Name: Badr AlKhamissi

ID: 900141572

CSCE 4930

Practical Machine Deep Learning

Assignment #4

PIMA using PCA

Description

In this assignment I implemented a classifier to differentiate between diabetes vs not diabetes using the PIMA dataset. I then implemented a PCA and used it with my best classifier to see whether it will get me better results or not.

Friday, April 21, 2017

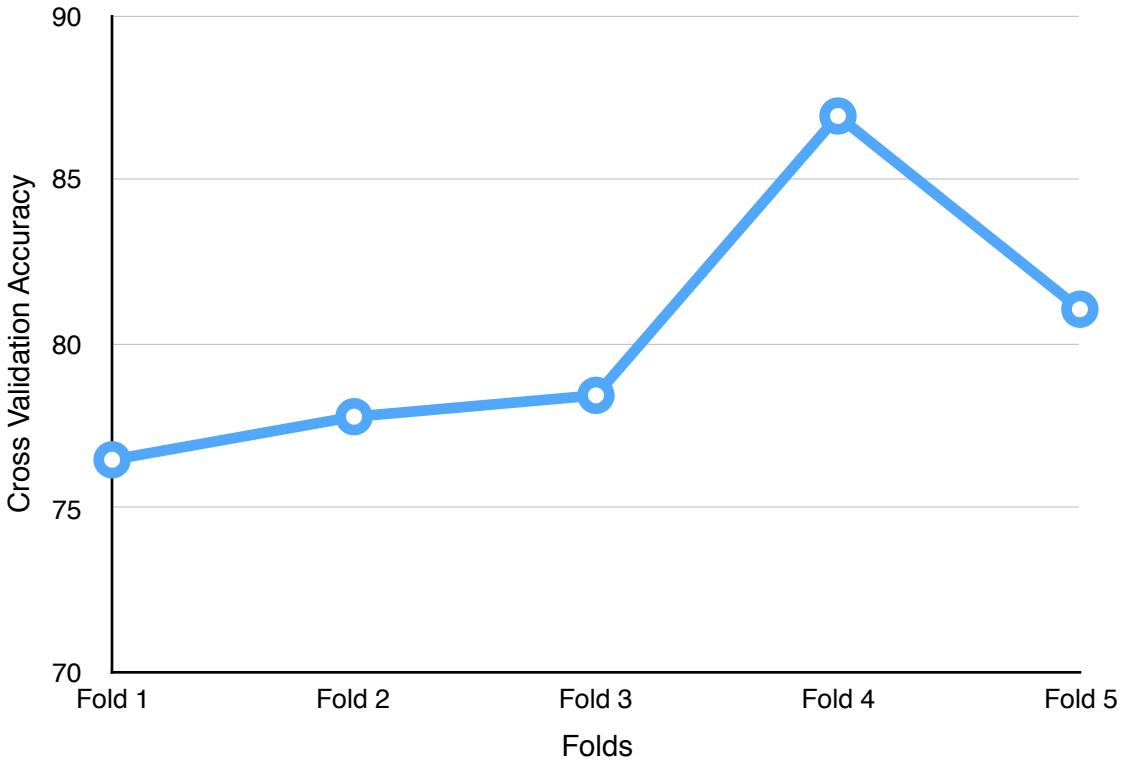
Cross Validation Accuracy without PCA

The best classifier I trained without PCA got me a 80.13% average cross validation accuracy. It's a fully connected model with the following architecture:

$$8(\text{input}) \rightarrow 4 \rightarrow 6 \rightarrow 7 \rightarrow 8 \rightarrow 1(\text{output})$$

all having an ELU activation function, with a sigmoid in the end for binary classification. I used *adam* as my optimizer and binary cross entropy as my loss function.

I preprocessed the data in the beginning by zero-centering the data then normalizing it using the training set. This improved the accuracy significantly

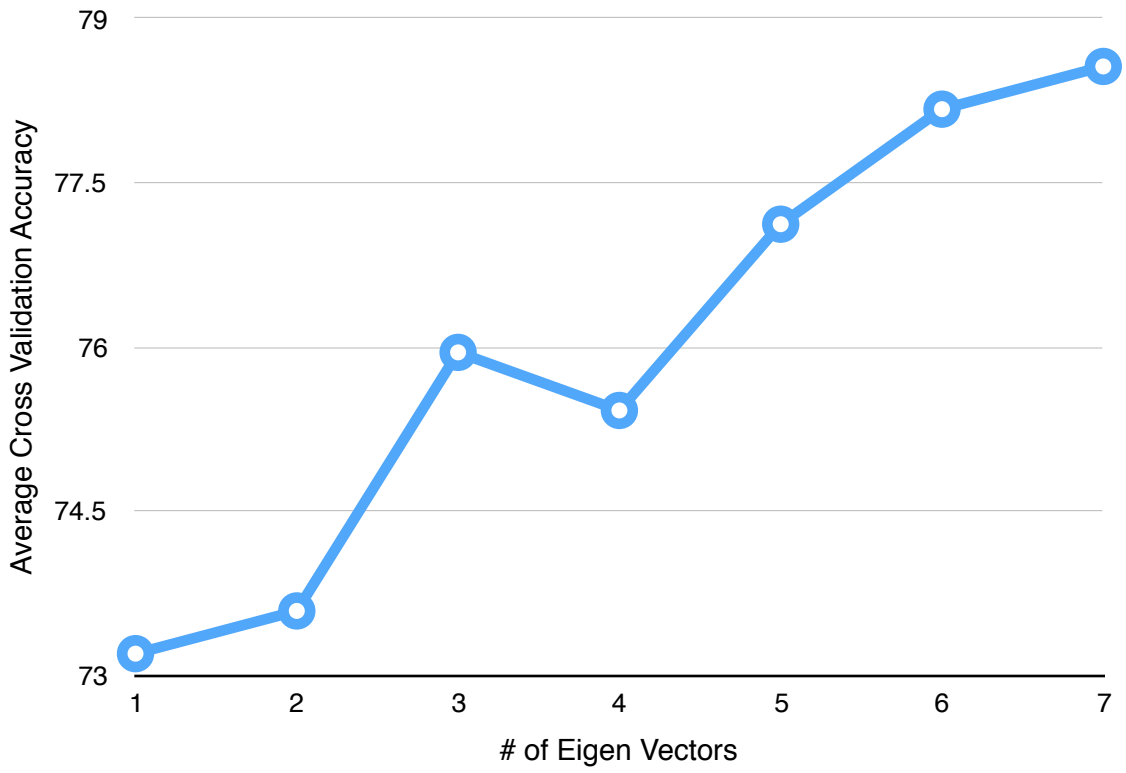


Average Cross Validation Accuracy with PCA

I chose the same architecture for my classifier to train all the reduced datasets obtained by Principal Component Analysis.

8(input) —> 4 —> 6 —> 7 —> 8 —> 1(output)

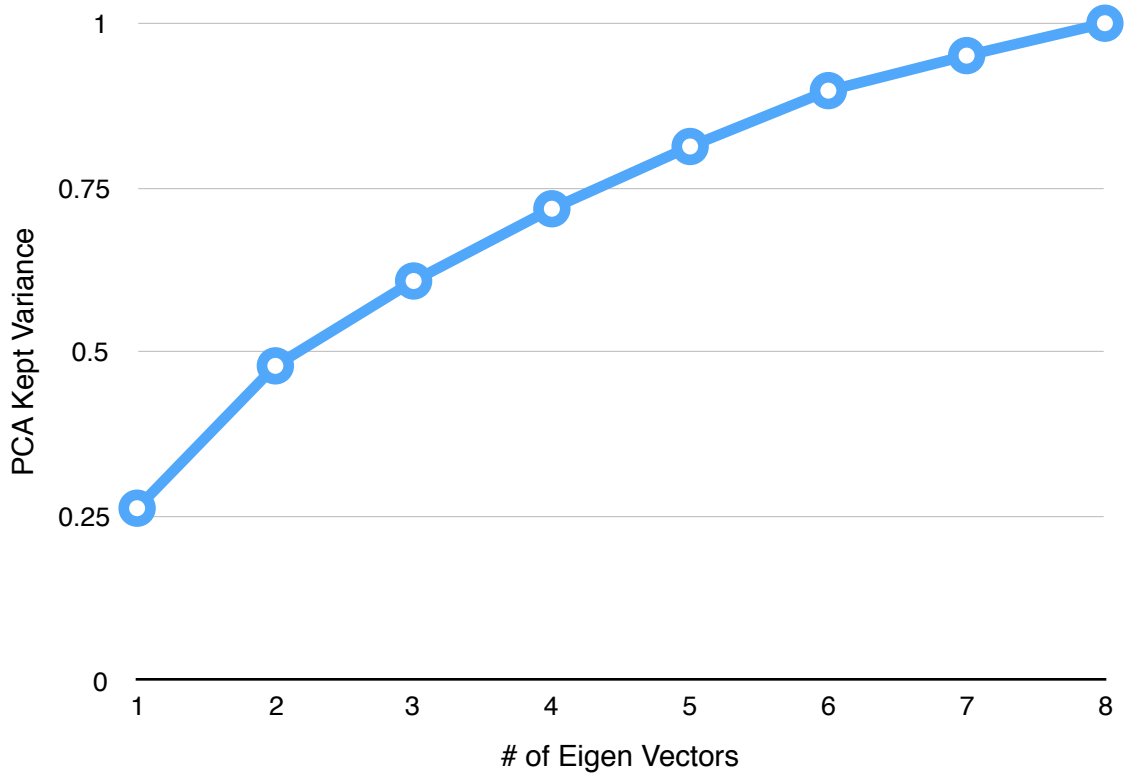
all hidden layers having an ELU activation function, with a sigmoid in the end for binary classification. I used *adam* as my optimizer and binary cross entropy as my loss function.



# of Eigen Vectors / ACCR

1	2	3	4	5	6	7
0.732	0.736	0.759	0.754	0.771	0.782	0.786

PCA Kept-Variance vs Eigen Vectors



# of Eigen Vectors / PCA Kept Variance

1	2	3	4	5	6	7	8
0.262	0.479	0.608	0.718	0.812	0.897	0.951	1.000

Comments

The PCA didn't enhance my classification accuracy, although reducing the data into 6 or 7 features got me a more or less similar accuracy, 78.6% compared to 80.13% with 8 features using the same network.

It's important to note also that the pre-processing step I did using the full features: subtracting the mean and dividing by the standard deviation improved the accuracy significantly, without this step I got an ACCR of 71.76% using the full features, which is a much less accuracy than the ones I got using PCA.