**Name: Chandramohan Natarajan**

**Student number: 190617866**

**Assignment number: 4**

**Module: ECS766**

**Exercise 0:** Now starting from this 4-attribute subset, find the best 3, 2, 1 attribute subset, filling in the table below. Which sized subset, and which set of attributes yields the best accuracy? [1 mark]

| Subset size | Attributes selected | Accuracy | Attribs removed |
|---|---|---|---|
| 5 | W, Hol, Vac, Health, P | 85.96% | None |
| 4 | W, Hol , Health, P | 85.47% | Vac |
| 3 | W, Hol, P | 91.23% | Vac, Health |
| 2 | W, P | 85.96% | Vac, Hol, Health |
| 1 | P | 80.70% | W, Vac, Hol, Health |

**It is quite palpable from the table that the attributes combination of Wage, Holidays and Pension gives maximum accuracy.**

**Exercise 1:** How many feature combinations did you try? How many combinations of features are there in total? Give an example of a combination of features that you did NOT try when doing backward selection. [1 mark]

**By calculation, there should be 31 different combinations with 5 attributes. The reason for choosing the attributes were the presence of instances kin the data set. Based on that, initially five attributes namely {W,P,H,Vac and Hol } are selected with 15 different combinations.**

{
w
W,p
W,p,hol
W,p,hol,v
W,p,hol,v,health
}
{
p
P,hol
P,hol,vac
P,hol,vac,health
}
{Hol
Hol,Vac
Hol,vac,Health}
{
vac,
Vac,health
}
{
Health
}

Example of a combination of features that NOT tried when doing backward selection are

{

Health

Heal,p

Health,p,hol

Heal,pen,hol,vac

}

{

W

W,hol

W,hol,vac

W,hol,vac,heal

}

---

**Exercise 2:** How many and which attributes are selected? Do they match the results from *Section 2*? [1 mark]

Out of the six attributes, information ranker points its ranks clearly which can be observed from the Run information.

**Run information**

Evaluator: weka.attributeSelection.InfoGainAttributeEval

Instances: 57

Attributes: 6

- wage-increase-first-year
- pension
- statutory-holidays
- vacation
- contribution-to-health-plan
- class

Evaluation mode: evaluate on all training data

Attribute Selection on all input data

Search Method:

Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 6 class):

Information Gain Ranking Filter

Ranked attributes:

**0.2948 1 wage-increase-first-year**

**0.1624 3 statutory-holidays**

0.1164 5 contribution-to-health-plan

0.1091 4 vacation

**0.0548 2 pension**

Selected attributes: 1,3,5,4,2 : 5

Information Gain Ranking method results in high importance to wage.

On the other hand, Greedy Backward Selection method selects 3 attributes, viz.

W, P, Hol. So, it is observed that W is considered as an important attribute in the both the methods and P is

important only according to the Greedy algorithm.

---

**Exercise 3:** Which attributes does it pick (and hence which ones are discarded?) [1 mark]

Altogether different attributes selected were eleven in number

They are sepal length, sepal width, petal length, petal width, class, copy of sepal width, copy of sepal length and its repeated copies.

But the selected attributes were petal length and petal width.

---

**Exercise 4:** Use the data used to produce the above plot to find out what number of PCs is required to explain 99% of the data variance (achieve 99% reconstruction accuracy). What # is this and does it match the value from **Q8**? Provide a short discussion. [1 mark]

The loaded face database 165 rows and 4096 columns.

Each row corresponds to face image and column corresponds to pixel image. Image is displayed by reshaping each of 1 x 4096 vectors to 64 x 64 matrix.

When we gradually increment the principal components, towards reconstruction, we should be able to observe the Reconstruction error which keeps changing resulting in clarity.

```
original size: 5406.72 (KB)
reduced size: 33.0 (KB)

Reconstruction error for nPCA = 25 is: 0.024254014179209613

------------------------------------------------------------

original size: 5406.72 (KB)
reduced size: 66.0 (KB)

Reconstruction error for nPCA = 50 is: 0.011474148236758946
```

---

```
original size: 5406.72 (KB)
reduced size: 72.6 (KB)

Reconstruction error for nPCA = 55 is: 0.01010096555425931

------------------------------------------------------------------------

original size: 5406.72 (KB)
reduced size: 99.0 (KB)

Reconstruction error for nPCA = 75 is: 0.006154662201600863

------------------------------------------------------------

original size: 5406.72 (KB)
reduced size: 132.0 (KB)

Reconstruction error for nPCA = 100 is: 0.0031827383679695577

------------------------------------------------------------

 original size: 5406.72 (KB)

reduced size: 165.0 (KB)

Reconstruction error for nPCA = 125 is: 0.0013731531541292148
```

99% reconstruction accuracy is achieved when principal components used cross 55. As principal components increase, the picture clarity increases and looks like real when PC cross 100.

---

**Exercise 5:** Which number of PCA dimensions gets the maximum face recognition accuracy? Is it better or worse than the accuracy when classifying the raw images? Why? (What factors contribute to this?) Provide a brief discussion. [1 mark]

The output when we run the dataset in NN Classifier gives us the accuracy score which can be found below.

```
Accuracy score of 1-NN when nPCA = 25 is 0.695
Accuracy score of 1-NN when nPCA = 26 is 0.695
Accuracy score of 1-NN when nPCA = 27 is 0.707
Accuracy score of 1-NN when nPCA = 28 is 0.707
Accuracy score of 1-NN when nPCA = 29 is 0.695
Accuracy score of 1-NN when nPCA = 30 is 0.695
```

It is pretty clear that the number of PCA dimensions where we get maximum face recognition accuracy is 28.

When we classify the raw images, we were not able to achieve the better accuracy because of number of components.

When the components keeps increasing, it shows that accuracy does not appreciate higher.

So it implies, by providing more components, we cannot expect appreciation in accuracy with these program structure.