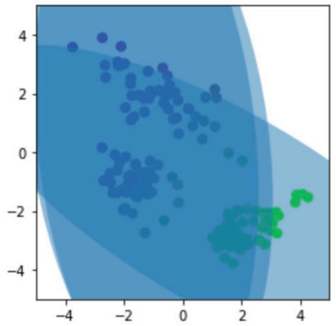
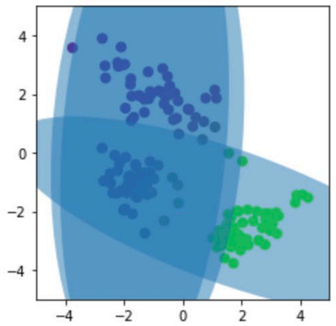
Zhao Nailin 1003003 HW3

1.

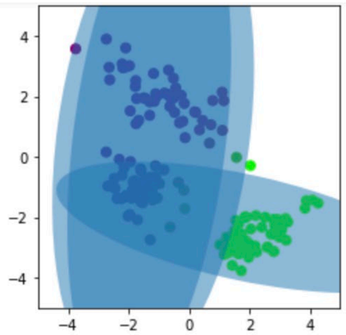
```
Threshold:  1
```



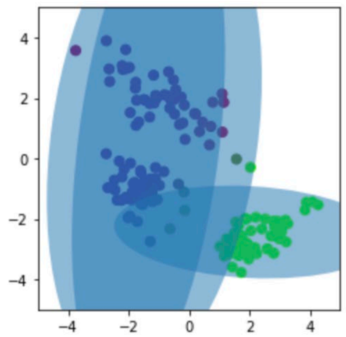
```
Likelihood:  -569.9078422712588
Diff:  inf
```



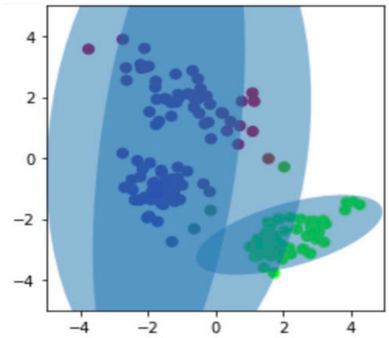
```
Likelihood:  -558.0367501491277
Diff:  11.871092122131131
```



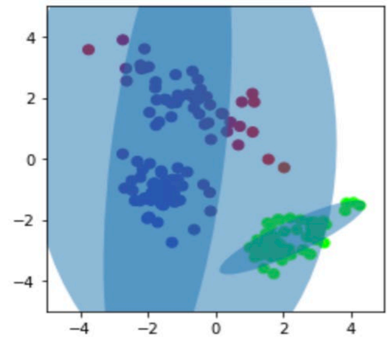
```
Likelihood:  -552.0453741917975
Diff:  5.991375957330206
```



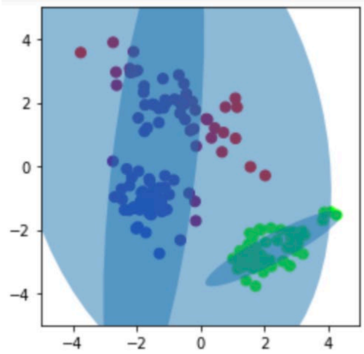
```
Likelihood:  -543.5782613458849
Diff:  8.467112845912538
```



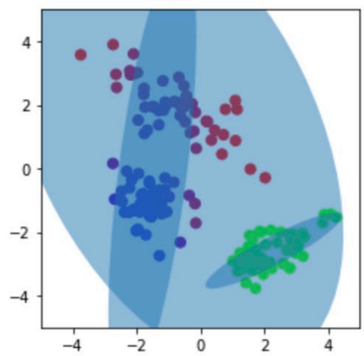
Likelihood: -528.1539581170985
Diff: 15.42430322878647



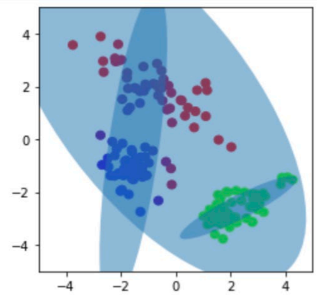
Likelihood: -516.7825467019251
Diff: 11.371411415173384



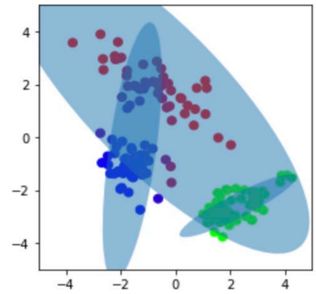
Likelihood: -513.0218062886651
Diff: 3.7607404132600095



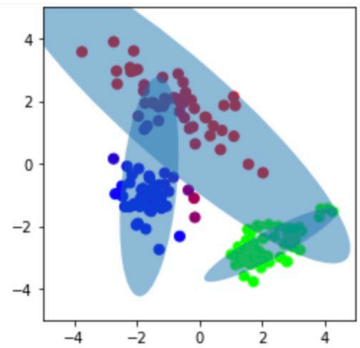
Likelihood: -509.11965953815053
Diff: 3.9021467505145324



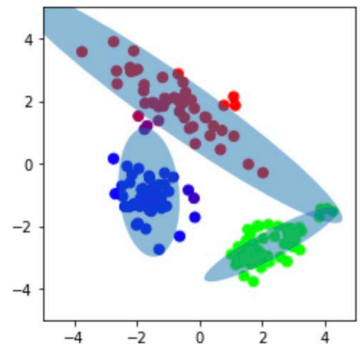
```
Likelihood:  -504.56227427994867
Diff:  4.557385258201862
```



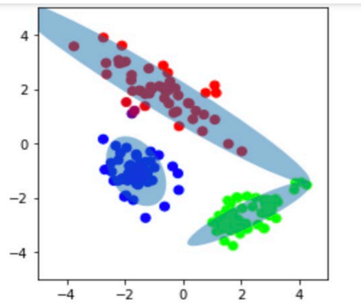
```
Likelihood:  -498.1708896815768
Diff:  6.391384598371872
```



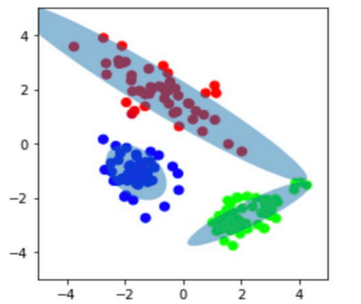
```
Likelihood:  -486.8543528577504
Diff:  11.316536823826368
```



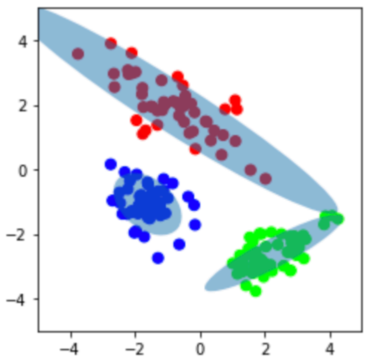
```
Likelihood:  -470.32012495477136
Diff:  16.53422790297907
```



```
Likelihood:  -464.62476830676417
Diff:  5.695356648007191
```

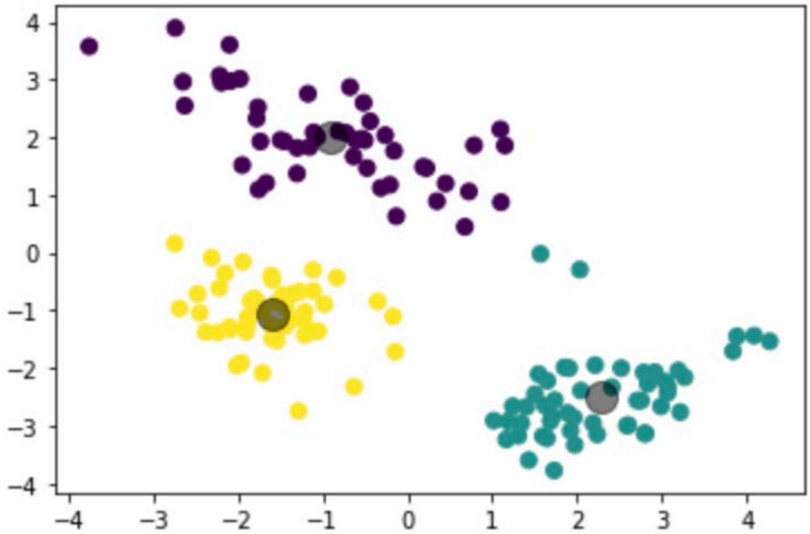


```
Likelihood:  -463.4186346944744
Diff:  1.2061336122897615
```



```
Likelihood:  -463.3425266051188
Diff:  0.07610808935561408
Satisfied threshold of  1
```

<matplotlib.collections.PathCollection at 0x7f8f6b8832b0>



2. a) $H(p,q) = -\sum_i p_i \log q_i$

$= -(\frac{1}{8}\log_2\frac{1}{4} + \frac{1}{2}\log_2\frac{1}{8} + \frac{1}{8}\log_2\frac{1}{8} + \frac{1}{8}\log_2\frac{1}{4} + \frac{1}{8}\log_2\frac{1}{4})$

$= -(-\frac{1}{4} - \frac{3}{2} - \frac{3}{8} - \frac{1}{4} - \frac{1}{4})$

$= \frac{21}{8}$ bits

$H(q,p) = -\sum_i q_i \log p_i$

$= -(\frac{1}{4}\log_2\frac{1}{8} + \frac{1}{8}\log_2\frac{1}{2} + \frac{1}{8}\log_2\frac{1}{8} + \frac{1}{4}\log_2\frac{1}{8} + \frac{1}{4}\log_2\frac{1}{8})$

$= -(-\frac{3}{4} - \frac{1}{8} - \frac{3}{8} - \frac{3}{4} - \frac{3}{4})$

$= \frac{22}{8} = \frac{11}{4}$ bits

$\because \frac{11}{4} \neq \frac{21}{8}$

$\therefore H(p,q) \neq H(q,p)$

b) $H(p) = -\sum_i p_i \log_2 p_i$

$= -(\frac{1}{8}\log_2\frac{1}{8} + \frac{1}{2}\log_2\frac{1}{2} + \frac{1}{8}\log_2\frac{1}{8} + \frac{1}{8}\log_2\frac{1}{8} + \frac{1}{8}\log_2\frac{1}{8})$

$= -(-\frac{3}{8} - \frac{1}{2} - \frac{3}{8} - \frac{3}{8} - \frac{3}{8})$

$= 2$ bits

$H(q) = -\sum_i q_i \log_2 q_i$

$= -(\frac{1}{4}\log_2\frac{1}{4} + \frac{1}{8}\log_2\frac{1}{8} + \frac{1}{8}\log_2\frac{1}{8} + \frac{1}{4}\log_2\frac{1}{4} + \frac{1}{4}\log_2\frac{1}{4})$

$= -(-\frac{1}{2} - \frac{3}{8} - \frac{3}{8} - \frac{1}{2} - \frac{1}{2})$

$= \frac{9}{4}$ bits

c) $D_{KL}(p|q) = H(p,q) - H(p,p)$

$= \frac{21}{8} - 2$

$= \frac{21}{8} - \frac{16}{8} = \frac{5}{8}$

3. a) $X = U\Sigma V^T = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$

$X^T = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$

$X^T X = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$

$= \begin{bmatrix} 1+1+0 & 1+0+0 \\ 1+0+0 & 1+0+1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$

$X^T X = V \Sigma^T \Sigma V^T$

$\begin{vmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{vmatrix} = 0$

$\therefore (2-\lambda)^2 - 1 = 0$

$(2-\lambda)^2 = 1$

$2-\lambda = \pm 1$

$\therefore \lambda_1 = 1, \quad \lambda_2 = 3$

When $\lambda_1 = 1$, $\begin{cases} 2x + y = x & ① \\ x + 2y = y & ② \end{cases}$

$\therefore x + y = 0$

$x = -y$

$\therefore v_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

When $\lambda_2 = 3$, $\begin{cases} 2x + y = 3x & ③ \\ x + 2y = 3y & ④ \end{cases}$

$\therefore x = y$

$\therefore v_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

$\therefore V = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$

$\Sigma = \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$

$\begin{vmatrix} 2-\lambda & 1 & 1 \\ 1 & 1-\lambda & 0 \\ 1 & 0 & 1-\lambda \end{vmatrix} = 0$

$\therefore (2-\lambda) \begin{vmatrix} 1-\lambda & 0 \\ 0 & 1-\lambda \end{vmatrix} - 1 \begin{vmatrix} 1 & 0 \\ 1 & 1-\lambda \end{vmatrix} + 1 \begin{vmatrix} 1 & 1-\lambda \\ 1 & 0 \end{vmatrix} = 0$

$(2-\lambda)((1-\lambda)^2 - 0) - (1-\lambda) + (0 - (1-\lambda)) = 0$

$(2-\lambda)(1-\lambda)^2 - 1 + \lambda - 1 + \lambda = 0$

$(2-\lambda)(1 - 2\lambda + \lambda^2) - 2 + 2\lambda = 0$

$\cancel{2} - 4\lambda + 2\lambda^2 - \lambda + 2\lambda^2 - \lambda^3 - \cancel{2} + 2\lambda = 0$

$-\lambda^3 + 4\lambda^2 - 3\lambda = 0$

$\lambda^3 - 4\lambda^2 + 3\lambda = 0$

$\lambda_1 = 0$ or $\lambda \neq 0$: $\lambda^2 - 4\lambda + 3 = 0$

$(\lambda - 3)(\lambda - 1) = 0$

$$\cancel{2} - 4\lambda + 2\lambda^2 - \lambda + 2\lambda^2 - \lambda^3 - \cancel{2} + 2\lambda = 0$$
$$-\lambda^3 + 4\lambda^2 - 3\lambda = 0$$
$$\lambda^3 - 4\lambda^2 + 3\lambda = 0$$
$$\lambda_1 = 0 \text{ or } \lambda \neq 0 : \lambda^2 - 4\lambda + 3 = 0$$
$$(\lambda - 3)(\lambda - 1) = 0$$

$$\therefore \lambda_1 = 0, \lambda_2 = 1, \lambda_3 = 3$$

When $\lambda_1 = 0$, $\begin{cases} 2x + y + z = 0 \\ x + y + 0 = 0 \\ x + 0 + z = 0 \end{cases}$

$$\therefore x = -z, \ y = z$$

$$\therefore U_1 = \begin{bmatrix} -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}$$

When $\lambda_2 = 1$, $\begin{cases} 2x + y + z = x \\ x + y + 0 = y \\ x + 0 + z = z \end{cases}$

$$\therefore x = 0, \ y = -z$$

$$\therefore U_2 = \begin{bmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$$

When $\lambda_3 = 3$, $\begin{cases} 2x + y + z = 3x \\ x + y + 0 = 3y \\ x + 0 + z = 3z \end{cases}$

$$\therefore x = -z$$
$$y = z$$

$$\therefore U_3 = \begin{bmatrix} \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{bmatrix}$$

$$\therefore U = \begin{bmatrix} \frac{2}{\sqrt{6}} & 0 & -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \end{bmatrix}$$

b) $V^T = V^{-1}$ ($\because$ orthogonal)

(Prove: $(XV)^T = (XV)^{-1}$)

$X = U\Sigma V^T$

$\therefore XV = U\Sigma V^T V$

$XV = U\Sigma I = U\Sigma$ ($\because V^T V = I$)

$\because$ U and V are orthogonal.

XV is orthogonal. (shown)

```python
import numpy as np
from sklearn import decomposition
from sklearn import datasets

X = datasets.load_diabetes().data

from sklearn.preprocessing import scale
scaled_X = scale(X)
U, S, V = np.linalg.svd(scaled_X)
print('V matrix' + '\n', V)
print("Singular values" + '\n', S)

print("3 most important components: ")
pca = decomposition.PCA(n_components=3)
comp = pca.fit_transform(scaled_X)
print(comp[:10])
```

```
V matrix
 [[-0.21643101 -0.18696711 -0.3031625  -0.2717397  -0.34325493 -0.35186062
   0.28243639 -0.42883325 -0.37861731 -0.32218282]
 [ 0.04437151 -0.38654811 -0.15628061 -0.13825564  0.57302669  0.45593985
   0.50624287 -0.06818423 -0.0261893  -0.0849466 ]
 [ 0.49466811 -0.10685833  0.1675317   0.51356804 -0.0685867  -0.26969438
   0.38602787 -0.38068121  0.0636315   0.27684271]
 [-0.4140095  -0.67986052  0.49982533 -0.01966734 -0.06839533 -0.16777384
  -0.07602005  0.0079212   0.26442742  0.08708624]
 [-0.68686389  0.37345612  0.12935936  0.48689014  0.12917415  0.11673143
   0.24499115 -0.14364377 -0.1516611   0.03138792]
 [ 0.2258505  -0.04173103  0.4031419   0.27276274 -0.00540864  0.1332572
  -0.1063716   0.0339454  -0.17873005 -0.80506447]
 [-0.10953821 -0.06760551 -0.51985787  0.32064908  0.07364908 -0.23054011
  -0.00753445  0.07123619  0.64731345 -0.35727279]
 [ 0.01493468  0.44293966  0.39294187 -0.47736435  0.12941351 -0.19131121
   0.32463641 -0.18058834  0.44966002 -0.1666087 ]
 [-0.00810057  0.00210552 -0.04237751 -0.0271941   0.04203984  0.35931549
  -0.48124771 -0.77381656  0.18945947  0.01527381]
 [-0.00326309 -0.00366069 -0.00824809  0.00322111 -0.70977447  0.56319605
   0.31744413  0.09059464  0.26446735 -0.0026109 ]]
Singular values
 [42.17466853 25.68276971 23.08755816 20.55043949 17.10806903 16.32182255
 15.39999097 13.84514267  5.88365535  1.94518745]
3 most important components:
[[ 0.58720767 -1.94682793  0.58923299]
 [-2.83161209  1.37208454  0.02791506]
 [ 0.27214757 -1.63489803  0.73927034]
 [ 0.04931005  0.38225333 -2.01303697]
 [-0.75645071  0.81196754 -0.05725853]
 [-3.96635524 -0.38105927 -0.33738317]
 [-1.99378667 -0.80553831 -0.71219915]
 [ 2.07586704  1.82792114  0.52492352]
 [ 0.60303259 -0.88125266 -0.07671973]
 [-0.21215262 -0.49290431 -0.81436321]]
```

5. a) $E[r_t] = \mu = \phi_0 + \phi_1 \mu + \phi_2 \mu$

$$\mu - \phi_1 \mu - \phi_2 \mu = \phi_0$$

$$\mu(1 - \phi_1 - \phi_2) = \phi_0$$

$$E[r_t] = \mu = \frac{\phi_0}{1 - \phi_1 - \phi_2}$$ (shown)

b) $$\rho(1) = \frac{\sum_{t=2}^{T} (r_t - \bar{r})(r_{t-1} - \bar{r})}{\sum_{t=1}^{T} (r_t - \bar{r})^2}$$

$$\hat{\rho}(s) = \frac{\sum_{t=s+1}^{T} (r_t - \bar{r})(r_{t-s} - \bar{r})}{\sum_{t=1}^{T} (r_t - \bar{r})^2}$$

From part a): $\phi_0 = \mu(1 - \phi_1 - \phi_2)$

$$r_t = \mu(1 - \phi_1 - \phi_2) + \phi_1 r_{t-1} + \phi_2 r_{t-2} + a_t$$

$$r_t = \mu - \phi_1 \mu - \phi_2 \mu + \phi_1 r_{t-1} + \phi_2 r_{t-2} + a_t$$