Exercise 1

a)

$$x^T A x = tr[A x x^T]$$

$(1\times d \;\; d\times d)\; d\times 1$ $\quad$ $(d\times d \;\; d\times 1)\; 1\times d$

$1\times d \;\; d\times 1$ $\quad$ $d\times 1 \;\; 1\times d$

1×1 $\quad$ $d\times d$

$$\begin{array}{ccc} x^T & A & x \\ [x_0 \;\; x_1 \; \dots \; x_d] & \begin{bmatrix} a_{00} & a_{01} & \dots & a_{0d} \\ a_{10} & a_{11} & \dots & a_{1d} \\ \vdots & & \ddots & \\ a_{d0} & a_{d1} & & a_{dd} \end{bmatrix} & \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_d \end{bmatrix} \end{array}$$

$$\left[x_0 \sum_{i=0}^{d} x_i a_{i0} + x_1 \sum_{i=0}^{d} x_i a_{i1} + x_d \sum_{i=0}^{d} x_i a_{id}\right]$$

$$\sum_{j=0}^{d} \left\{ x_j \sum_{i=0}^{d} (x_i a_{ij}) \right\} = x^T A x$$

$$\begin{array}{ccc} A & x & x^T \\ \begin{bmatrix} a_{00} & a_{01} & \dots & a_{0d} \\ a_{10} & a_{11} & \dots & a_{1d} \\ \vdots & & \ddots & \\ a_{d0} & a_{d1} & & a_{dd} \end{bmatrix} & \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_d \end{bmatrix} & [x_0 \;\; x_1 \; \dots \; x_d] \end{array}$$

$$\underset{d\times 1}{\begin{bmatrix} \sum_{i=0}^{d} a_{0i} x_i \\ \sum_{i=0}^{d} a_{1i} x_i \\ \vdots \\ \sum_{i=0}^{d} a_{di} x_i \end{bmatrix}} \underset{1\times d}{[x_0 \;\; x_1 \; \dots \; x_d]} = \begin{bmatrix} x_0 \sum_{i=0}^{d} a_{0i} x_i & x_1 \sum_{i=0}^{d} a_{0i} x_i & \dots & x_d \sum_{i=0}^{d} a_{0i} x_i \\ x_0 \sum_{i=0}^{d} a_{1i} x_i & x_1 \sum_{i=0}^{d} a_{1i} x_i & \dots & x_d \sum_{i=0}^{d} a_{1i} x_i \\ \vdots & & \ddots & \\ x_0 \sum_{i=0}^{d} a_{di} x_i & x_1 \sum_{i=0}^{d} a_{di} x_i & \dots & x_d \sum_{i=0}^{d} a_{di} x_i \end{bmatrix}$$

diagonal

$$\sum_{j=0}^{d} x_j \left\{ \sum_{i=0}^{d} (a_{ji} x_i) \right\} = tr[A x x^T]$$

equivalent

b)

b) $p(D|\Sigma) = \prod_{n=1}^{N} \left\{ \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \cdot \exp\left\{-\frac{1}{2}(x_n-\mu)^T\Sigma^{-1}(x_n-\mu)\right\}\right\}$

const w/respect to n, repeated mult = exponentiation

repeated ~~exp~~ mult goes to addition in exponent

$$= \left(\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}\right)^N \cdot \exp\left\{\sum_{n=1}^{N} \left\{-\frac{1}{2}(x_n-\mu)^T\Sigma^{-1}(x_n-\mu)\right\}\right\}$$

$$= \quad '' \quad '' \quad -\frac{1}{2}\sum_{n=1}^{N}\left\{(x_n-\mu)^T\Sigma^{-1}(x_n-\mu)\right\}$$

in form $x^TAx = tr[Axx^T]$ from part a

$$= \left(\frac{1}{(2\pi)^{Nd/2}|\Sigma|^{N/2}}\right) \cdot \exp\left\{-\frac{1}{2}\sum_{n=1}^{N} tr\left[\Sigma^{-1}(x_n-\mu)(x_n-\mu)^T\right]\right\}$$

$$\cdot \exp\left\{-\frac{1}{2} tr\left[\sum_{n=1}^{N}\Sigma^{-1}(x_n-\mu)(x_n-\mu)^T\right]\right\}$$

$$\left(\frac{1}{(2\pi)^{Nd/2}} \cdot |\Sigma|^{-N/2}\right) \cdot \exp\left\{-\frac{1}{2} tr\left[\Sigma^{-1}\sum_{n=1}^{N}(x_n-\mu)(x_n-\mu)^T\right]\right\}$$

c)

c) $A = \Sigma^{-1}\tilde{\Sigma} \rightarrow A\tilde{\Sigma}^{-1} = \Sigma^{-1} \rightarrow |A\tilde{\Sigma}^{-1}| = |\Sigma^{-1}|$

$$|\Sigma^{-1}| = \frac{1}{|\tilde{\Sigma}|}\prod_{i=1}^{d}\lambda_i = |\tilde{\Sigma}^{-1}|\prod_{i=1}^{d}\lambda_i = \prod_{i=1}^{d}\lambda_i|\tilde{\Sigma}^{-1}|$$

$$-\frac{1}{2}tr\left[\Sigma^{-1}\sum_{n=1}^{N}(x_n-\mu)(x_n-\mu)^T\right]$$

$$= -\frac{1}{2}tr\left[\Sigma^{-1}\cdot\frac{N}{N}\cdot\sum_{n=1}^{N}(x_n-\mu)(x_n-\mu)^T\right]$$

$$= -\frac{1}{2}tr\left[\Sigma^{-1}\cdot N\cdot\tilde{\Sigma}\right] = -\frac{N}{2}tr[A] = -\frac{N}{2}\sum_{i=1}^{d}\lambda_i$$

$$P(D|\Sigma) = \frac{1}{(2\pi)^{Nd/2}|\tilde{\Sigma}|^{N/2}}\left(\prod_{i=1}^{d}\lambda_i\right)^{N/2}\cdot\exp\left\{-\frac{N}{2}\sum_{i=1}^{d}\lambda_i\right\}$$

d)

d) $\arg\max_{\lambda} p(D|\Sigma)$

$= \arg\min_{\lambda} -\ln(p(D|\Sigma))$

$= \arg\min_{\lambda} -\ln\left\{\frac{1}{2\pi^{N/2}|\tilde{\Sigma}|^{N/2}}\left(\prod_{i=1}^{d}\lambda_i\right)^{N/2}\exp\left\{\frac{-N}{2}\sum_{i=1}^{d}\lambda_i\right\}\right\}$

$= \arg\min_{\lambda} -\ln\left\{\frac{1}{2\pi^{N/2}|\tilde{\Sigma}|^{N/2}}\left(\prod_{i=1}^{d}\lambda_i\right)^{N/2}\right\} - \frac{N}{2}\sum_{i=1}^{d}\lambda_i$

$\arg\min_{\lambda} -\sum_{i=1}^{d}\ln\left\{\frac{1}{2\pi^{N/2}|\tilde{\Sigma}|^{N/2}}\lambda_i^{N/2}\right\} - \frac{N}{2}\sum_{i=1}^{d}\lambda_i$

$\nabla_{\lambda_i}\left(-\sum_{i=1}^{d}\ln\left\{\frac{1}{2\pi^{N/2}|\tilde{\Sigma}|^{N/2}}\lambda_i^{N/2}\right\} - \frac{N}{2}\sum_{i=1}^{d}\lambda_i\right) = 0$

$-\frac{N}{2}\ln(k\lambda_i) - \frac{N}{2}\lambda_i = 0$

$\lambda_i = 1 \ \forall_i$

e)

e) From c, we know $\Sigma^{-1} = \prod_{i=1}^{d}\lambda_i\tilde{\Sigma}^{-1}$

$\prod_{i=1}^{d}\lambda_i = 1$ so $\Sigma^{-1} = \tilde{\Sigma}^{-1} \rightarrow \Sigma = \tilde{\Sigma}$

f) Use gradient ascent on equation 2 by changing the values of Σ

g)

Exercise 2

a)

2a) $P_{Y|X}(C_i|\vec{x}) = \dfrac{P_{X|Y}(\vec{x}|C_i)\,P_Y(C_i)}{P_X(\vec{x})}$

$$P_{X|Y}(\vec{x}|C_i) = \frac{1}{(2\pi)^{d/2}|\Sigma_i|^{1/2}} \exp\left\{-\frac{1}{2}(\vec{x}-\vec{\mu}_i)^T\Sigma_i^{-1}(\vec{x}-\vec{\mu}_i)\right\}$$

$$-\log P_{X|Y}(\vec{x}|C_i) = \frac{1}{2}(\vec{x}-\vec{\mu}_i)^T\Sigma_i^{-1}(\vec{x}-\mu_i) + \frac{d}{2}\log 2\pi + \frac{1}{2}\log|\Sigma_i|$$

$$i^* = \arg\max_i P_{Y|X}(C_i|\vec{x}) = \arg\max_i \log(P_{Y|X}(C_i|\vec{x}))$$

$$= \arg\max_i \log\left\{\frac{P_{X|Y}(\vec{x}|C_i)\,P_Y(C_i)}{P_X(\vec{x})}\right\}$$

$$= \arg\max_i\left\{\log(P_{X|Y}(\vec{x}|C_i)) + \log P_Y(\overset{\pi_i}{C_i}) - \cancel{\log P_X(\vec{x})}\right\}$$

$$i^* = \arg\max_i\left\{-\frac{1}{2}(\vec{x}-\mu_i)^T\Sigma_i^{-1}(x-\mu_i) - \frac{1}{2}\log|\Sigma_i| + \log\pi_i\right\}$$

$$-\frac{1}{2}(\vec{x}-\mu_1)^T\Sigma_1^{-1}(x-\mu_1) + \log\pi_1 - \frac{1}{2}\log|\Sigma_1| \underset{C_0}{\overset{C_1}{\gtrless}} -\frac{1}{2}(\vec{x}-\mu_0)\Sigma_0^{-1}(\vec{x}-\mu_0) + \log\pi_0 - \frac{1}{2}\log|$$

b)

mu_1:
```
[[0.44080734]
 [0.43871359]]
```

mu_0:
```
[[0.48249575]
 [0.4864399 ]]
```

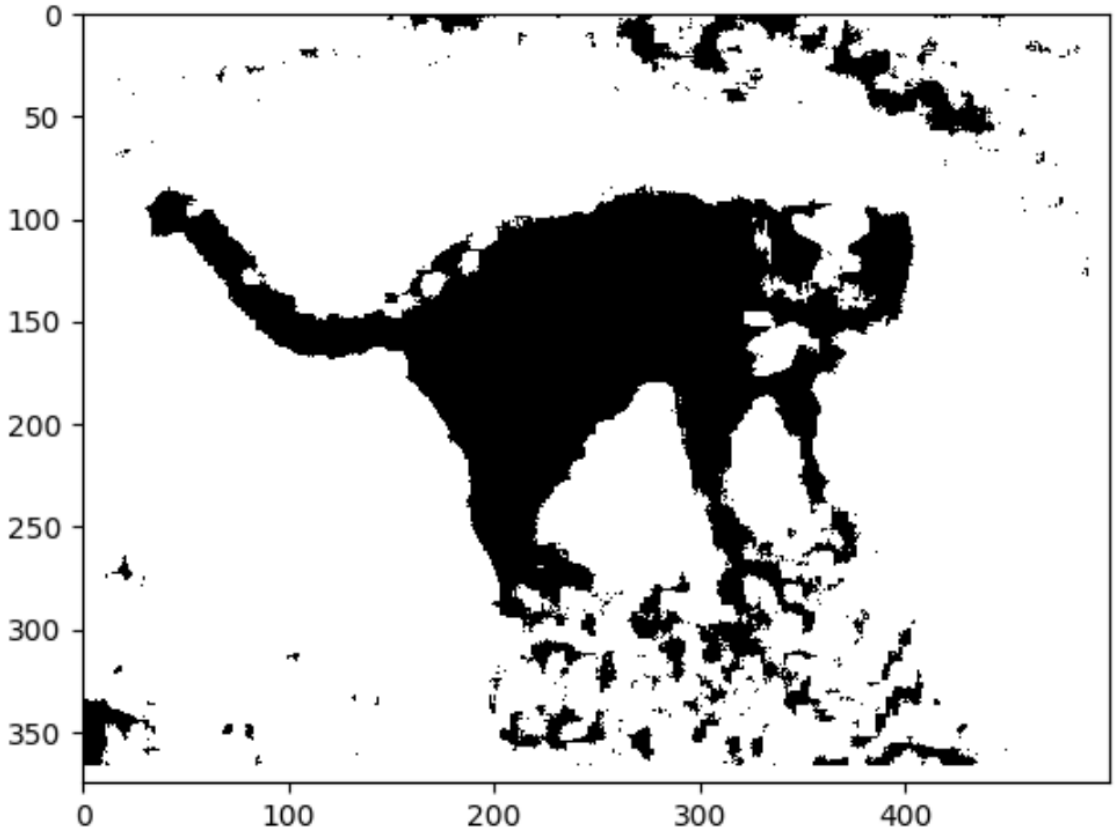
sigma_1:
```
[[0.04305652 0.03533616]
 [0.03533616 0.042466  ]]
```

sigma_0:
```
[[0.06447725 0.03691294]
 [0.03691294 0.06622764]]
```

pi_1:
0.171349288935137

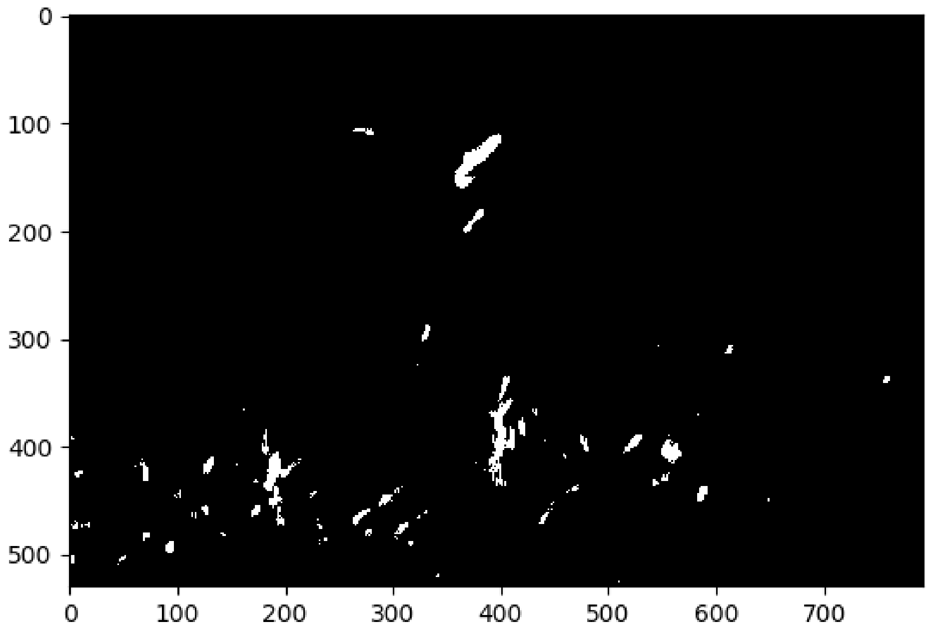
pi_0:
0.828650711064863

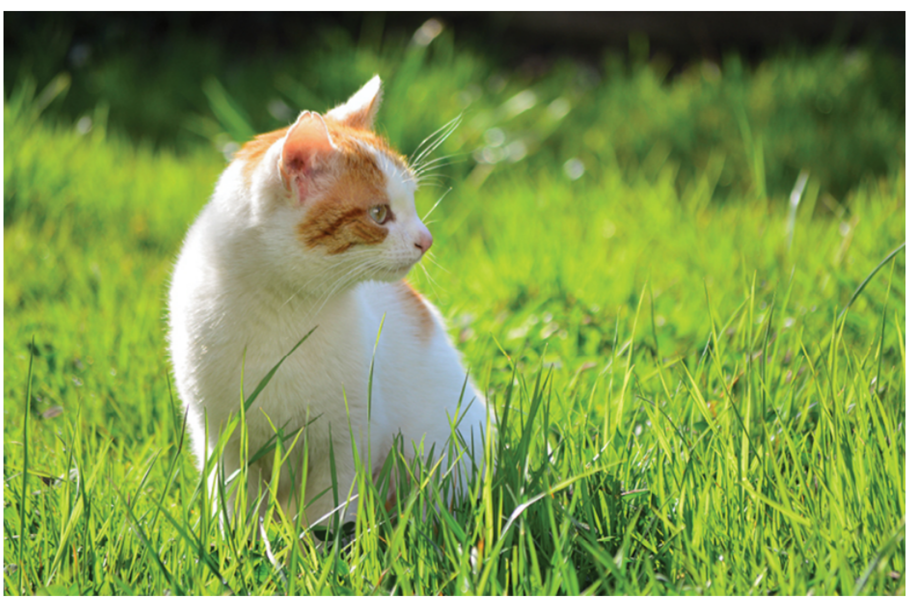
c)



d) MAE: 0.20

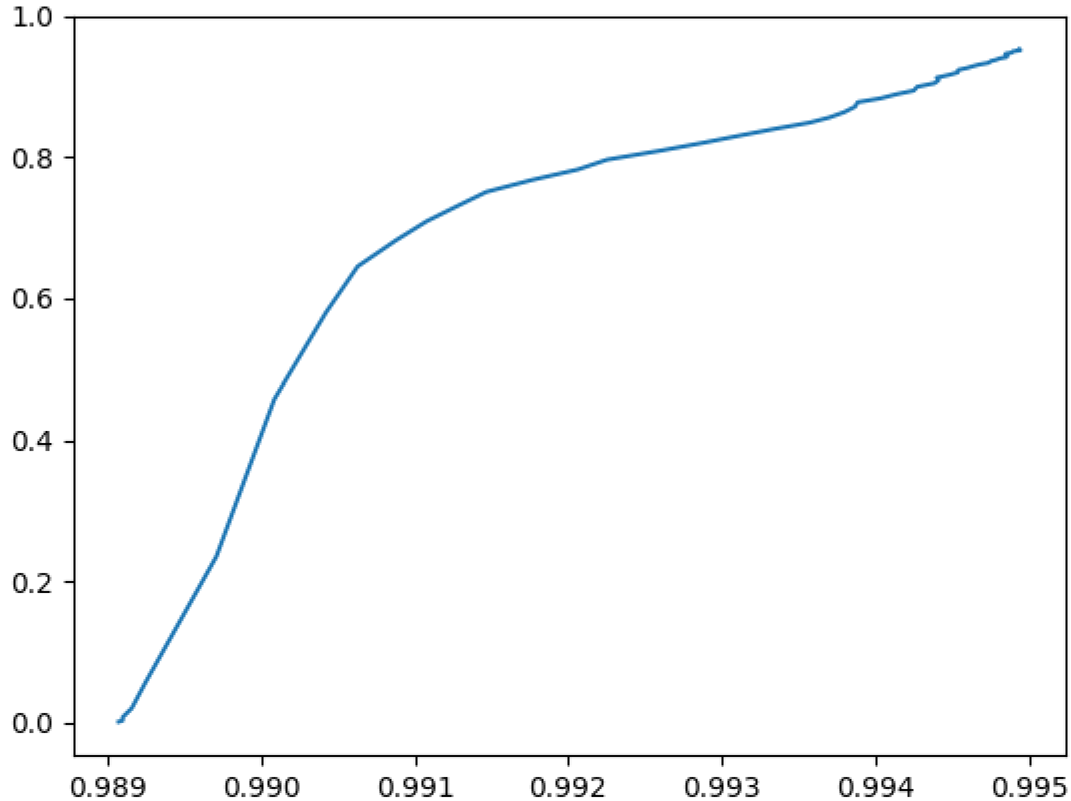
e) Did not fit well, data was probably overfit to original cat picture.



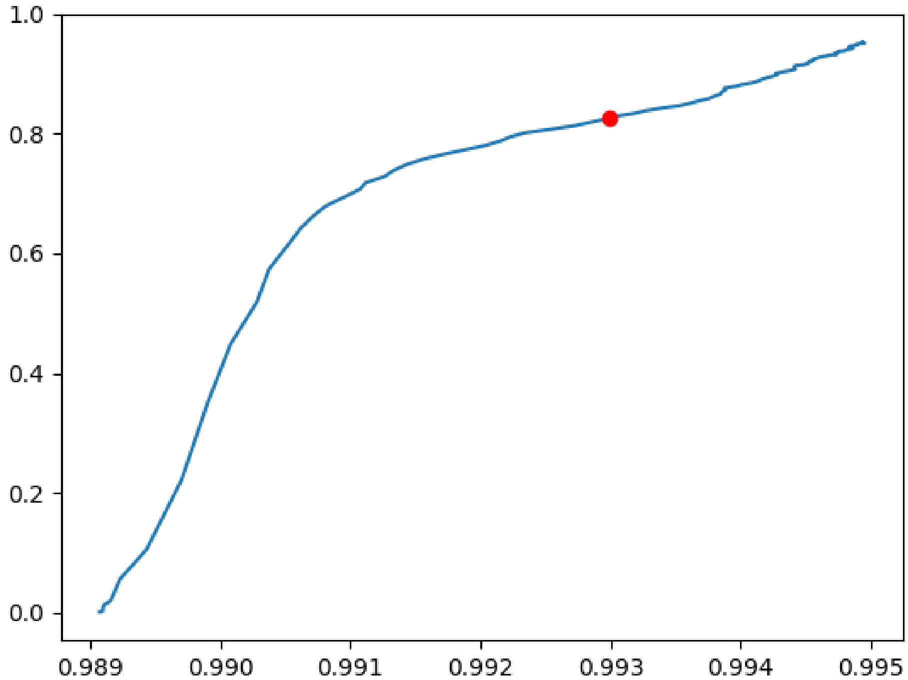


Exercise 3

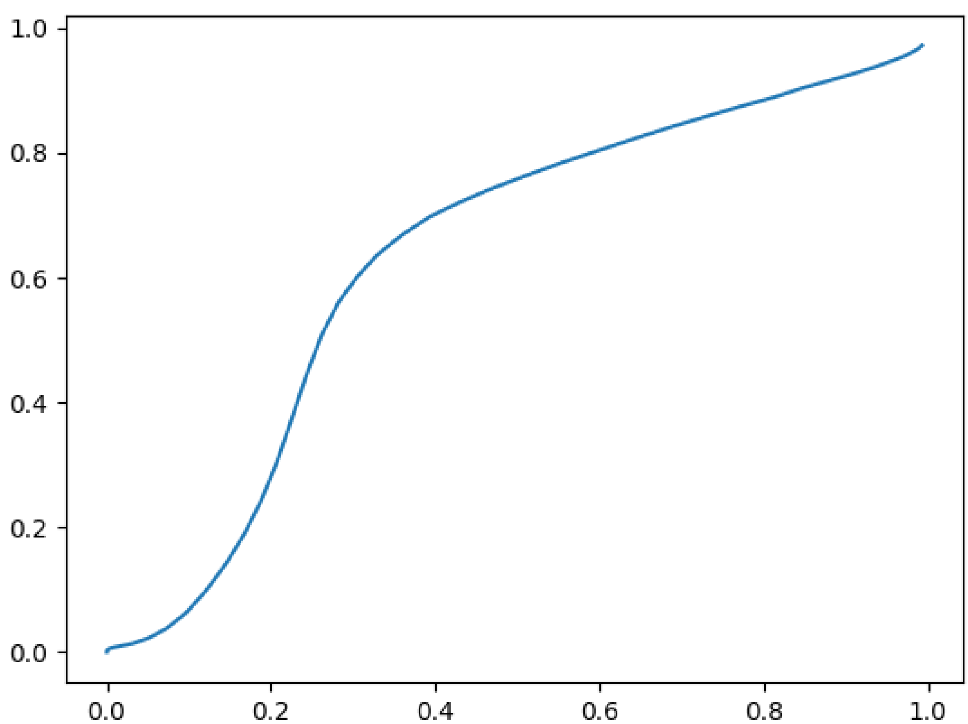
a)
We used tau=1 in exercise 2
b)



c)


1.0
0.8
0.6
0.4
0.2
0.0
0.989
0.990
0.991
0.992
0.993
0.994
0.995

d)


1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0