CS 189 HW4 Write-Up

Qingyang Zhao

TOTAL POINTS

112 / 134

QUESTION 1

1 Getting Started 4 / 4

✓ + 2 pts Part A

✓ + 2 pts Part B

+ 0 pts Empty

QUESTION 2

MLE or Multivariate Gaussian 30 pts

2.1 a 10 / 10

✓ + 10 pts Correct

+ 8 pts Mostly Correct

+ 5 pts Made Progress

+ 2 pts Trivial Progress

+ 0 pts Wrong/Missing

2.2 b 10 / 10

✓ + 10 pts Correct

+ 5 pts Correct MLE for mu

+ 5 pts Correct MLE for covariance

+ 3 pts Minor error or insufficient work in finding mu

+ 3 pts Minor error or insufficient work in finding covariance

+ 0 pts Blank or Incorrect

2.3 c 10 / 10

✓ + 10 pts Correct

+ 8 pts Minor error

+ 5 pts Progress

+ 2 pts Small amount of progress

+ 0 pts Completely incorrect/blank

QUESTION 3

Tikhonov Regularization and Weighted Least Squares 40 pts

3.1 a 10 / 10

✓ + 10 pts Correct

+ 8 pts Minor Mistake / Almost Complete

+ 5 pts Reasonable Progress

+ 2 pts Minimal Effort

+ 0 pts Blank or Incorrect

3.2 b 10 / 10

✓ + 10 pts Correct

+ 8 pts Minor math error

+ 3 pts Attempted but did not reach correct form

+ 0 pts Incorrect or blank

3.3 c 0 / 0

+ 10 pts Correct - demonstrated basis change

+ 8 pts Minor Math error

+ 2 pts Only conceptual explanation

✓ + 0 pts Incorrect/Missing/Insufficient for EC

3.4 d 10 / 10

✓ + 10 pts Correct explanation, graphs correct

+ 7 pts Correct explanation, missing graphs

+ 5 pts Missing explanation, correct graphs

+ 0 pts Incorrect

3.5 e 5 / 10

+ 10 pts Correct

+ 8 pts Small Error

✓ + 5 pts Partial

+ 2 pts Some Start

+ 0 pts Empty

No graphs and code

QUESTION 4

Total Least Squares 50 pts

4.1 a 10 / 10

✓ + 10 pts Correct

+ 8 pts almost correct

+ 5 pts half way through

+ 2 pts minimal effort (for example, only the first step)

+ **0 pts** no effort

4.2 b **8 / 10**

+ **10 pts** Correct

✓ + **8 pts** **Mostly correct**

+ **5 pts** Partially Correct

+ **2 pts** Some Progress

+ **0 pts** Incorrect

You don't need to use ridge regression to find the optimal alpha. You came to the proper conclusion that w is a scaled version of the d+1 eigenvector. All you need to do is normalize this eigenvector by the last element in this vector to generate a -1 as the last entry.

4.3 c **10 / 10**

✓ + **10 pts** **Correct**

+ **0 pts** Incorrect

+ **5 pts** Partially correct

4.4 d **5 / 10**

+ **0 pts** Incorrect/Blank

+ **2 pts** Minimal Progress

✓ + **5 pts** **Good Progress**

+ **8 pts** Missing minor part or Off by a bit

+ **10 pts** Correct

4.5 e **0 / 10**

+ **10 pts** Correct

+ **8 pts** Mostly Correct

+ **6 pts** Significant Progress

+ **4 pts** Some Progress

+ **2 pts** Trivial Progress

✓ + **0 pts** **No Progress**

QUESTION 5

5 Your Own Question 10 / 10

✓ + **10 pts** **Correct**

+ **0 pts** No Answer

HW4-1

a) Names Email Address

① Wan jhun0324@berkeley.edu

Description of Team: Best Group Ever

How did I work?

Comments:

b) I certify that all solutions are entirely in my words and that I have not looked at another student's solutions. I have credited all external sources in this write up.

Qingyang Zhao

1 Getting Started 4 / 4

✓ + 2 pts Part A

✓ + 2 pts Part B

+ 0 pts Empty

HW04-2

a) $p_x(x_1, x_2, x_3 \cdots x_n; \Sigma, \mu) = \prod_{i=1}^{n} \underbrace{\frac{1}{\sqrt{(2\pi)^d |\Sigma|}}}_{a_0} e^{-\frac{1}{2}(x_i-\mu)^T \Sigma^{-1}(x_i-\mu)}$

$= a_0^n \, e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^T \Sigma^{-1}(x_i-\mu)}$

$\log p_x(x_1, x_2 \cdots x_n; \Sigma, \mu) = -\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu) - n\log\left(\sqrt{(2\pi)^d|\Sigma|}\right)$

b) $\underset{\mu, \Sigma}{\text{maximize}} \; -\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu) - n\log\left(\sqrt{(2\pi)^d|\Sigma|}\right) = f(\mu, \Sigma)$

$\frac{\partial f(\mu, \Sigma)}{\partial \mu} = -\sum_{i=1}^{n}(x_i-\mu)^T\Sigma^{-1} = 0 \qquad \mu = \frac{1}{n}\sum_{i=1}^{n} x_i$

$\frac{\partial f(\mu, \Sigma)}{\partial \Sigma} = \frac{1}{2}\sum_{i=1}^{n}\Sigma^{-1}(x_i-\mu)(x_i-\mu)^T\Sigma^{-1} - \frac{1}{2}n\cdot(\Sigma)^{-1} = 0$

$\Sigma = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)^T = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \frac{1}{n}\sum_{i=1}^{n}x_i\right)\left(x_i - \frac{1}{n}\sum_{i=1}^{n}x_i\right)^T$

c)
```
import numpy.matlib
mu= 1/(len(samples))* sum(samples)
mu_mat = np.matlib(mu, len(samples),1)
var = 1/((len(samples)) * np.dot(samples.T-mu.T, samples - mu)
```

	Estimate_variance	μ
$\Sigma_1 = \begin{bmatrix} 20 & 0 \\ 0 & 10 \end{bmatrix}$	$\begin{bmatrix} 24.79 & -2.17 \\ -2.17 & 11.71 \end{bmatrix}$	$\begin{bmatrix} 13.75 \\ 5.32 \end{bmatrix}$
$\Sigma_2 = \begin{bmatrix} 20 & 14 \\ 14 & 10 \end{bmatrix}$	$\begin{bmatrix} 22.27 & 15.11 \\ 15.11 & 10.48 \end{bmatrix}$	$\begin{bmatrix} 14.91 \\ 4.90 \end{bmatrix}$
$\Sigma_3 = \begin{bmatrix} 20 & -14 \\ -14 & 10 \end{bmatrix}$	$\begin{bmatrix} 18.27 & -12.65 \\ -12.65 & 8.95 \end{bmatrix}$	$\begin{bmatrix} 15.44 \\ 4.66 \end{bmatrix}$

2.1 a 10 / 10

- ✓ + **10 pts** Correct
- + **8 pts** Mostly Correct
- + **5 pts** Made Progress
- + **2 pts** Trivial Progress
- + **0 pts** Wrong/Missing

HW04-2

a) $$P_x(x_1, x_2, x_3 \cdots x_n; \Sigma, \mu) = \prod_{i=1}^{n} \underbrace{\frac{1}{\sqrt{(2\pi)^d |\Sigma|}}}_{a_0} e^{-\frac{1}{2}(x_i-\mu)^T \Sigma^{-1}(x_i-\mu)}$$

$$= a_0^n \, e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^T \Sigma^{-1}(x_i-\mu)}$$

$$\log P_x(x_1, x_2 \cdots x_n; \Sigma, \mu) = -\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^T \Sigma^{-1}(x_i-\mu) - n\log\left(\sqrt{(2\pi)^d |\Sigma|}\right)$$

b) $$\underset{\mu, \Sigma}{\text{maximize}} \; -\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^T \Sigma^{-1}(x_i-\mu) - n\log\left(\sqrt{(2\pi)^d |\Sigma|}\right) = f(\mu, \Sigma)$$

$$\frac{\partial f(\mu, \Sigma)}{\partial \mu} = -\sum_{i=1}^{n}(x_i-\mu)^T \Sigma^{-1} = 0 \qquad \mu = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\frac{\partial f(\mu, \Sigma)}{\partial \Sigma} = \frac{1}{2}\sum_{i=1}^{n} \Sigma^{-1}(x_i-\mu)(x_i-\mu)^T \Sigma^{-1} - \frac{1}{2}n \cdot (\Sigma)^{-1} = 0$$

$$\Sigma = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)^T = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \frac{1}{n}\sum_{i=1}^{n} x_i\right)\left(x_i - \frac{1}{n}\sum_{i=1}^{n} x_i\right)^T$$

c)
```
import numpy.matlib
mu= 1/(len(samples))* sum (samples)
mu_mat = np.matlib(mu, len(samples), 1)
var = 1/((len(samples)) * np.dot (samples.T-mu.T, samples - mu)
```

	Estimate_variance	μ
$\Sigma_1 = \begin{bmatrix} 20 & 0 \\ 0 & 10 \end{bmatrix}$	$\begin{bmatrix} 24.79 & -2.17 \\ -2.17 & 11.71 \end{bmatrix}$	$\begin{bmatrix} 13.75 \\ 5.32 \end{bmatrix}$
$\Sigma_2 = \begin{bmatrix} 20 & 14 \\ 14 & 10 \end{bmatrix}$	$\begin{bmatrix} 22.27 & 15.11 \\ 15.11 & 10.48 \end{bmatrix}$	$\begin{bmatrix} 14.91 \\ 4.90 \end{bmatrix}$
$\Sigma_3 = \begin{bmatrix} 20 & -14 \\ -14 & 10 \end{bmatrix}$	$\begin{bmatrix} 18.27 & -12.65 \\ -12.65 & 8.95 \end{bmatrix}$	$\begin{bmatrix} 15.44 \\ 4.66 \end{bmatrix}$

2.2 b **10 / 10**

✓ + **10 pts** Correct

+ **5 pts** Correct MLE for mu

+ **5 pts** Correct MLE for covariance

+ **3 pts** Minor error or insufficient work in finding mu

+ **3 pts** Minor error or insufficient work in finding covariance

+ **0 pts** Blank or Incorrect

HW04-2

a) $p_x(x_1, x_2, x_3 \cdots x_n ; \Sigma, \mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} e^{-\frac{1}{2}(x_i-\mu)^T \Sigma^{-1}(x_i-\mu)}$, $a_0 = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}}$

$= a_0^n \, e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^T \Sigma^{-1}(x_i-\mu)}$

$\log p_x(x_1, x_2 \cdots x_n ; \Sigma, \mu) = -\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^T \Sigma^{-1}(x_i-\mu) - n\log\sqrt{(2\pi)^d |\Sigma|}$

b) $\underset{\mu, \Sigma}{\text{maximize}} \; -\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^T \Sigma^{-1}(x_i-\mu) - n\log(\sqrt{(2\pi)^d |\Sigma|}) = f(\mu, \Sigma)$

$\frac{\partial f(\mu, \Sigma)}{\partial \mu} = -\sum_{i=1}^{n}(x_i-\mu)^T \Sigma^{-1} = 0 \qquad \mu = \frac{1}{n}\sum_{i=1}^{n} x_i$

$\frac{\partial f(\mu, \Sigma)}{\partial \Sigma} = \frac{1}{2}\sum_{i=1}^{n} \Sigma^{-1}(x_i-\mu)(x_i-\mu)^T \Sigma^{-1} - \frac{1}{2}n \cdot (\Sigma)^{-1} = 0$

$\Sigma = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)^T = \frac{1}{n}\sum_{i=1}^{n}(x_i - \frac{1}{n}\sum_{i=1}^{n} x_i)(x_i - \frac{1}{n}\sum_{i=1}^{n} x_i)^T$

c)
```
import numpy.matlib
mu= 1/(len(samples))* sum (samples)
mu_mat = np.matlib(mu, len(samples), 1)
var = 1/((len(samples)) * np.dot (samples.T-mu.T, samples - mu)
```

	Estimate_variance	μ
$\Sigma_1 = \begin{bmatrix} 20 & 0 \\ 0 & 10 \end{bmatrix}$	$\begin{bmatrix} 24.79 & -2.17 \\ -2.17 & 11.71 \end{bmatrix}$	$\begin{bmatrix} 13.75 \\ 5.32 \end{bmatrix}$
$\Sigma_2 = \begin{bmatrix} 20 & 14 \\ 14 & 10 \end{bmatrix}$	$\begin{bmatrix} 22.27 & 15.11 \\ 15.11 & 10.48 \end{bmatrix}$	$\begin{bmatrix} 14.91 \\ 4.90 \end{bmatrix}$
$\Sigma_3 = \begin{bmatrix} 20 & -14 \\ -14 & 10 \end{bmatrix}$	$\begin{bmatrix} 18.27 & -12.65 \\ -12.65 & 8.95 \end{bmatrix}$	$\begin{bmatrix} 15.44 \\ 4.66 \end{bmatrix}$

2.3 C 10 / 10

- ✓ + 10 pts Correct
- + 8 pts Minor error
- + 5 pts Progress
- + 2 pts Small amount of progress
- + 0 pts Completely incorrect/blank

HW04 - 3

a) $Z \sim N(0,1)$

$wx + Z$ $Y \mid X = x, W = w \sim N(wx, 1)$

Density function

$$P_{Y|X,W} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y - w^T x)^2}$$

b) According to Total Bayesian Rule

Notation:

$P_{\vec{w}|\{Y_i, X_i\}}$ Denotes pdf of $\vec{w}$ given all data points

$\{X_i\}$, $\{Y_i\}$ $\{X_i, Y_i\}$ denotes set of $X_{i=1\ldots n}$, $Y_{i=1\ldots n}$, $\{X_{i=1\ldots n}, Y_{i=1\ldots n}\}$

$$P_{\vec{w}|\{Y_i\}\{X_i\}} \propto P_{\{Y_i\}|w\{X_i\}} P_w$$

$$P_{\vec{w}|\{Y_i, X_i\}} \propto \left(\prod_{i=1}^{n} e^{-\frac{1}{2}(y_i - w^T x_i)^2}\right) e^{-\frac{1}{2} w^T w}$$

$P_{\vec{w}|\{Y_i, X_i\}}$ is also a Gaussian $\propto e^{-\frac{1}{2}\left(\sum_{i=1}^{n}(y_i - w^T x_i)^2 + w^T \Sigma^{-1} w\right)}$

convert this in a matrix form.

corresponds to a generalised Gaussian form, which is

$$f(w) = c \cdot \exp\left(-\frac{1}{2} w^T A w + b^T w\right)$$

we have our model as $P_{\vec{w}|\{Y_i, X_i\}} \propto \exp\left(-\frac{1}{2}\left(w^T (X^T X + \Sigma^{-1}) w - 2y^T X w\right)\right)$

The $w \mid \{Y_i, X_i\} \sim N\left((X^T X + \Sigma^{-1})^{-1} X^T y,\ (X^T X + \Sigma^{-1})^{-1}\right)$

c) with $Z \sim (\mu_z, \Sigma_z)$

$$w \mid \{X_i, Y_i\} \sim N\left((X^T \Sigma_z^{-1} X + \Sigma^{-1})^{-1} X^T \Sigma_z^{-1} (Y - \mu_z),\ (X^T \Sigma_z^{-1} X + \Sigma^{-1})\right)$$

$Y_i \mid \{X_i\}, w \sim (\mu_z + w^T X, \Sigma_z)$

$$-\frac{1}{2}\left(w^T X^T \Sigma_z^{-1} X w - 2(Y - \mu_z)^T \Sigma_z^{-1} X w + w^T \Sigma^{-1} w\right)$$

Thus mean is $\boxed{(X^T \Sigma_z^{-1} X + \Sigma^{-1})^{-1} X^T \Sigma_z^{-1} (Y - \mu_z)}$

variance is $(X^T \Sigma_z^{-1} X + \Sigma^{-1})$

3.1 a 10 / 10

✓ + 10 pts Correct

+ 8 pts Minor Mistake / Almost Complete

+ 5 pts Reasonable Progress

+ 2 pts Minimal Effort

+ 0 pts Blank or Incorrect

HW04 - 3

a) $Z \sim N(0,1)$

$wx+Z$ $Y \mid X=x, W=w \sim N(wx, 1)$

Dencity function

$$P_{Y|X,W} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-w^Tx)^2}$$

b) According to Total Bayesian Rule

Notation: $P_{\vec{w}|\{Y_i,X_i\}}$ Denotes pdf of $\vec{w}$ given all data points

$\{X_i\}, \{Y_i\}$ $\{X_i, Y_i\}$ denotes set of $X_{i=1\cdots n}$, $Y_{i=1\cdots n}$, $\{X_{i=1\cdots n}, Y_{i=1\cdots n}\}$

$$P_{\vec{w}|\{Y_i\}\{X_i\}} \propto P_{\{Y_i\}|w\{X_i\}} P_w$$

$$P_{\vec{w}|\{Y_i,X_i\}} \propto \left(\prod_{i=1}^{n} e^{-\frac{1}{2}(y_i-w^Tx_i)^2}\right) e^{-\frac{1}{2}w^Tw}$$

$P_{\vec{w}|\{Y_i,X_i\}}$ is also a Gaussian $\propto$ $e^{-\frac{1}{2}\left(\sum_{i=1}^{n}(y_i-w^Tx_i)^2+w^T\Sigma^{-1}w\right)}$

convert this in a matrix form.

corresponds to a generalised Gaussian form, which is

$$f(w) = c \cdot \exp\left(-\frac{1}{2}w^TAw + b^Tw\right)$$

we have our model as $P_{\vec{w}|\{Y_i,X_i\}} \propto \exp\left(-\frac{1}{2}\left(w^T(X^TX+\Sigma^{-1})w - 2y^TXw\right)\right)$

The $w|\{Y_i,X_i\} \sim N\left((X^TX+\Sigma^{-1})^{-1}X^Ty,\ (X^TX+\Sigma^{-1})^{-1}\right)$

c) with $Z \sim (\mu_z, \Sigma_z)$

$$w|\{X_i,Y_i\} \sim N\left((X^T\Sigma_z^{-1}X+\Sigma^{-1})^{-1}X^T\Sigma_z^{-1}(Y-\mu_z),\ (X^T\Sigma_z^{-1}X+\Sigma^{-1})\right)$$

$Y_i|\{X_i\}, W \sim (\mu_z + w^TX, \Sigma_z)$

$$-\frac{1}{2}\left(w^TX^T\Sigma_z^{-1}Xw - 2(Y-\mu_z)^T\Sigma_z^{-1}Xw + w^T\Sigma^{-1}w\right)$$

Thus mean is $\boxed{(X^T\Sigma_z^{-1}X+\Sigma^{-1})^{-1}X^T\Sigma_z^{-1}(Y-\mu_z)}$

variance is $(X^T\Sigma_z^{-1}X+\Sigma^{-1})$

3.2 b 10 / 10

✓ + **10 pts** Correct

+ **8 pts** Minor math error

+ **3 pts** Attempted but did not reach correct form

+ **0 pts** Incorrect or blank

HW04 - 3

a) $Z \sim N(0,1)$

$wx+Z$ $Y \mid X=x, W=w \sim N(wx, 1)$

Dencity function

$$P_{Y|X,W} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-w^T x)^2}$$

b) According to Total Bayesian Rule

Notation:

$P_{\vec{w}|\{Y_i,X_i\}}$ Denotes pdf of $\vec{w}$ given all data points

$\{X_i\}, \{Y_i\}$ $\{X_i, Y_i\}$ denotes set of $X_{i=1\dots n}$, $Y_{i=1\dots n}$, $\{X_{i=1\dots n}, Y_{i=1\dots n}\}$

$$P_{\vec{w}|\{Y_i\}\{X_i\}} \propto P_{\{Y_i\}|w\{X_i\}} P_w$$

$$P_{\vec{w}|\{Y_i,X_i\}} \propto \left(\prod_{i=1}^{n} e^{-\frac{1}{2}(y_i - w^T x_i)^2}\right) e^{-\frac{1}{2} w^T w}$$

$P_{\vec{w}|\{Y_i,X_i\}}$ is also a Gaussian $\propto$ $e^{-\frac{1}{2}\left(\sum_{i=1}^{n}(y_i - w^T x_i)^2 + w^T \Sigma^{-1} w\right)}$

convert this in a matrix form.

corresponds to a generalised Gaussian form, which is

$$f(w) = C \cdot \exp\left(-\frac{1}{2} w^T A w + b^T w\right)$$

we have our model as $P_{\vec{w}|\{Y_i,X_i\}} \propto \exp\left(-\frac{1}{2}\left(w^T(X^T X + \Sigma^{-1})w - 2y^T X w\right)\right)$

The $w|\{Y_i, X_i\} \sim N\left((X^T X + \Sigma^{-1})^{-1} X^T y,\ (X^T X + \Sigma^{-1})^{-1}\right)$

c) with $Z \sim (\mu_z, \Sigma_z)$

$$w|\{X_i, Y_i\} \sim N\left((X^T \Sigma_z^{-1} X + \Sigma^{-1})^{-1} X^T \Sigma_z^{-1}(Y-\mu_z),\ (X^T \Sigma_z^{-1} X + \Sigma^{-1})\right)$$

$Y_i|\{X_i\}, W \sim (\mu_z + w^T X, \Sigma_z)$

$$-\frac{1}{2}\left(w^T X^T \Sigma_z^{-1} X w - 2(Y-\mu_z)^T \Sigma_z^{-1} X w + w^T \Sigma^{-1} w\right)$$

Thus mean is $\boxed{(X^T \Sigma_z^{-1} X + \Sigma^{-1})^{-1} X^T \Sigma_z^{-1} (Y-\mu_z)}$

variance is $(X^T \Sigma_z^{-1} X + \Sigma^{-1})$

3.3 C **0 / 0**

- \+ **10 pts** Correct - demonstrated basis change
- \+ **8 pts** Minor Math error
- \+ **2 pts** Only conceptual explanation
- ✓ + **0 pts** **Incorrect/Missing/Insufficient for EC**

H04-2)

d) As samples goes large, the confidence of the estimation increase, (variance smaller).

As variance becomes correlated, estimation becomes worse.

e) ① when sample number gets large, even though the prior are strongly correlated, the errors still good.

② when sample points are small test errors are big, while less correlated priors have small errors.

3.4 d 10 / 10

- ✓ + 10 pts Correct explanation, graphs correct
- + 7 pts Correct explanation, missing graphs
- + 5 pts Missing explanation, correct graphs
- + 0 pts Incorrect

H04-2)

d) As samples goes large, the confidence of the estimation increase, (variance smaller).

As variance becomes correlated, estimation becomes worse.

e) ① when sample number gets large, even though the prior are strongly correlated, the errors still good.

② When sample points are small test errors are big, while less correlated priors have small errors.

3.5 e 5 / 10

- + 10 pts Correct
- + 8 pts Small Error
- ✓ + 5 pts **Partial**
- + 2 pts Some Start
- + 0 pts Empty

No graphs and code

HW04 - 4

a)

Optimization Problem: $\underset{\hat{A}\,\hat{y}}{\text{minimize}} \; \|\hat{A}, \hat{y}\|_F$

s.t. $(A+\hat{A})\vec{w} = \vec{y} + \hat{y}$

$$\left[A+\hat{A}, \vec{y}+\hat{y}\right]\begin{bmatrix}\vec{w}\\-1\end{bmatrix} = 0$$

$\begin{bmatrix}\vec{w}\\-1\end{bmatrix} \in R^{d+1}$ and lies in the null space of $[A+\hat{A}, y+\hat{y}]$

Therefore, $\begin{bmatrix}\vec{w}\\-1\end{bmatrix}$ has nonzero solutions iff $\text{rank}[A+\hat{A}, \vec{y}+\hat{y}] < d+1$

(Because, $\begin{bmatrix}\vec{w}\\-1\end{bmatrix}$ lies in the null space of $[A+\hat{A}, y+\hat{y}]$, and singular matrix will result in infinite solutions)

$\text{rank}[A+\hat{A}] = d$, which means $\text{rank}[A+\hat{A}, y+\hat{y}] \geq d$

Thus, $\text{rank}[A+\hat{A}, \vec{y}+\hat{y}] = d$

b) $[A+\hat{A}, y+\hat{y}]\vec{x} = \vec{0}$ According the theorem if $[A, \vec{y}] = U\Sigma V^T$

$[A+\hat{A}, \vec{y}+\hat{y}] = U\begin{bmatrix}\Sigma_{1\dots d} & \\ & 0\end{bmatrix}V^T$

$$U\begin{bmatrix}\Sigma_{1\dots d} & \\ & 0\end{bmatrix}V^T\vec{x} = \vec{0} \qquad [\vec{u}_1\, \vec{u}_2 \cdots \vec{u}_d\, \vec{u}_{d+1}]\begin{bmatrix}\Sigma_{1\dots d} & \\ & 0\end{bmatrix}\begin{bmatrix}\vec{v}_1^T\\ \vdots \\ \vec{v}_d^T \\ \vec{v}_{d+1}^T\end{bmatrix}\vec{x} = \vec{0}$$

$\vec{x} = \alpha \vec{v}_{d+1}$

(Because $\underset{i \neq d+1}{\vec{v}_i^T} \cdot \vec{v}_{d+1} = 0$, and the last term in Σ is 0)

Thus, $\vec{w} = \alpha \vec{v}_{d+1}$

$$[A, \vec{y}]^T[A, \vec{y}] = \begin{bmatrix}\vec{v}_1 \cdots \vec{v}_d\, \vec{v}_{d+1}\end{bmatrix}\begin{bmatrix}\Sigma_d^2 & \\ & 0\end{bmatrix}\begin{bmatrix}\vec{v}_1^T\\ \vdots \\ \vec{v}_d^T \\ \vec{v}_{d+1}^T\end{bmatrix}$$

$$\begin{bmatrix}A^TA & A^T\vec{y}\\ \vec{y}^TA & \vec{y}^T\vec{y}\end{bmatrix}\begin{bmatrix}\vec{w}\\-1\end{bmatrix} = \sigma_{d+1}^2\begin{bmatrix}\vec{w}\\-1\end{bmatrix} \qquad \vec{w} = (A^TA - \sigma_{d+1}^2 I)^{-1}A^T\vec{y}$$

4.1 a 10 / 10

- ✓ + **10 pts** Correct
- + **8 pts** almost correct
- + **5 pts** half way through
- + **2 pts** minimal effort (for example, only the first step)
- + **0 pts** no effort

HW04 - 4

a)

Optimization Problem: $\underset{\hat{A}\,\hat{y}}{\text{minimize}} \ \|\hat{A}, \hat{y}\|_F$

s.t. $(A+\hat{A})\vec{\omega} = \vec{y} + \hat{y}$

$$\left[A+\hat{A}, \vec{y}+\hat{y}\right]\begin{bmatrix}\vec{\omega}\\ -1\end{bmatrix} = 0$$

$\begin{bmatrix}\vec{\omega}\\ -1\end{bmatrix} \in R^{d+1}$ and lies in the null space of $[A+\hat{A}, y+\hat{y}]$

Therefore, $\begin{bmatrix}\vec{\omega}\\ -1\end{bmatrix}$ has nonzero solutions iff $\text{rank}[A+\hat{A}, \vec{y}+\hat{y}] < d+1$

(Because, $\begin{bmatrix}\vec{\omega}\\ -1\end{bmatrix}$ lies in the null space of $[A+\hat{A}, y+\hat{y}]$, and singular matrix will result in infinite solutions)

$\text{rank}[A+\hat{A}] = d$, which means $\text{rank}[A+\hat{A}, y+\hat{y}] \geq d$

Thus, $\text{rank}[A+\hat{A}, \vec{y}+\hat{y}] = d$

b) $\left[A+\hat{A}, \vec{y}+\hat{y}\right]\vec{x} = \vec{0}$ According the theorem if $[A, \vec{y}] = U\Sigma V^T$

$[A+\hat{A}, \vec{y}+\hat{y}] = U\begin{bmatrix}\Sigma_{1\ldots d} & \\ & 0\end{bmatrix}V^T$

$$U\begin{bmatrix}\Sigma_{1\ldots d} & \\ & 0\end{bmatrix}V^T\vec{x} = \vec{0} \qquad [\vec{u}_1\, \vec{u}_2 \cdots \vec{u}_d\, \vec{u}_{d+1}]\begin{bmatrix}\Sigma_{1\ldots d} & \\ & 0\end{bmatrix}\begin{bmatrix}\vec{v}_1^T\\ \vdots \\ \vec{v}_d^T\\ \vec{v}_{d+1}^T\end{bmatrix}\vec{x} = \vec{0}$$

$\vec{x} = \alpha\vec{v}_{d+1}$

(Because $\underset{i \neq d+1}{\vec{v}_i^T} \cdot \vec{v}_{d+1} = 0$, and the last term in Σ is 0)

Thus, $\vec{\omega} = \alpha\vec{v}_{d+1}$

$$[A, \vec{y}]^T[A, \vec{y}] = \begin{bmatrix}\vec{v}_1 \cdots \vec{v}_d\ \vec{v}_{d+1}\end{bmatrix}\begin{bmatrix}\Sigma_d^2 & \\ & 0\end{bmatrix}\begin{bmatrix}\vec{v}_1^T\\ \vdots \\ \vec{v}_d^T\\ \vec{v}_{d+1}^T\end{bmatrix}$$

$$\begin{bmatrix}A^TA & A^T\vec{y}\\ \vec{y}^TA & \vec{y}^T\vec{y}\end{bmatrix}\begin{bmatrix}\vec{\omega}\\ -1\end{bmatrix} = \sigma_{d+1}^2\begin{bmatrix}\vec{\omega}\\ -1\end{bmatrix} \qquad \vec{\omega} = (A^TA - \sigma_{d+1}^2 I)^{-1}A^T\vec{y}$$

4.2 b 8 / 10

+ 10 pts Correct

✓ + 8 pts **Mostly correct**

+ 5 pts Partially Correct

+ 2 pts Some Progress

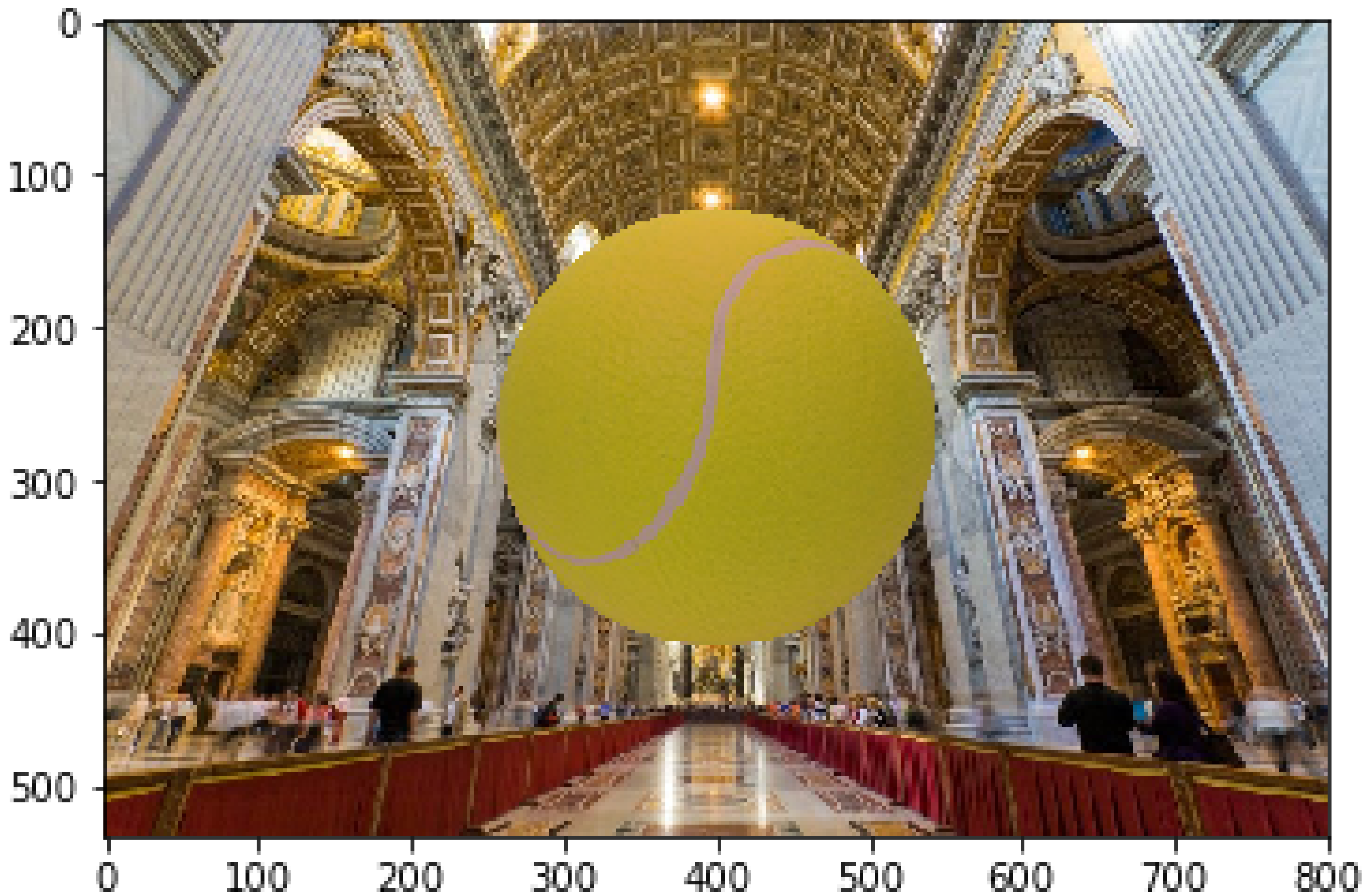
+ 0 pts Incorrect

You don't need to use ridge regression to find the optimal alpha. You came to the proper conclusion that w is a scaled version of the d+1 eigenvector. All you need to do is normalize this eigenvector by the last element in this vector to generate a -1 as the last entry.

c) $\Gamma_i = \begin{bmatrix} 202.318 & 162.420 & 149.071 \\ -27.666 & -17.889 & -12.924 \\ -5.152 & -4.514 & -4.243 \\ -1.086 & 0.429 & 1.155 \\ -3.141 & -3.703 & -3.744 \\ 23.677 & 23.157 & 21.946 \\ -3.822 & 0.576 & 1.816 \\ 4.735 & 1.468 & -1.123 \\ -9.727 & -5.757 & -4.840 \end{bmatrix}$

d) $r_i = \begin{bmatrix} 10866.508 & 951.322 & 632.526 \\ -484.851 & 901.990 & 0.273 \\ -1911.399 & 220.869 & -37.091 \\ 32.285 & 4.367 & 0.770 \\ 5.174 & -1.506 & 0.025 \\ -2.927 & 0.708 & -0.028 \\ -8.763 & 6.716 & 0.059 \\ -8.980 & -0.565 & -0.281 \\ -6.844 & 0.278 & 0.486 \end{bmatrix}$ $\quad \gamma = -V_{XY} V_{YY}^{-1}$ (See my own Questions)

e)



0
100
200
300
400
500
0
100
200
300
400
500
600
700
800

4.3 C 10 / 10

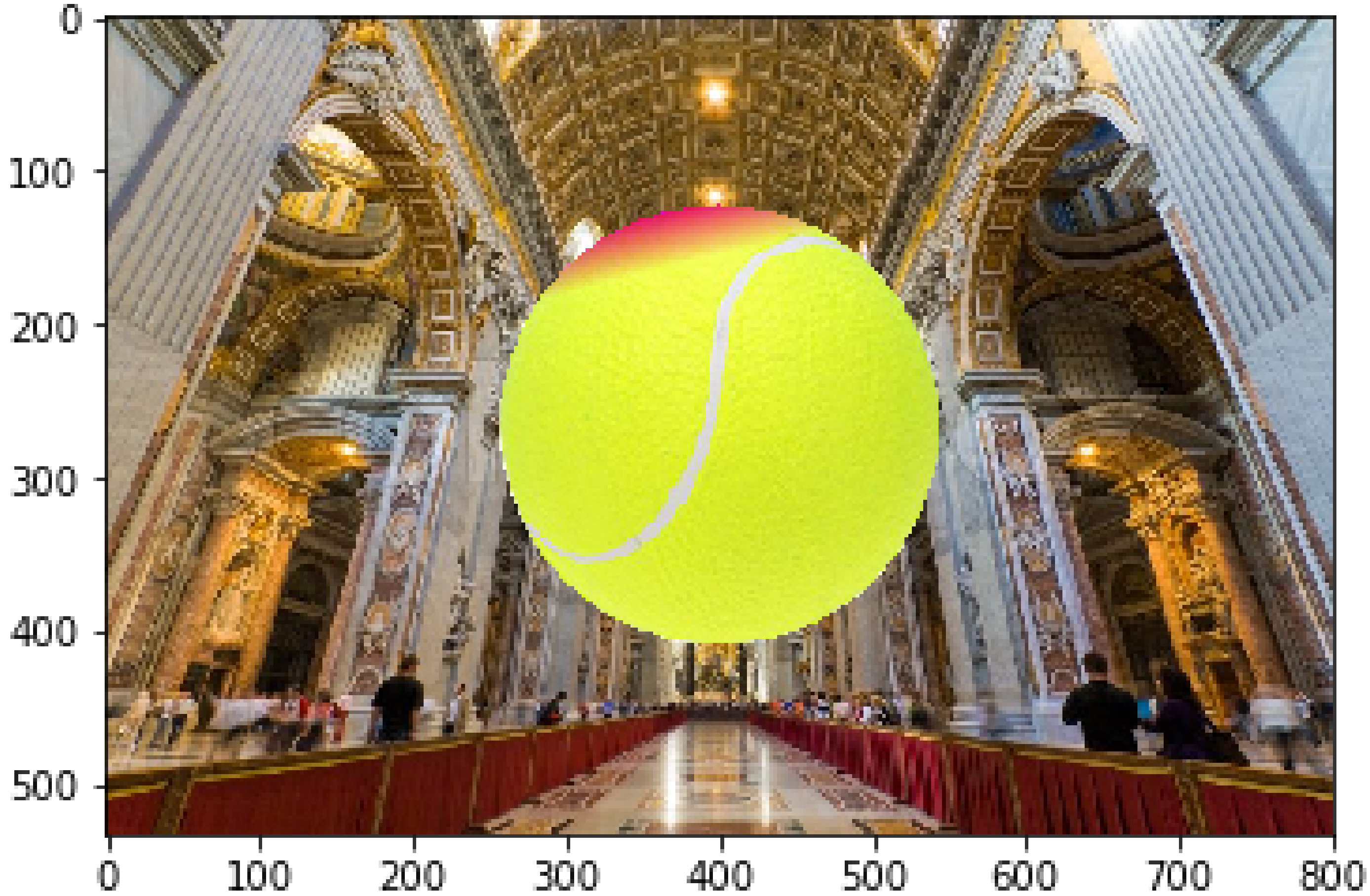
- ✓ + 10 pts Correct
- + 0 pts Incorrect
- + 5 pts Partially correct

c) $r_i = \begin{bmatrix} 202.318 & 162.420 & 149.071 \\ -27.666 & -17.889 & -12.924 \\ -5.152 & -4.516 & -4.243 \\ -1.086 & 0.429 & 1.155 \\ -3.141 & -3.703 & -3.744 \\ 23.677 & 23.157 & 21.946 \\ -3.822 & 0.576 & 1.816 \\ 4.735 & 1.468 & -1.123 \\ -9.727 & -5.757 & -4.840 \end{bmatrix}$

d) $r_i = \begin{bmatrix} 10866.508 & 951.322 & 632.526 \\ -484.851 & 901.990 & 0.273 \\ -1911.399 & 220.869 & -37.091 \\ 32.285 & 4.367 & 0.770 \\ 5.174 & -1.506 & 0.025 \\ -2.927 & 0.708 & -0.028 \\ -8.763 & 6.716 & 0.059 \\ -8.980 & -0.565 & -0.281 \\ -6.844 & 0.278 & 0.486 \end{bmatrix}$

$\gamma = -V_{XY} V_{YY}^{-1}$ (See my own Questions)

e)


0
100
200
300
400
500
0
100
200
300
400
500
600
700
800

4.4 d **5 / 10**

+ **0 pts** Incorrect/Blank
+ **2 pts** Minimal Progress
✓ + **5 pts** **Good Progress**
+ **8 pts** Missing minor part or Off by a bit
+ **10 pts** Correct

c) $r_i = \begin{bmatrix} 202.318 & 162.420 & 149.071 \\ -27.666 & -17.889 & -12.924 \\ -5.152 & -4.516 & -4.243 \\ -1.086 & 0.429 & 1.155 \\ -3.141 & -3.703 & -3.744 \\ 23.677 & 23.157 & 21.946 \\ -3.822 & 0.576 & 1.816 \\ 4.735 & 1.468 & -1.123 \\ -9.727 & -5.757 & -4.840 \end{bmatrix}$

d) $r_i = \begin{bmatrix} 10866.508 & 951.322 & 632.526 \\ -484.851 & 901.990 & 0.273 \\ -1911.399 & 220.869 & -37.091 \\ 32.285 & 4.367 & 0.770 \\ 5.174 & -1.506 & 0.025 \\ -2.927 & 0.708 & -0.028 \\ -8.763 & 6.716 & 0.059 \\ -8.980 & -0.565 & -0.281 \\ -6.844 & 0.278 & 0.486 \end{bmatrix}$

$\gamma = -V_{XY} V_{YY}^{-1}$ (See my own Questions)

e)

4.5 e 0 / 10

+ **10 pts** Correct
+ **8 pts** Mostly Correct
+ **6 pts** Significant Progress
+ **4 pts** Some Progress
+ **2 pts** Trivial Progress
+ ✓ + **0 pts** No Progress

HO4-5 My Own Question

Total Least Square: when $\begin{cases} Y \in R^{n\times k} \\ X \in R^{n\times d} \end{cases}$, what is w? $w \in R^{d\times k}$

$$\underset{\vec{w}}{\text{minimize}} \; \| [\hat{X}, \hat{Y}] \|_F^2$$

$$\text{s.t.} \quad [X+\hat{X}, \; Y+\hat{Y}] \begin{bmatrix} w \\ -I_k \end{bmatrix} = 0$$

$$\underset{n\times(d+k)}{[X \; Y]} = [\underset{n\times d}{U_X} \; \underset{n\times k}{U_Y}] \begin{bmatrix} \overset{d\times d}{\Sigma_X} & 0 \\ 0 & \underset{k\times k}{\Sigma_Y} \end{bmatrix} \begin{bmatrix} \overset{d\times d}{V_{XX}} & V_{XY} \\ V_{YX} & \underset{k\times k}{V_{YY}} \end{bmatrix}^T$$

According to E-Y theorem

$$[X+\hat{X}, \; Y+\hat{Y}] = [U_X \; U_Y] \begin{bmatrix} \Sigma_X & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_{XX} & V_{XY} \\ V_{YX} & V_{YY} \end{bmatrix}^T$$

$$[\hat{X} \; \hat{Y}] = -[U_X \; U_Y] \begin{bmatrix} 0 & 0 \\ 0 & \Sigma_Y \end{bmatrix} \begin{bmatrix} V_{XX} & V_{XY} \\ V_{YX} & V_{YY} \end{bmatrix}^T$$

$$= -[0 \;\; U_Y \Sigma_Y] \begin{bmatrix} V_{XX} & V_{XY} \\ V_{YX} & V_{YY} \end{bmatrix}^T = -U_Y \Sigma_Y \begin{bmatrix} V_{XY} \\ V_{YY} \end{bmatrix}^T$$

$$= -[X \; Y] \begin{bmatrix} \overset{d\times k}{V_{XY}} \\ \underset{k\times k}{V_{YY}} \end{bmatrix} \begin{bmatrix} \overset{d\times k}{V_{XY}} \\ V_{YY} \end{bmatrix}^T$$

$$[\hat{X} \; \hat{Y}] \underset{(d+k)\times k}{\begin{bmatrix} V_{XY} \\ V_{YY} \end{bmatrix}} = -[X \; Y] \begin{bmatrix} V_{XY} \\ V_{YY} \end{bmatrix} \qquad [\hat{X}+X \;\; \hat{Y}+Y] \begin{bmatrix} w \\ -I_k \end{bmatrix} = 0$$

$$w = -V_{XY} V_{YY}^{-1}$$

reference: https://en.wikipedia.org/wiki/total-least-squares

5 Your Own Question 10 / 10

✓ + **10 pts** Correct

+ **0 pts** No Answer