Face Recognition

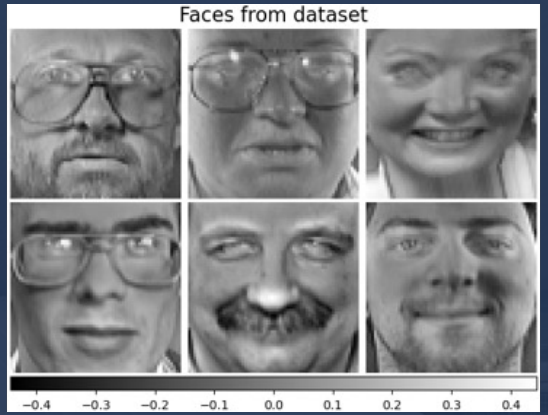
STAT 542 Midterm Project
Group 11

Outline

- Dataset Introduction
- PCA
- Models
- Results Comparison
- Difficulties

Dataset Introduction

The Olivetti Face Dataset



Background

- April 1992- April 1994
- AT&T Laboratories Cambridge

Description

- 10 images for each 40 distinct subjects
- 64x64 pixels
- 256 gray levels and stored as 8-bit integers, which is then converted to floating point values on the interval [0,1]

*Note that 80-20 train test split is used for this project.

Principal Components Analysis (PCA)

PCA

Methodology

- Each Principal Component z_i is a linear combination of the original variables $x_1, x_2, \ldots, x_m$ with weights given by each column u_i of matrix U:

$$z_i = u_{1i}x_1 + u_{2i}x_2 + \cdots + u_{mi}x_m,$$

while U is the rotation matrix and Z is a version of the data rotated in such a way that the resulting principal components are orthogonal.

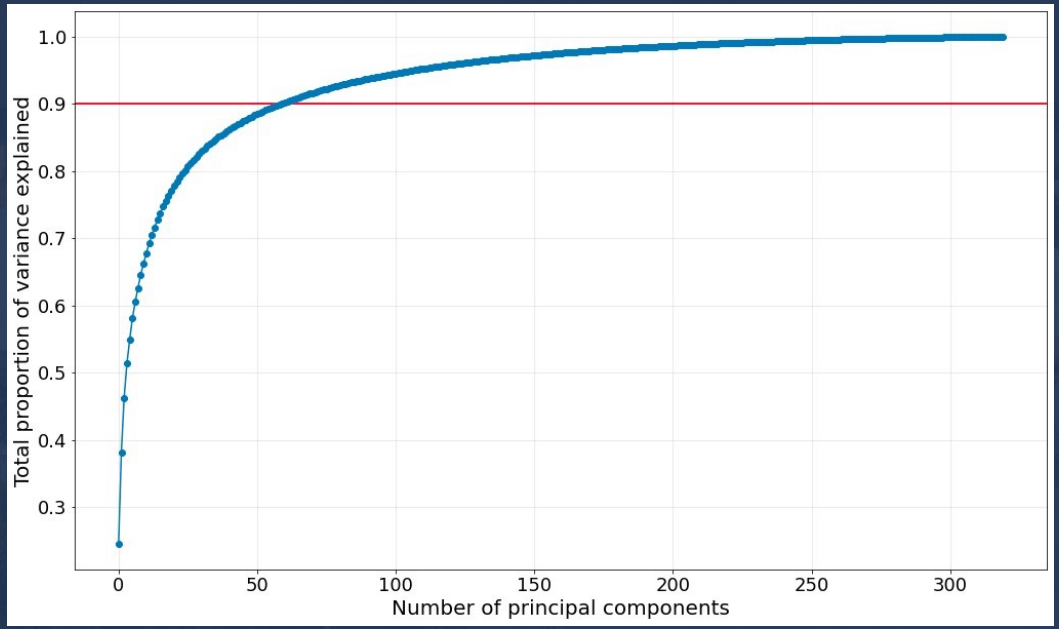
PCA

Methodology

- Creates new variables in decreasing order of the total variation contained which are orthogonal to the previous ones
- A dimensionality-reduction method that is often used on large data sets containing a lot of parameters per observation

PCA

- First 60 principal components explain over 90% of the total variation



Model 1-
Logistic Regression

Logistic Regression

Methodology

- For each class k,

$$P(Y = k|X) = \frac{e^{\beta_{0k}+\beta_{1k}X_1+\cdots+\beta_{pk}X_p}}{\sum_{l=1}^{K} e^{\beta_{0l}+\beta_{1l}X_1+\cdots+\beta_{pl}X_p}},$$

and we maximize the log-likelihood function of β to obtain the conditional probability of each class.

Logistic Regression

Results

	Accuracy	Time Consumption
With PCA	0.975	0.093 sec
Without PCA	0.975	5.121 sec

Model 2-
Linear Discriminant Analysis (LDA)

LDA

Methodology

- We assign x to the class with the largest *discriminant score,* and the *discriminant score* is denoted as

$$\delta(x) = x \cdot \frac{\mu_k}{\sigma^2} - \frac{\mu_k^2}{2\sigma^2} \log(\pi_k)$$

- A technique used to find a linear combination of features that best separates the classes in a dataset

LDA

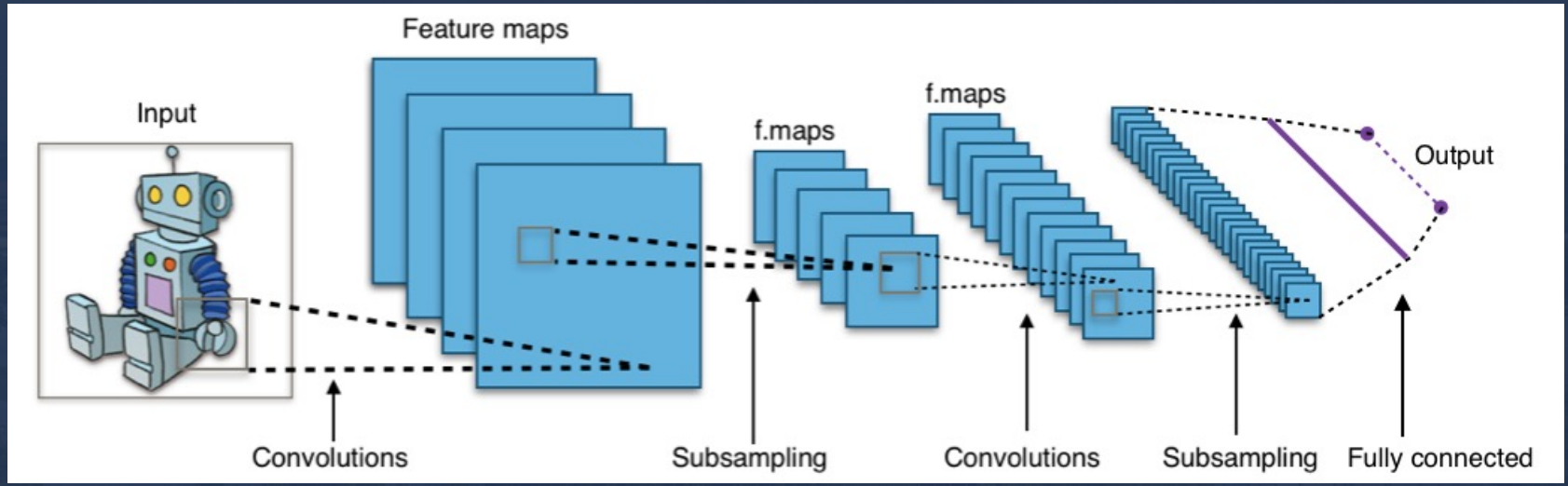
Results

	Accuracy	Time Consumption
With PCA	0.975	0.048 sec
Without PCA	0.986	0.491 sec

Model 3- Convolutional Neural Network (CNN)

CNN

Methodology



CNN

Parameters Tuned	Trials
Nodes & Layers	2 hidden layers with 32, 64 nodes 3 hidden layers with 32, 64, 128 nodes
Batch Size	32, 64
Epoch	25, 60

Optimal model

- Nodes & layers: 2 hidden layers with 32, 64 nodes respectively
- Kernel size of convolutional layer: 3x3
- Subsampling method: max-pooling (size 2x2)
- Activation function: *relu*
- Optimizer: *adam*

CNN

Results

	2 layers with 32, 64 nodes	3 layers with 32, 64, 128 nodes
Batch size=32	0.9750	0.9625
Batch size=64	0.9375	0.9250

Notes

- The optimal model has an accuracy of *0.975*

Results Comparison

Results Comparison

	Logistic Regression	LDA	CNN
Accuracy	0.975	0.986	0.975

- All 3 models did well on this dataset
- LDA has slightly higher accuracy

Difficulties

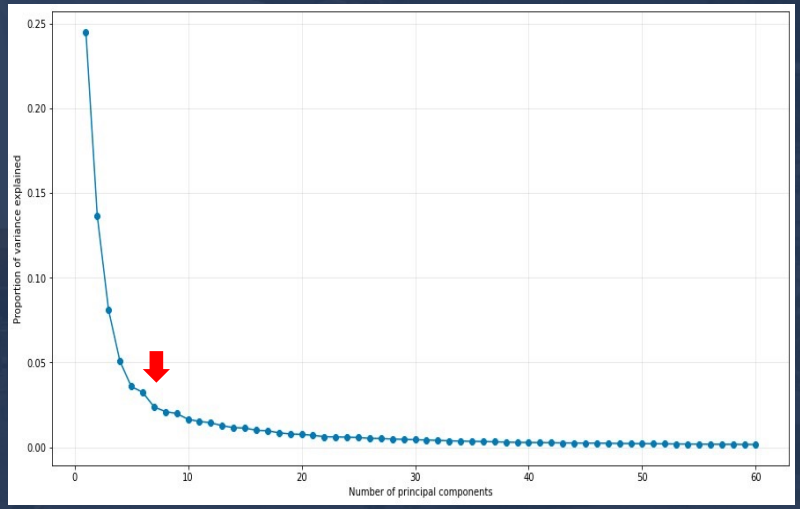
Difficulties

Problem 1

- We decided to use the first 7 principal components initially by looking at the scree plot, but then found that the accuracy would drop too much compared to that without using PCA

Solution:

- We decided to choose the number of principal components which can achieve a cumulative percentage of total variation larger than 90% instead



Difficulties

Problem 2

- Overfitting issues: the accuracy of the training set could be trained to 1

Solution:

- Drop the 3^{rd} layer with 128 neurons

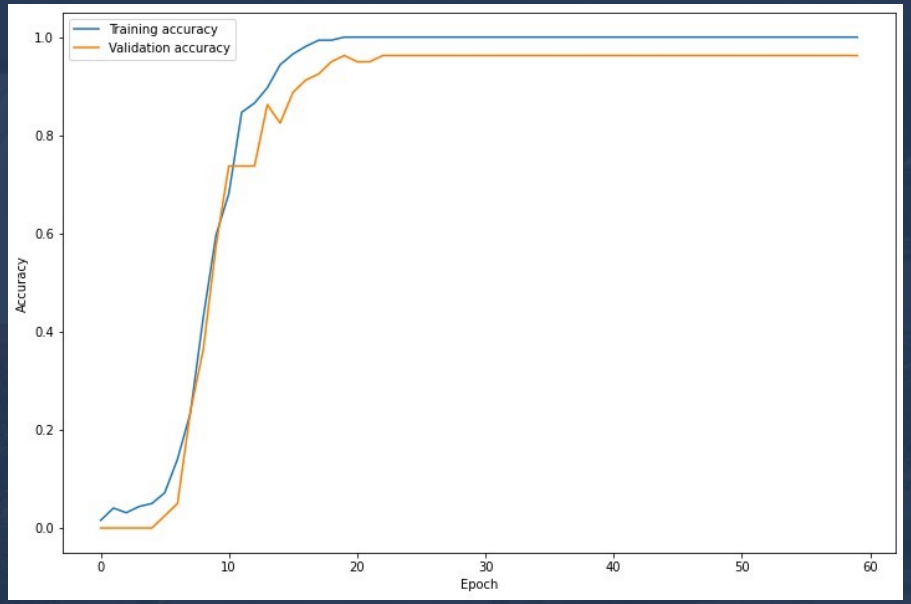
Difficulties

Problem 3

- Classification accuracy has already settled when *Epoch*=60

Solution:

- Choose *Epoch*=25



Thank you