

Homework 1 – Jonas Strand Aasberg

CQ1:

Found in Excel directly:

Sample average: 467,4876

Sample Standard deviation: 288,0880

CQ2:

- a) Regression, because it provides a linear set of outputs
- b) This is Extrapolation because it predicts something outside the original sample area

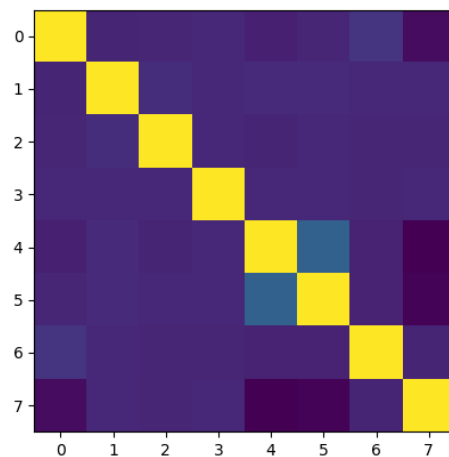
CQ3:

```
1  def sum_to_n(arr, n):
2      res = []
3      for a in arr:
4          for b in range(round(len(arr) / 2)):
5              if a + arr[b] == n:
6                  res.append([a, arr[b]])
7      return res
8
9
10 # print(sum_to_n([1, 2, 3, 4],5))
11 # print(sum_to_n([1, 4, 5, 3, 2],6))
12 # print(sum_to_n([1, 2, 5, 6, 3],7))
```

CQ4:

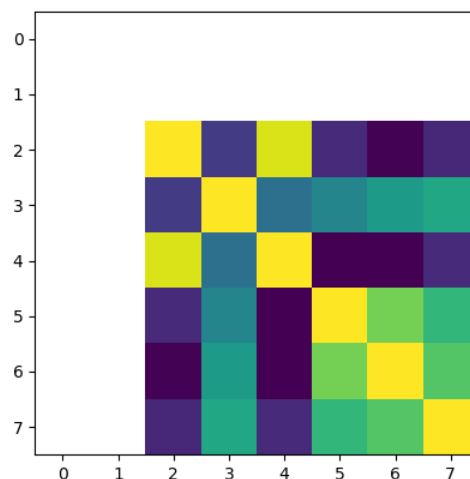
- a) Features A,B and C are categorical because they only take an integer value and in my mind are most likely related to year day and month. The day and month value fluctuates between 1 and 7 and a and 31 respectively.
- b) Features E and F are most strongly correlated shown from this snippet of code and plot. Indexes 5 and 4 indicate feature F and E respectively. As they are positively correlated they seem to increase and decrease in some degree of sync with each other.

```
14 import pandas as pd
15 import matplotlib.pyplot as plt
16 from mpl_toolkits.mplot3d import Axes3D
17
18 data = pd.read_csv(r"C:\Users\jonas\Documents\UQ\COMP4702\Homework\HW01\mystery.csv")
19
20 print(data.describe())
21 print(data.corr())
22 print(data.cov())
23 plt.imshow(data.corr())
24 plt.show()
```



- c) The features E and H are negatively correlated to some degree.
- d) F is a value fluctuating between 180 and -180 this leads me to believe it is either a measurement of degrees or a part of polar coordinates.
- e)

i)



- ii) D and F have the lowest correlation value: -0.019961
- iii) They are the two least correlated which means they have the lowest degree of linear relation. The low value indicates a low or almost non existent linear relationship.