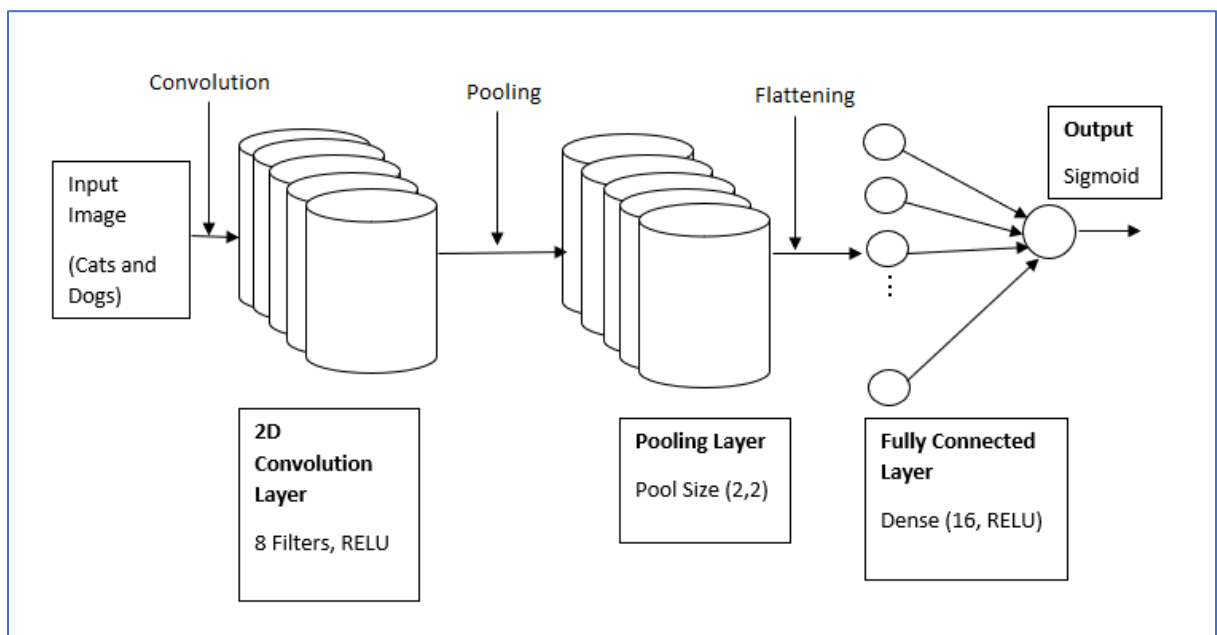


Figure.3 CNN Layer Architecture for Bush/Williams Model (Phase 4)

Comparison of Phase 3 and Phase 4-

I could see that the phase 4 results are same or better than phase 4 results considering test F1 score. Williams model in phase 4 gave us higher F1 score when used pre-trained model while Bush model gave us similar results which would be improved if tried training multiple times on datasets. We could also see that use of pre-trained model on training new data gave us better F1 score if trained multiple times with no change in random state. Random state with different seed values also affect the F1 score.