Assignment 2

Rik Vegter (s3147495)
Jordi Timmermans (s3870758)
Hari Vidharth (s4031180)
Ayça Avcı (s4505972)

Group 12

September 23, 2020

2 Part A

2.1 Exercises

2.1.1 Eigenvalues and Eigenvectors

Calculating the eigenvalues of A

$Det(A - \lambda * I) = 0$

$A = \begin{bmatrix} 3 & 4 \\ 5 & 8 \end{bmatrix}$

$Det(\begin{bmatrix} 3-\lambda & 4 \\ 5 & 8-\lambda \end{bmatrix}) = (3-\lambda) * (8-\lambda) - 5 * 4$

$\lambda^2 - 11\lambda + 4 = 0$

$(\lambda - \frac{11}{2})^2 - \frac{105}{4} = 0$

$\lambda_1 = \frac{1}{2}(11 + \sqrt{105})$

$\lambda_2 = \frac{1}{2}(11 - \sqrt{105})$

Calculating the eigenvectors of A

$\begin{bmatrix} 3 & 4 \\ 5 & 8 \end{bmatrix} \begin{bmatrix} x_{11} \\ x_{12} \end{bmatrix} = \frac{1}{2}(11 + \sqrt{105}) \begin{bmatrix} x_{11} \\ x_{12} \end{bmatrix}$

Calculating $\vec{v_1}$

$3 * x_{11} + 4 * x_{12} = \frac{1}{2}(11 + \sqrt{105}) * x_{11}$

$8 * x_{12} = 5 + \sqrt{105})x_{11}$

$\vec{v_{11}} = \begin{bmatrix} 5 + \sqrt{105} \\ 8 \end{bmatrix}$

Normalizing $\vec{v_1}$

$||\vec{v_1}|| \frac{\begin{bmatrix} 5 + \sqrt{105} \\ 8 \end{bmatrix}}{\sqrt{8^2 + (5 + \sqrt{105})^2}} \approx \begin{bmatrix} 0.89 \\ 0.46 \end{bmatrix}$

Calculating $\vec{v_2}$

$\begin{bmatrix} 3 & 4 \\ 5 & 8 \end{bmatrix} \begin{bmatrix} x_{21} \\ x_{22} \end{bmatrix} = \frac{1}{2}(11 - \sqrt{105}) \begin{bmatrix} x_{21} \\ x_{22} \end{bmatrix}$

$3 * x_{21} + 4 * x_{22} = \frac{1}{2}(11 - \sqrt{105}) * x_{21}$

$8 * x_{22} = (5 - \sqrt{105})x_{21}$

$\vec{v_2} = \begin{bmatrix} 5 - \sqrt{105} \\ 8 \end{bmatrix}$

Normalizing $\vec{v_2}$

$||\vec{v_2}|| \frac{\begin{bmatrix} 5 - \sqrt{105} \\ 8 \end{bmatrix}}{\sqrt{8^2 + (5 - \sqrt{105})^2}} \approx \begin{bmatrix} -0.55 \\ 0.84 \end{bmatrix}$

Calculating the eigenvalues of B

$Det(B - \lambda * I) = 0$

$B = \begin{bmatrix} 4 & 2 \\ 3 & 1 \end{bmatrix}$

$Det(\begin{bmatrix} 4-\lambda & 2 \\ 3 & 1-\lambda \end{bmatrix}) = (4-\lambda) * (1-\lambda) - 2 * 3$

$\lambda^2 - 5\lambda - 2 = 0$

$(\lambda - \frac{5}{2})^2 - \frac{33}{4} = 0$

$\lambda_1 = \frac{1}{2}(5 + \sqrt{33})$

$\lambda_2 = \frac{1}{2}(5 - \sqrt{33})$

Calculating eigenvectors of B

$\begin{bmatrix} 4 & 2 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} x_{11} \\ x_{12} \end{bmatrix} = \frac{1}{2}(5 + \sqrt{33}) \begin{bmatrix} x_{11} \\ x_{12} \end{bmatrix}$

Calculating $\vec{v_1}$

$4 * x_{11} + 2 * x_{12} = \frac{1}{2}(5 + \sqrt{33}) * x_{11}$

$4 * x_{12} = (-3 + \sqrt{33}) x_{11}$

$\vec{v_{11}} = \begin{bmatrix} -3 + \sqrt{33} \\ 4 \end{bmatrix}$

Normalizing $\vec{v_1}$

$||\vec{v_1}|| \frac{\begin{bmatrix} -3 + \sqrt{33} \\ 4 \end{bmatrix}}{\sqrt{4^2 + (-3+\sqrt{33})^2}} \approx \begin{bmatrix} 0.57 \\ 0.82 \end{bmatrix}$

Calculating $\vec{v_2}$

$\begin{bmatrix} 4 & 2 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} x_{21} \\ x_{22} \end{bmatrix} = \frac{1}{2}(5 - \sqrt{33}) \begin{bmatrix} x_{21} \\ x_{22} \end{bmatrix}$

$4 * x_{21} + 2 * x_{22} = \frac{1}{2}(5 - \sqrt{33}) * x_{21}$

$4 * x_{22} = (-3 - \sqrt{33}) x_{21}$

$\vec{v_2} = \begin{bmatrix} -3 - \sqrt{33} \\ 4 \end{bmatrix}$

Normalizing $\vec{v_2}$

$||\vec{v_2}|| \frac{\begin{bmatrix} -3 - \sqrt{33} \\ 4 \end{bmatrix}}{\sqrt{4^2 + (-3-\sqrt{33})^2}} \approx \begin{bmatrix} -0.91 \\ 0.42 \end{bmatrix}$

2.1.2 ANOVA

$\overline{x}_{High} = \frac{9+7+6.5+8+7.5+7+9.5+8+6.5}{9} = \frac{69}{9}$

$\overline{x}_{Medium} = \frac{7.5+8+6+7+6.5+7.5}{6} = \frac{85}{12}$

$\overline{x}_{Low} = \frac{8+6+6+6.5+6.5}{5} = \frac{33}{5}$

$\sigma^2_{High} = \frac{(9-\frac{69}{9})^2+(7-\frac{69}{9})^2+(6.5-\frac{69}{9})^2+(8-\frac{69}{9})^2+(7.5-\frac{69}{9})^2+(7-\frac{69}{9})^2+(9.5-\frac{69}{9})^2+(8-\frac{69}{9})^2+(6.5-\frac{69}{9})^2}{8} \frac{9}{8}$

$\sigma^2_{Medium} = \frac{(7.5-\frac{85}{12})^2+(8-\frac{85}{12})^2+(6-\frac{85}{12})^2+(7-\frac{85}{12})^2+(6.5-\frac{85}{12})^2+(7.5-\frac{85}{12})^2}{5} = \frac{13}{24}$

$\sigma^2_{Low} = \frac{(8-\frac{33}{5})^2+(6-\frac{33}{5})^2+(6-\frac{33}{5})^2+(6.5-\frac{33}{5})^2+(6.5-\frac{33}{5})^2}{4} = \frac{27}{40}$

$\overline{\overline{x}} = \frac{69+42.5+33}{20} = \frac{289}{40}$

Calculating variance between groups

$SS(B) = 9 * (\frac{69}{9} - \frac{289}{40})^2 + 6 * (\frac{85}{12} - \frac{289}{40})^2 + 5 * (\frac{33}{5} - \frac{289}{40})^2 = \frac{919}{240} \approx 3.83$

$MS(B) = \frac{SS(B)}{k-1} \approx 1.91$

Calculating variance within groups

$SS(W) = 8 * \frac{9}{8} + 5 * \frac{13}{24} + 4 * \frac{27}{40} = \frac{1729}{120} \approx 14.41$

$MS(W) = \frac{SS(W)}{N-k} \approx 0.85$

Calculating F statistic

$F = \frac{SS(B)}{SS(W)} \approx 2.26$

Source	SS	df	MS	F
Between	3.83	2	1.91	2.26
Within	14.41	17	0.85	NA
Total	18.24	19	NA	NA

2.2 Implementation

2.2.1 Implementation of the F-statistic

See matlab file (**myOneWayANOVA.m**)

2.2.2 Principal Component Analysis

See matlab file (**mypca.m**)

2.2.3 Application: Face Recognition

(a) The result of the ANOVA approach can be seen in figure 1. It can be seen that with relatively few number of features, a good accuracy can be achieved. The accuracy peaks between around 250 and 450 features, after which it decreases slightly.

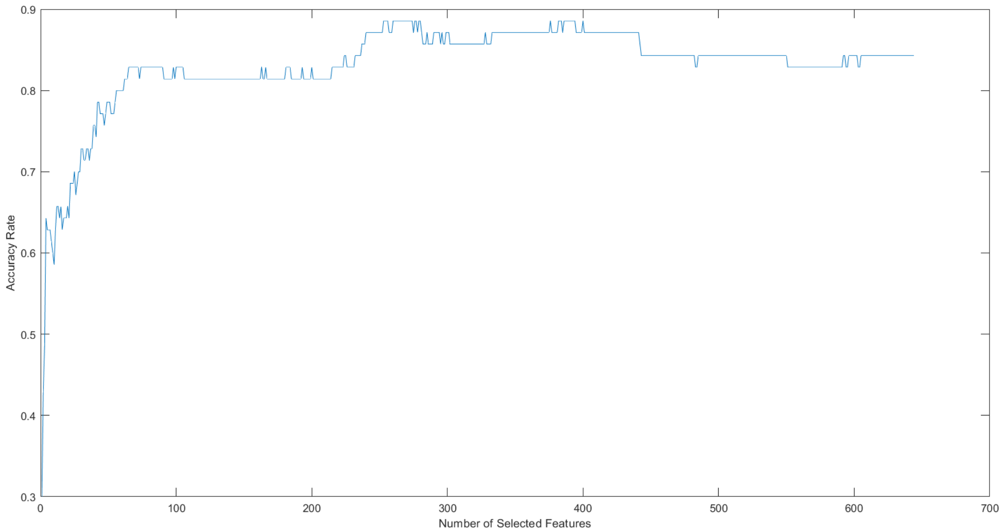

Figure 1: Accuracy ANOVA approach

(b) The result of the PCA approach can be seen in figure 2. The accuracy of this approach keeps increasing with the number of features for the first ≈ 30 features. After this the accuracy graph flattens out.
In this file, we added 3 lines which are 51, 55 and 56. We added the following code in the respective lines:

```
            trainingVectors = pc(:, 1:K)' * trainingData;
            testData = testingData - repmat(meanTrainingFace,1,size(testingData
                ,2));
            accuracy(idx) = classify(testingVectors,trainingVectors,
                trainingFilenameList,testingFilenameList);
```

Comparing the two methods we find that for this approach ANOVA performs best accuracy-wise with a peak accuracy of 89%, compared to the 71% of the PCA approach as can be seen in 1 and 2. Run time wise PCA performs best with a run time of 0.1s compared to ANOVA's run time which is 4.2s
Another difference, which may or may not be considered a disadvantage of PCA compared to ANOVA, in this case is that using PCA, one needs to preprocess the data. Other advantage of the ANOVA in feature selection is that it makes comparison on means between the groups that are interested in and determines whether any of the means are statistically significantly differs from each other. What PCA does is to select variables according to their magnitude of coefficients. The other disadvantage of it is that it does not consider higher order interactions between variables in the data set and only the linear relationships are taken into account.

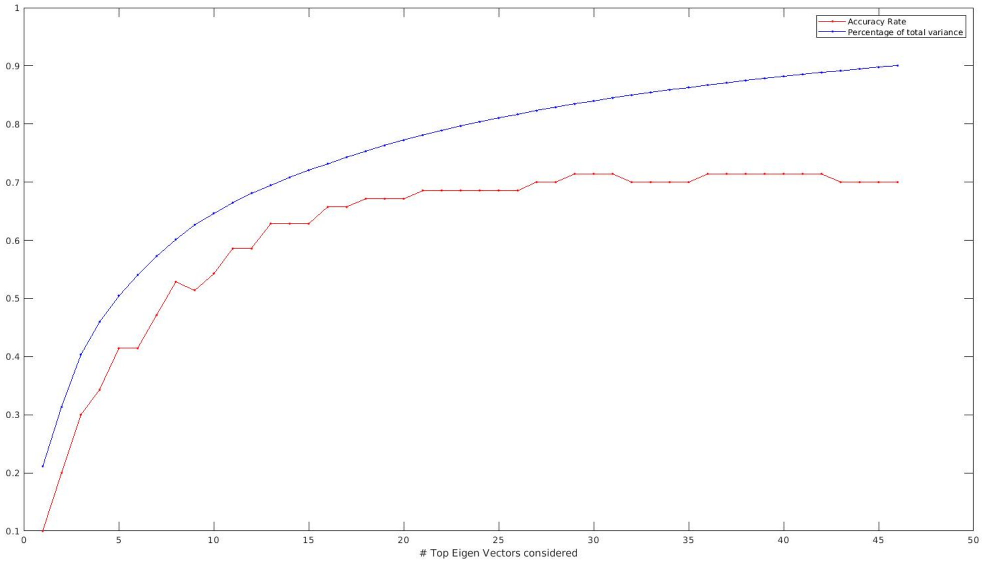

Figure 2: Accuracy PCA approach

3 Part B

3.1 Exercises

3.1.1 Classification metrics and unbalanced classes

(a) $recall = TRP = \frac{TP}{TP+FN} = \frac{15}{15+20} = \frac{15}{35} \approx 0.43$
$specificity = TNR = \frac{TN}{TN+FP} = \frac{1965}{1965+0} = 1$
$precision = \frac{TP}{TP+FP} = \frac{15}{15+0} = 1$
$classification\ accuracy = \frac{TP+TN}{TP+FP+TN+FN} = \frac{15+1965}{15+0+1965+20} = \frac{99}{100} = 0.99$
$F_1\ score = \frac{2*precision*recall}{precision+recall} = \frac{2*1*\frac{15}{35}}{1+\frac{15}{35}} = \frac{3}{5} = 0.6$

Ground truth: Class A = 15 + 20 = 35, Class B = 1965 + 0 = 1965
ML Classifier: Class A = 15 + 0 = 15, Class B = 1965 + 20 = 1985
Total: 15 + 20 + 1965 + 0 = 2000

$Expected\ Accuracy = \frac{15*35/2000}{1985*1965/2000}/2000 \approx 0.975$

$Observed\ Accuracy = \frac{15+1965}{2000} = 0.99$

$Cohen\ Kappa = \frac{0.99-0.9752625}{1-0.975} \approx 0.598$

(b) The total number of negative classes (majority) is 1965, the total number of positive classes (minority) is 35, the IR is thus 1965 : 35.

(c) In this case, the only wrong classifications happen by classifying positives as negatives. The metric to evaluate this would be the recall or true positive rate. This indicates what ratio of the positives you actually classify correct.

3.1.2 Feature selection with Information Gain

First we sort the table based on output to make counting easier, which can be seen in table 1.

Sky condition	Temperature	Humidity	Windy	Cycle
Cloudy	High	High	False	Yes
Rain	Mid	High	False	Yes
Rain	Low	Low	False	Yes
Cloudy	Low	Low	True	Yes
Sunny	Low	Low	False	Yes
Rain	Mid	Low	False	Yes
Sunny	Mid	Low	True	Yes
Cloudy	Mid	High	True	Yes
Cloudy	High	Low	False	Yes
Sunny	High	High	False	No
Sunny	High	High	True	No
Rain	Low	Low	True	No
Sunny	Mid	High	False	No
Rain	Mid	High	True	No

Table 1: Table ordered on Cycle variable

Calculating the Shannon's entropy of the output variable
We have 9 times "Yes" and 5 times "No" out of a total of 14 instances, which gives:
$H(Y) = -\frac{9}{14} * log_2(\frac{9}{14}) - \frac{9}{14} * log_2(\frac{9}{14}) \approx 0.94$
Calculating the conditional entropy and information gain of the Sky condition variable
We have 5 times Sunny (2 yes, 3 no), 4 times Cloudy (4 yes, 0 no) and 5 times Rainy (3 yes, 2 no). From this follows:
$H(Y|X) = \frac{5}{14} * (-\frac{3}{5} * log_2(\frac{3}{5}) - \frac{2}{5} * log_2(\frac{2}{5})) + \frac{4}{14} * 0 + \frac{5}{14} * (-\frac{3}{5} * log_2(\frac{3}{5}) - \frac{2}{5} * log_2(\frac{2}{5})) \approx 0.69$
The information gain (H(Y)-H(Y—X)) is approximately $0.94 - 0.69 = 0.25$
Calculating the conditional entropy and information gain of the Temperature variable We have 4 times High (2 yes, 2 no), 6 times Mid (4 yes, 2 no) and 4 times Low (3 yes, 1 no). From this follows:
$H(Y|X) = \frac{4}{14} * (-2 * \frac{2}{4} * log_2(\frac{2}{4})) + \frac{6}{14} * (-\frac{4}{6} * log_2(\frac{4}{6}) - \frac{2}{6} * log_2(\frac{2}{6})) + \frac{4}{14} * (-\frac{3}{4} * log_2(\frac{3}{4}) - \frac{1}{4} * log_2(\frac{1}{4}))$
≈ 0.91
The information gain is approximately $0.94 - 0.91 = 0.03$
Calculating the conditional entropy and information gain of the Humidity variable We have 7 times High (3 Yes, 4 No), 7 times Low (6 Yes, 1 No). From this follows:
$H(Y|X) = \frac{7}{14} * (-\frac{3}{7} * log_2(\frac{3}{7}) - \frac{4}{7} * log_2(\frac{4}{7})) + \frac{7}{14} * (-\frac{6}{7} * log_2(\frac{6}{7}) - \frac{1}{7} * log_2(\frac{1}{7})) \approx 0.79$
The information gain is approximately $0.94 - 0.79 = 0.15$
Calculating the conditional entropy for the Windy variable
We have 8 times False (6 Yes, 2 No) and 6 times True (3 Yes, 3 No). From this follows:
$H(Y|X) = \frac{8}{14} * (-\frac{6}{8} * log_2(\frac{6}{8}) - \frac{2}{8} * log_2(\frac{2}{8})) + \frac{6}{14} * (-2 * \frac{3}{6} * log_2(\frac{3}{6})) \approx 0.89$
The information gain is approximately $0.94 - 0.89 = 0.05$
The features, ranked on relevance from high to low can be found in table x.

Variable	Information gain
Sky condition	0.25
Humidity	0.15
Windy	0.05
Temperature	0.03

Table 2: Caption

3.2 Implementation

3.2.1 Implementation of single imputation

See function *MyImpute.m* and see the *Readme*

3.2.2 Relief feature selection

See function *MyRelief.m*
Using the Relief function, we obtained the weight vector:

$$W = \begin{bmatrix} 0.214 \\ -0.143 \\ -0.143 \\ 0.071 \end{bmatrix}$$

This means that our ranking of feature relevance from high to low is: Sky condition, Windy, and tied: Humidity and temperature.
Using the information gain, we obtained the following ranking: Sky condition, Humidity, Windy, Temperature.
Comparing the two we see that in both cases the Sky condition feature is the most relevant feature and Temperature is the least relevant.