1 a) Model 2: $S := S_1 = S_2$

$$g_i(x) = \frac{-d}{2}\log(2\pi) - \frac{1}{2}\log(|\Sigma|) - \frac{1}{2}(x-\mu_i)^T\Sigma^{-1}(x-\mu_i) + \log(P(C_i))$$

$$\mathcal{L}(\Sigma^{-1}|x) = \sum_{t=1}^{N} \log(g_i(x^t)) \equiv \sum_{t=1}^{N} g_i(x^t)$$

$$= \sum_{t=1}^{N}\left[-\frac{d}{2}\log(2\pi) - \frac{1}{2}\log(|\Sigma|) - \frac{1}{2}(x^t-\mu_i)^T\Sigma^{-1}(x^t-\mu_i) + \log(P(C_i))\right]$$

$$\equiv \sum_{t=1}^{N}\left[-\frac{1}{2}\log(|\Sigma|) + -\frac{1}{2}(x^t-\mu_i)^T\Sigma^{-1}(x^t-\mu_i)\right]$$

$$= -\frac{N}{2}\log(|\Sigma|) - \frac{1}{2}\sum_{t=1}^{N}(x^t-\mu_i)^T\Sigma^{-1}(x^t-\mu_i)$$

$$\frac{\partial \mathcal{L}}{\partial \Sigma^{-1}} = -\frac{N}{2}\frac{\partial(\log(|\Sigma|))}{\partial \Sigma^{-1}} - \frac{1}{2}\sum_{t=1}^{N}\frac{\partial\left[(x^t-\mu_i)^T\Sigma^{-1}(x^t-\mu_i)\right]}{\partial \Sigma^{-1}}$$

$$= -\frac{N}{2}(-\Sigma^T) - \frac{1}{2}\sum_{t=1}^{N}(x^t-\mu_i)^T(x^t-\mu_i) = 0$$

$$\frac{N}{\cancel{2}}\Sigma = \frac{1}{\cancel{2}}\sum_{t=1}^{N}(x^t-\mu_i)^T(x^t-\mu_i)$$

$$\boxed{S = \Sigma = \frac{\sum_{t=1}^{N}(x^t-\mu_i)^T(x^t-\mu_i)}{N}}$$

1a) <u>Model 3:</u> $S_1 = \alpha_1 I \qquad S_2 = \alpha_2 I$

$$g_i(x) = \frac{-d}{2}\log(2\pi) - \frac{1}{2}\log(|\Sigma_i|) - \frac{1}{2}(x-\mu_i)^T\Sigma_i^{-1}(x-\mu_i) + \log(P(C_i)).$$

$$\frac{-1}{2}\log(|\Sigma_i|) = \frac{-1}{2}\log(|\alpha_i I|) = \frac{-1}{2}\log(\alpha_i^d) = \frac{-d}{2}\log(\alpha_i)$$

$$\Sigma_i^{-1} = (\alpha_i I)^{-1} = (\alpha_i^{-1})I$$

$$g_i(x) = \frac{-d}{2}\log(2\pi) - \frac{d}{2}\log(\alpha_i) - \frac{1}{2}\alpha^{-1}(x-\mu_i)^T(x-\mu_i) + \log(P(C_i))$$

$$\mathcal{L}(\alpha_i|x) \equiv \sum_{t=1}^{N}\log(g_i(x^t))$$
$$\equiv \frac{-1}{2}\sum_{t=1}^{N}\log(d\log(\alpha_i) + \alpha_i^{-1}(x^t-\mu_i)^T(x^t-\mu_i))$$
$$\equiv \frac{-1}{2}\sum_{t=1}^{N}\left[d\log(\alpha_i) + \alpha_i^{-1}(x^t-\mu_i)^T(x^t-\mu_i)\right]$$
$$= \frac{-1}{2}\cdot Nd\log(\alpha_i) + \alpha_i^{-1}\sum_{t=1}^{N}(x^t-\mu_i)^T(x^t-\mu_i)$$

$$\frac{\partial\mathcal{L}}{\partial\alpha_i} = \frac{-1}{2}\cdot Nd\,\frac{\partial(\log(\alpha_i))}{\partial\alpha_i} + \frac{\partial(a_i^{-1})}{\partial a_i}\sum_{t=1}^{N}(x^t-\mu_i)^T(x-\mu_i)$$
$$= \frac{-Nd}{2\alpha_i} + \frac{-1}{\alpha_i^2}\cdot\sum_{t=1}^{N}(x^t-\mu_i)^T(x-\mu_i) = 0.$$

$$\boxed{\alpha_i = \frac{2}{Nd}\sum_{t=1}^{N}(x^t-\mu_i)^T(x-\mu_i)}$$

1b) data 1, model 1:

```
pc1 =

    0.3000


pc2 =

    0.7000


m1 =

    0.4306
    2.0235
    3.1758
   -2.4272
   -2.5234
    3.2378
   -5.5208
   -6.6921


m2 =

    4.5841
    6.4933
    6.4265
    1.6891
    2.2943
    8.3626
   -0.1658
   -1.8048
```

```
S1 =

    1.8640    0.2267    0.7462    0.9977    0.4178    1.2227    1.1337   -1.1909
    0.2267    3.5370    0.3019   -0.1300    1.5294    0.9958   -0.1878    3.1664
    0.7462    0.3019    7.8426    1.2902   -0.4143    1.7198    0.3431    0.2243
    0.9977   -0.1300    1.2902    4.0886    0.9166    0.7222    1.0326    1.9150
    0.4178    1.5294   -0.4143    0.9166    3.9976    0.9693   -0.5271    3.3238
    1.2227    0.9958    1.7198    0.7222    0.9693    3.9339   -0.1894    2.2238
    1.1337   -0.1878    0.3431    1.0326   -0.5271   -0.1894    4.0757   -1.6529
   -1.1909    3.1664    0.2243    1.9150    3.3238    2.2238   -1.6529   16.5256


S2 =

    3.4237    2.0692    2.5707    2.6127    1.7732    1.8303    2.6792    2.9340
    2.0692    5.7835    2.1793    2.7182    3.1575    2.8877    2.7564    5.8537
    2.5707    2.1793    8.7126    3.3752    2.8256    2.2294    2.7535    5.1835
    2.6127    2.7182    3.3752    8.1683    3.5774    2.6607    2.0204    8.3968
    1.7732    3.1575    2.8256    3.5774    5.5677    2.9061    3.2500    4.8208
    1.8303    2.8877    2.2294    2.6607    2.9061    3.7294    2.2349    4.4766
    2.6792    2.7564    2.7535    2.0204    3.2500    2.2349    8.2148    4.3055
    2.9340    5.8537    5.1835    8.3968    4.8208    4.4766    4.3055   19.8477


err =

    0.2200
```

Data 2, model 1:

```
pc1 =

    0.3000


pc2 =

    0.7000
```

```
m1 =

    1.0658
    2.6548
    3.2977
   -1.6793
   -1.4987
    4.3959
   -4.2138
   -4.9679


m2 =

    2.8221
    4.4669
    4.8537
    0.5192
    0.3764
    6.2585
   -2.6611
   -3.8175
```

```
S1 =

    1.3355   -0.2649    0.6608    0.2725    0.9973    0.4623   -0.2879    0.2938
   -0.2649    2.5453    2.0657    0.9204    1.0862   -0.4459   -0.0506    2.6920
    0.6608    2.0657    6.9133    1.5145    0.8146    0.1927    2.6834    2.7168
    0.2725    0.9204    1.5145    3.0526    0.6584   -0.4640    0.2391    2.3010
    0.9973    1.0862    0.8146    0.6584    2.1990   -0.0037   -0.5659    0.7485
    0.4623   -0.4459    0.1927   -0.4640   -0.0037    1.2053    0.6576   -0.1986
   -0.2879   -0.0506    2.6834    0.2391   -0.5659    0.6576    4.0462   -0.7298
    0.2938    2.6920    2.7168    2.3010    0.7485   -0.1986   -0.7298   10.4733


S2 =

    2.9283    1.0700    2.9092    0.5704   -0.0910    0.0666    0.8741    1.7632
    1.0700    4.9901    2.4537    0.1758    2.7315   -0.3235    0.4943    0.0507
    2.9092    2.4537   11.6572    0.0882    0.8252    0.1275    0.0219   -0.8332
    0.5704    0.1758    0.0882    5.3310    0.0541    1.0103    4.3004    2.2958
   -0.0910    2.7315    0.8252    0.0541    4.2550    1.0248    0.5376    1.6681
    0.0666   -0.3235    0.1275    1.0103    1.0248    2.6892    1.7066    2.1595
    0.8741    0.4943    0.0219    4.3004    0.5376    1.7066   11.3170    2.7119
    1.7632    0.0507   -0.8332    2.2958    1.6681    2.1595    2.7119   12.9317


err =

    0.2300
```

Data 3, model 1:

```
pc1 =

    0.3000

pc2 =

    0.7000
```

```
m1 =

    0.9747
    2.6233
    3.1770
   -1.4652
   -1.3053
    4.5160
   -4.3197
   -5.5215

m2 =

    1.4916
    3.1655
    3.6504
   -0.8162
   -0.3515
    5.1345
   -3.2770
   -4.7293
```

```
S1 =

    0.2555    0.0816    0.0591   -0.0825   -0.0058   -0.0738   -0.0155    0.0778
    0.0816    0.4116   -0.0821   -0.1016   -0.0118   -0.0866    0.1126   -0.0733
    0.0591   -0.0821    0.5861   -0.0364   -0.0894   -0.0322    0.0797    0.0128
   -0.0825   -0.1016   -0.0364    0.4254   -0.0344   -0.1398   -0.1010    0.0721
   -0.0058   -0.0118   -0.0894   -0.0344    0.4183    0.0516   -0.0964    0.0137
   -0.0738   -0.0866   -0.0322   -0.1398    0.0516    0.5746    0.0684   -0.0237
   -0.0155    0.1126    0.0797   -0.1010   -0.0964    0.0684    0.5261   -0.0294
    0.0778   -0.0733    0.0128    0.0721    0.0137   -0.0237   -0.0294    0.3849

S2 =

    2.7546    0.4051   -0.3791    0.5899   -0.4070    0.1636    0.3523   -0.4814
    0.4051    2.5008   -0.3447   -0.0362   -0.2919    0.0578    0.1757    0.0869
   -0.3791   -0.3447    2.1297   -0.2523   -0.0483   -0.0299    0.0352   -0.2797
    0.5899   -0.0362   -0.2523    2.9260   -0.8652    0.1012   -0.3376   -0.3223
   -0.4070   -0.2919   -0.0483   -0.8652    3.4593    0.0735   -0.7518    0.1914
    0.1636    0.0578   -0.0299    0.1012    0.0735    2.8776    0.6332    0.3468
    0.3523    0.1757    0.0352   -0.3376   -0.7518    0.6332    2.9057    0.2744
   -0.4814    0.0869   -0.2797   -0.3223    0.1914    0.3468    0.2744    2.5750

err =

    0.1100
```

Data 1, model 2:

```
pc1 =

    0.3000

pc2 =

    0.7000
```

```
m1 =

    0.4306
    2.0235
    3.1758
   -2.4272
   -2.5234
    3.2378
   -5.5208
   -6.6921

m2 =

    4.5841
    6.4933
    6.4265
    1.6891
    2.2943
    8.3626
   -0.1658
   -1.8048
```

```
S =

    2.9558    1.5164    2.0233    2.1282    1.3666    1.6480    2.2156    1.6965
    1.5164    5.1096    1.6161    1.8637    2.6690    2.3201    1.8731    5.0475
    2.0233    1.6161    8.4516    2.7497    1.8536    2.0765    2.0304    3.6957
    2.1282    1.8637    2.7497    6.9444    2.7792    2.0791    1.7240    6.4523
    1.3666    2.6690    1.8536    2.7792    5.0966    2.3251    2.1169    4.3717
    1.6480    2.3201    2.0765    2.0791    2.3251    3.7908    1.5076    3.8007
    2.2156    1.8731    2.0304    1.7240    2.1169    1.5076    6.9731    2.5180
    1.6965    5.0475    3.6957    6.4523    4.3717    3.8007    2.5180   18.8511

err =

    0.1700
```

Data 2, model 2:

```
pc1 =

    0.3000

pc2 =

    0.7000
```

```
m1 =

    1.0658
    2.6548
    3.2977
   -1.6793
   -1.4987
    4.3959
   -4.2138
   -4.9679

m2 =

    2.8221
    4.4669
    4.8537
    0.5192
    0.3764
    6.2585
   -2.6611
   -3.8175
```

```
S =

    2.4505    0.6695    2.2347    0.4810    0.2355    0.1853    0.5255    1.3224
    0.6695    4.2567    2.3373    0.3992    2.2379   -0.3602    0.3308    0.8431
    2.2347    2.3373   10.2340    0.5161    0.8221    0.1471    0.8204    0.2318
    0.4810    0.3992    0.5161    4.6474    0.2353    0.5680    3.0820    2.2974
    0.2355    2.2379    0.8221    0.2353    3.6382    0.7162    0.2065    1.3922
    0.1853   -0.3602    0.1471    0.5680    0.7162    2.2440    1.3919    1.4521
    0.5255    0.3308    0.8204    3.0820    0.2065    1.3919    9.1358    1.6794
    1.3224    0.8431    0.2318    2.2974    1.3922    1.4521    1.6794   12.1942

err =

    0.5500
```

Data 3, model 2:

```
pc1 =

    0.3000

pc2 =

    0.7000
```

```
m1 =

    0.9747
    2.6233
    3.1770
   -1.4652
   -1.3053
    4.5160
   -4.3197
   -5.5215

m2 =

    1.4916
    3.1655
    3.6504
   -0.8162
   -0.3515
    5.1345
   -3.2770
   -4.7293
```

```
S =

    2.0048    0.3080   -0.2476    0.3882   -0.2866    0.0924    0.2419   -0.3136
    0.3080    1.8741   -0.2660   -0.0558   -0.2079    0.0145    0.1568    0.0388
   -0.2476   -0.2660    1.6666   -0.1875   -0.0607   -0.0305    0.0486   -0.1920
    0.3882   -0.0558   -0.1875    2.1758   -0.6160    0.0289   -0.2666   -0.2040
   -0.2866   -0.2079   -0.0607   -0.6160    2.5470    0.0670   -0.5552    0.1381
    0.0924    0.0145   -0.0305    0.0289    0.0670    2.1867    0.4638    0.2357
    0.2419    0.1568    0.0486   -0.2666   -0.5552    0.4638    2.1918    0.1832
   -0.3136    0.0388   -0.1920   -0.2040    0.1381    0.2357    0.1832    1.9180

err =

    0.4500
```

Data 1 model 3:

```
pc1 =

    0.3000

pc2 =

    0.7000
```

```
m1 =

    0.4306
    2.0235
    3.1758
   -2.4272
   -2.5234
    3.2378
   -5.5208
   -6.6921

m2 =

    4.5841
    6.4933
    6.4265
    1.6891
    2.2943
    8.3626
   -0.1658
   -1.8048
```

```
a1 =

    3.4399

a2 =

   11.1033

err =

    0.3400
```

Data 2, model 3:

```
pc1 =

    0.3000


pc2 =

    0.7000
```

```
m1 =

    1.0658
    2.6548
    3.2977
   -1.6793
   -1.4987
    4.3959
   -4.2138
   -4.9679


m2 =

    2.8221
    4.4669
    4.8537
    0.5192
    0.3764
    6.2585
   -2.6611
   -3.8175
```

```
a1 =

    2.3828


a2 =

    9.8174


err =

    0.3800
```

Data 3, model 3:

```
pc1 =

    0.3000


pc2 =

    0.7000
```

```
m1 =

    0.9747
    2.6233
    3.1770
   -1.4652
   -1.3053
    4.5160
   -4.3197
   -5.5215


m2 =

    1.4916
    3.1655
    3.6504
   -0.8162
   -0.3515
    5.1345
   -3.2770
   -4.7293
```

```
a1 =

    0.2687


a2 =

    3.8725


err =

    0.0700
```

1c) Error Rates:

	#	error
Model 1:	test_data1	0.22
	test-data2	0.23
	test-data3	0.11
Model 2:	test_data1	0.17
	test_data2	0.55
	test_data3	0.45
Model 3:	test_data1	0.34
	test_data2	0.38
	test_data3	0.07

ata3
model3

Based on the results, it seems most likely that test_data3 was distributed with S_1 very different from S_2, which is why models 1 and 3 were much more effective than model 2. It was also likely to have $S_i \approx \alpha_i I$, which meant that Model 3, a less complex model, had less error than model 1.

ta1
model2

It also likely that test-data1 had $S_1 \approx S_2$, which meant that Model 2 (less complex than model 1) had less error. It is also likely that there were many non-zero covariances, which would explain model 3's high error rate.

data2
model1

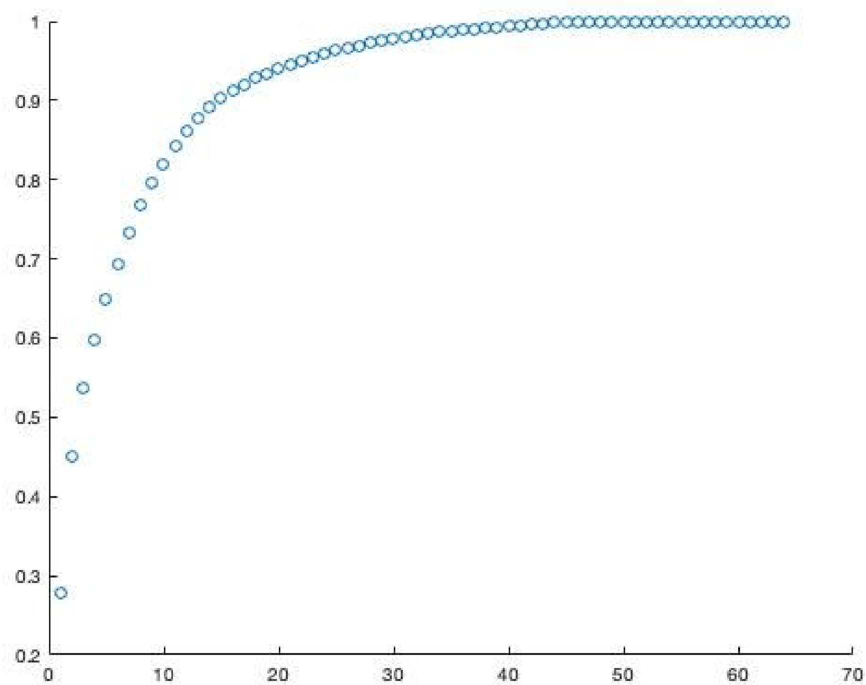
None of the models were particularly good at learning test-data2, which means that $S_1 \neq S_2$ and $S_1 \neq \alpha_i I$, but it is possible that a different simplification would allow better fit than model 1.

2a)

k	error
1	0.0539
3	0.0404
5	0.0438
7	0.0539

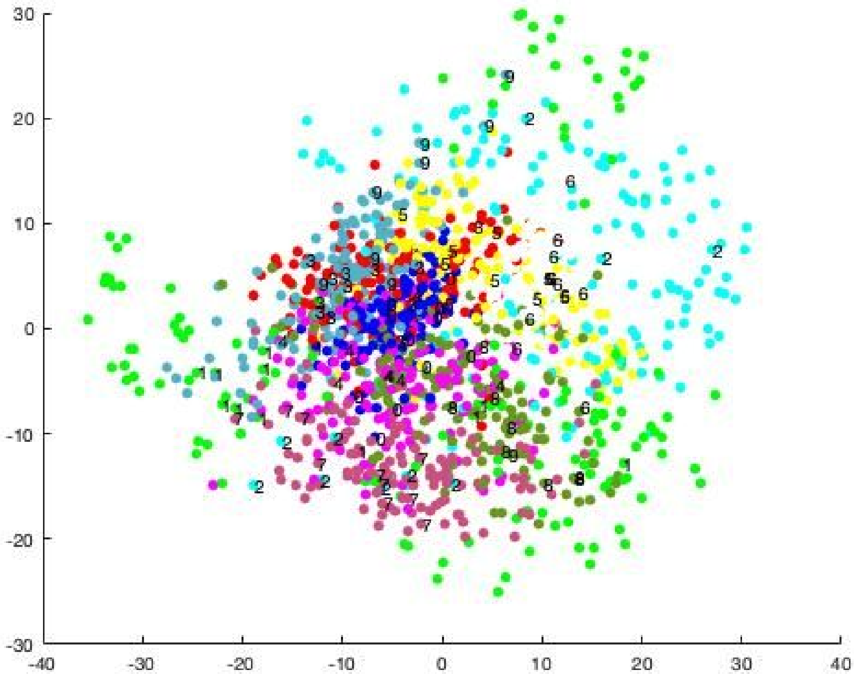
2b)

k	error
1	0.064
3	0.0505
5	0.0505
7	0.0572

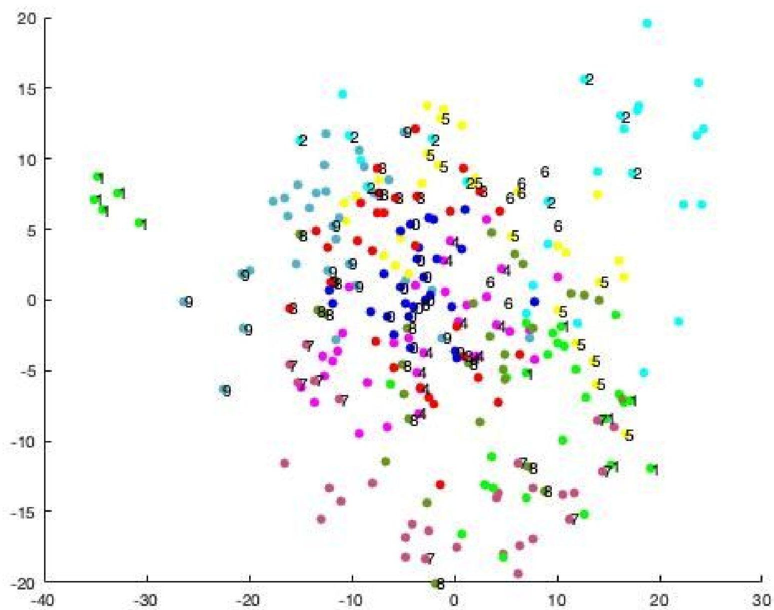


Thus we choose K = 15.

2c)
Training set:



Test set:

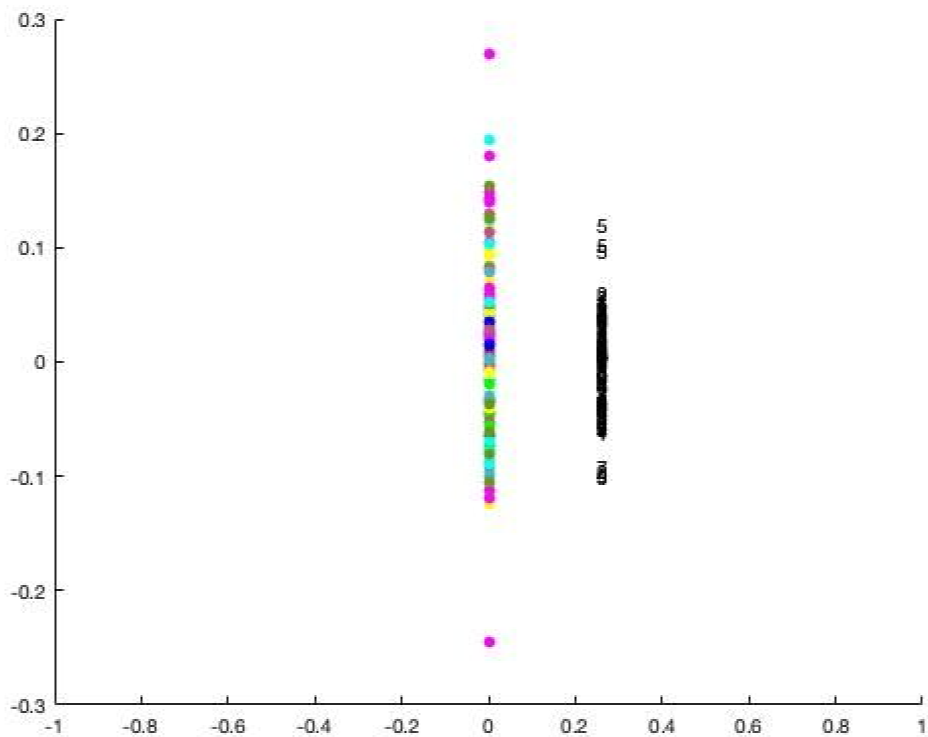


2d)

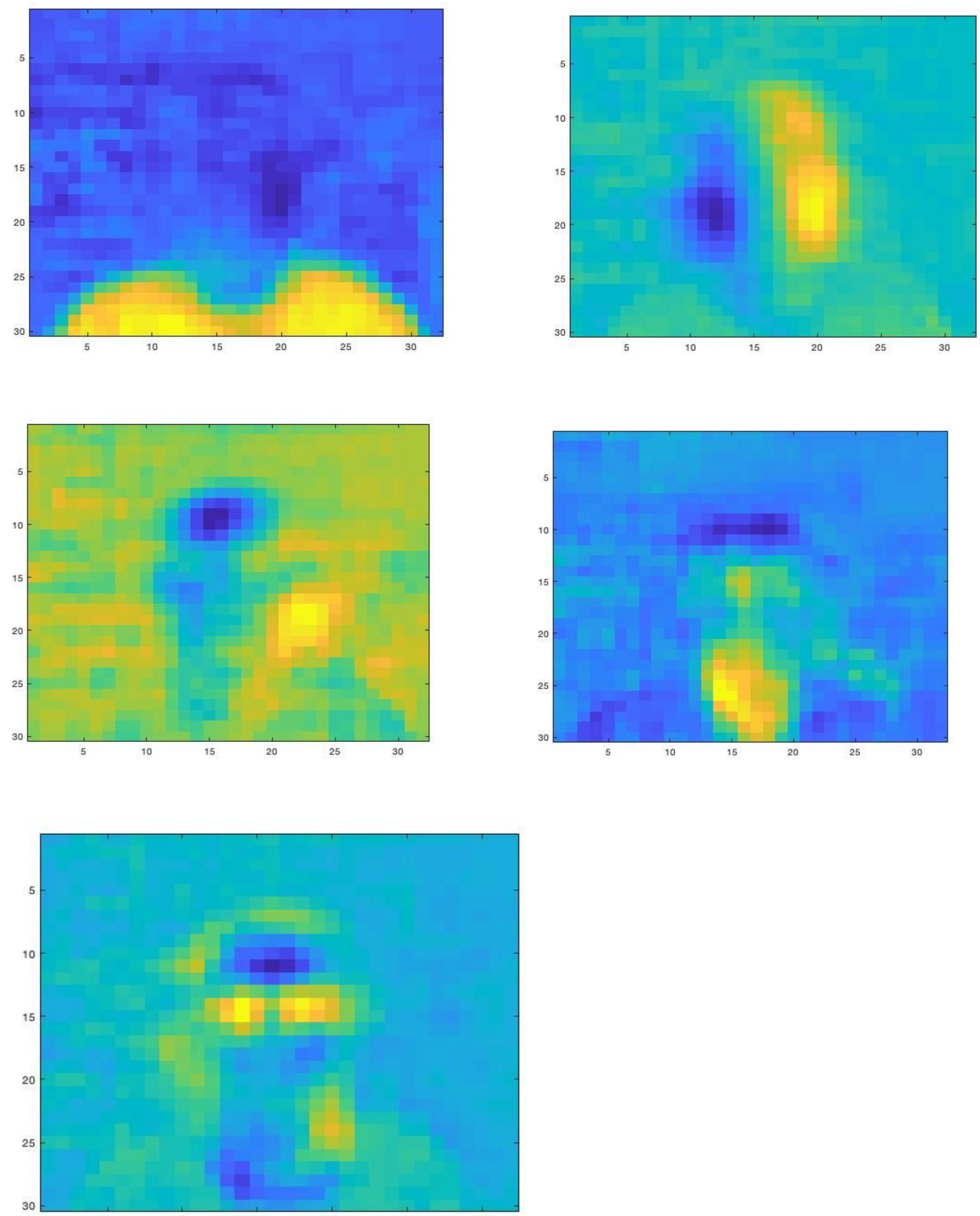
l	k	error
2	1	0.862
2	3	0.8721
2	5	0.8519
4	1	0.8249
4	3	0.8485
4	5	0.8485
9	1	0.771
9	3	0.7677
9	5	0.7407

I know this error is way too high, I’m not sure what went wrong. My myLDA function I think is mostly correct at least so hopefully I can get some partial credit for this part.

2e)

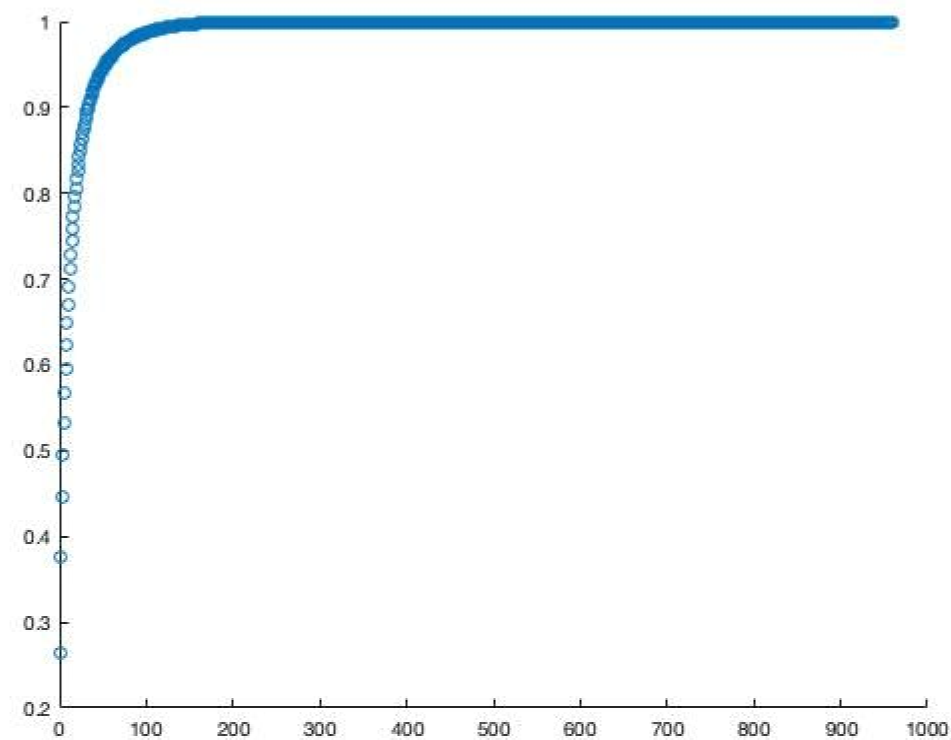


3a)


5 10 15 20 25 30
5 10 15 20 25 30
5 10 15 20 25 30
5 10 15 20 25 30
5 10 15 20 25 30
5 10 15 20 25 30
5 10 15 20 25 30
5 10 15 20 25 30
5 10 15 20 25 30
5 10 15 20 25 30

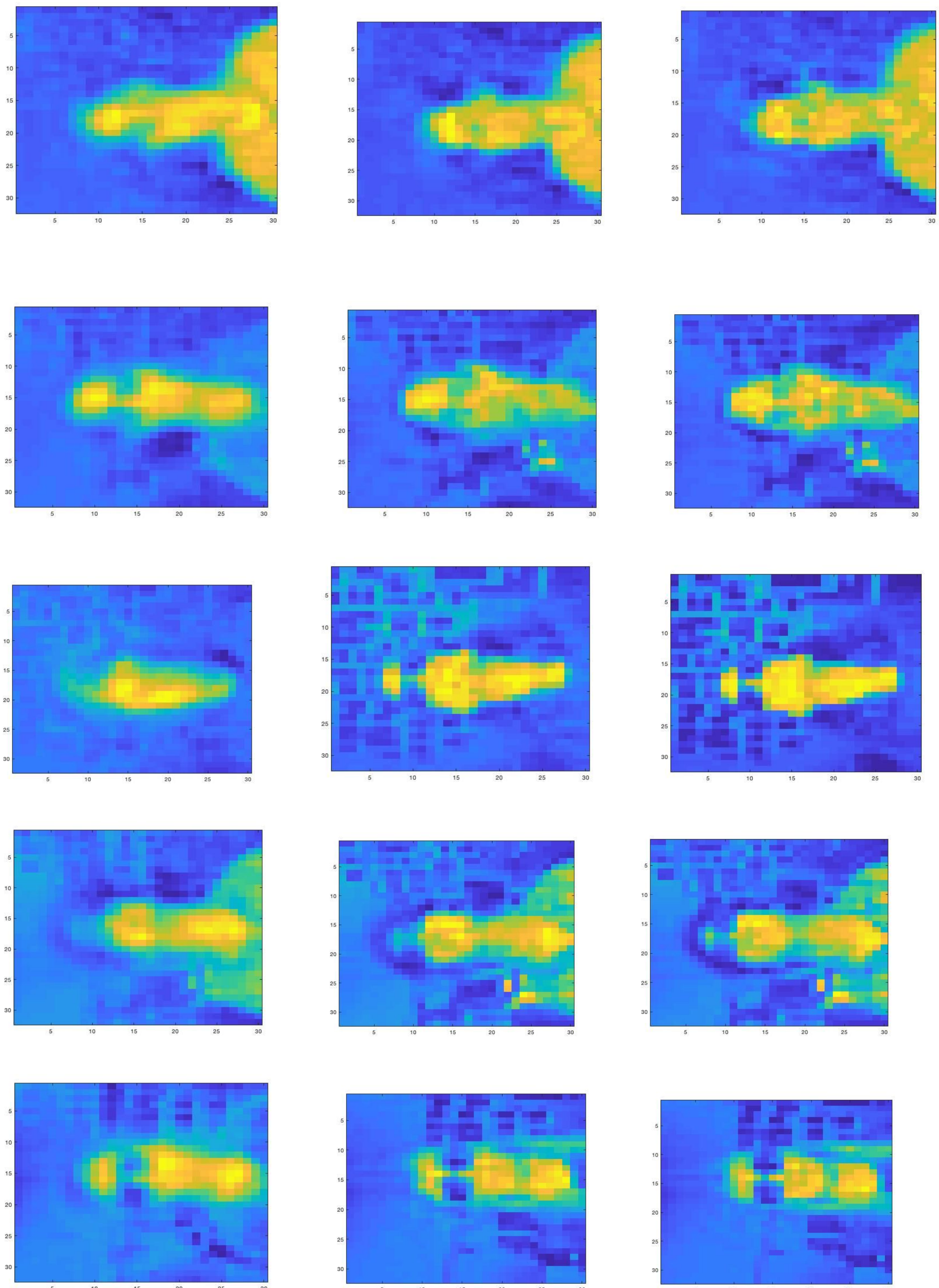
3b)

k	error
1	0.0887
3	0.2339
5	0.4274
7	0.4274



Choose K = 34

3c)



The reconstructed images are more detailed and less pixelated when the value of k is higher. The higher k is, the higher the picture quality is. This is because the original x vector loses information when it is projected to a lower dimension, and that information cannot be regained during back projection. The smaller k is, the lower the dimension that x gets projected to, and thus the more information is lost - resulting in lower picture quality.