CS 189 HW2 Write-Up

Qingyang Zhao

TOTAL POINTS

91 / 124

QUESTION 1

Getting Started 4 pts

1.1 Team members 2 / 2

✓ + 2 pts **Correct**

+ 0 pts Blank

1.2 Statement and signature 2 / 2

✓ + 2 pts **Correct**

+ 0 pts Empty or No Signature

QUESTION 2

Geometry of Ridge Regression 40 pts

2.1 a 0 / 0

✓ - 0 pts -

2.2 b 0 / 0

✓ - 0 pts -

2.3 c 3 / 10

+ 10 pts Correct

+ 5 pts Correct Derivation

✓ + 3 pts **Correct behavior for lambda to infinity**

+ 2 pts Second order condition verified (Hessian)

+ 3 pts Incomplete or minor error in derivation

+ 1 pts Incomplete or minor error in lambda to infinity behavior

+ 0 pts Incorrect or Blank

2.4 d 10 / 10

✓ + 6 pts **Correct eigenvalues**

✓ + 4 pts **Correct comment on improving the inversion operation**

+ 4 pts Minor mistake in computing eigenvalues

+ 0 pts Incorrect or blank

2.5 e 10 / 10

✓ + 10 pts **Set lambda = 0.5 with correct justification**

+ 0 pts Incorrect lambda

+ 5 pts Set lambda = 0.5 with incorrect justification

+ 0 pts Missing problem

2.6 f 4 / 10

+ 10 pts Correct

+ 9 pts very much on the right track, but minor mistake/but missing some logical steps

+ 6 pts either justification of ridge regression or OLS is wrong

✓ + 4 pts **show effort, but mostly incorrect**

+ 0 pts Incorrect

2.7 g 10 / 0

✓ + 10 pts **Correct**

+ 8 pts Mostly Correct

+ 5 pts Made Progress

+ 2 pts Trivial Progress

+ 0 pts No Progress

2.8 h 0 / 0

✓ - 0 pts -

QUESTION 3

Polynomials and Invertibility 20 pts

3.1 a 0 / 0

✓ - 0 pts -

3.2 b 0 / 0

✓ - 0 pts -

3.3 c 0 / 0

✓ - 0 pts -

3.4 d 0 / 0

✓ - 0 pts -

3.5 e 0 / 0

✓ - 0 pts -

3.6 f 0 / 0

✓ - 0 pts -

3.7 g 10 / 10

✓ + 10 pts **Correctly identifies that columns produced**

by same degree will be identical to each other and therefore lin dependent.

+ 7 pts Showed linear dependence but incorrect notion of feature matrix (Saying all columns are same. Only columns of same degree are identical).

+ 4 pts Tries to show linear independence but argument or interpretation is not quite sound

+ 0 pts Completely Wrong/Answer Missing

3.8 h 0 / 10

+ 10 pts Fully correct

+ 6 pts Relates solution to picking univariate monomials from the previous parts but doesn't have correct vectors

+ 2 pts Hits some related points in explanation but doesn't choose correct vectors

✓ **+ 0 pts Blank or incorrect**

QUESTION 4

Polynomials and Approximation 30 pts

4.1 a 10 / 10

✓ **+ 10 pts part i and part ii correct**

+ 5 pts part i correct

+ 5 pts part ii correct

+ 5 pts expansions correct

+ 3 pts partially correct part i

+ 3 pts partially correct part ii

+ 0 pts Incorrect or blank

4.2 b 0 / 0

✓ **- 0 pts -**

4.3 c 0 / 0

✓ **- 0 pts -**

4.4 d 10 / 10

✓ **+ 10 pts Correct**

+ 9 pts Mostly correct/minor mistake

+ 6 pts A significant mistake

+ 4 pts show effort, but incorrect

+ 0 pts Incorrect

4.5 e 10 / 10

+ 0 pts Incorrect

✓ **+ 10 pts Correct**

+ 5 pts Partially correct

4.6 f 0 / 0

✓ **- 0 pts Correct**

4.7 g 0 / 0

✓ **+ 0 pts Incorrect**

+ 10 pts Correct

+ 5 pts Partially correct

QUESTION 5

James and his Giant Peaches 20 pts

5.1 a 0 / 0

✓ **- 0 pts -**

5.2 b 0 / 0

✓ **- 0 pts -**

5.3 c 0 / 0

✓ **- 0 pts -**

5.4 d 0 / 10

+ 10 pts Correct

+ 5 pts Partially Correct

+ 3 pts Penalty

✓ **+ 0 pts Incorrect**

No plot for d?

5.5 e 0 / 0

✓ **- 0 pts -**

5.6 f 0 / 0

✓ **- 0 pts -**

5.7 g 0 / 10

+ 10 pts Correct

+ 8 pts Small Mistake

+ 5 pts Partial

+ 3 pts Penalty

✓ **+ 0 pts Incorrect/Empty**

QUESTION 6

6 Your Own Question 10 / 10

✓ **+ 10 pts Correct**

+ 0 pts No Answer

HW2-1

a) Names Email Address

(1) Wan jhun0324@berkeley.edu

Description of Team:

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

1.1 Team members 2 / 2

✓ + **2 pts** Correct

+ **0 pts** Blank

H02-1

a)

Names	Email Address
(1) Wan	jhun0324@berkeley.edu

Description of Team:

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

1.2 Statement and signature 2 / 2

- ✓ + 2 pts Correct
- + 0 pts Empty or No Signature

As well as

HO2-2 / HO2-6 My own Question: What is the relationship between Ridge & Lagrange?

$\boxed{a)}$ The Conclusion is as β becomes large, λ becomes small.

If $\beta \geq \beta_0$, λ becomes 0.

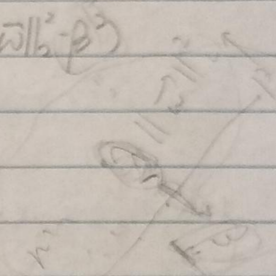
Where $\beta_0 = \left\| \underset{\vec{w}}{\text{argmin}} \|\vec{y} - X\vec{w}\|_2^2 \right\|_2$

(Relation between Lagrange & Ridge Regression)

$\boxed{\text{Explanation 1}}$: Optimization Problem

Note: $\underset{\vec{w}}{\text{minimize}}\ f(\vec{w})$ s.t. $g(\vec{w}) \leq 0$ — Standard form.

$$\underset{\vec{w}}{\text{minimize}} \|\vec{y} - X\vec{w}\|_2^2 \quad \text{s.t. } \|\vec{w}\|_2^2 \leq \beta^2 \xrightarrow[\text{Multiplier}]{\text{Using Lagrange}} L(\vec{w}, \lambda) = \|\vec{y} - X\vec{w}\|_2^2 + \lambda(\|\vec{w}\|_2^2 - \beta^2)$$



The Optimal Conditions: ① $\nabla L(\vec{w}, \lambda) = 0$

$$\frac{\partial L(\vec{w}, \lambda)}{\partial \vec{w}} = 0$$

② $\|w\|_2^2 \leq \beta^2$

③ $\lambda(\|\vec{w}\|_2^2 - \beta^2) = 0$

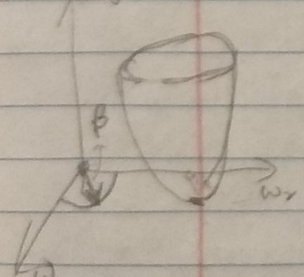
If $\lambda = 0$, $\|w\|_2^2 \leq \beta^2$
$\lambda > 0$, $\|\vec{w}\|_2^2 = \beta^2$ > That's where we need to check.

Instead of check it, think of this problem reversely.

First, note objective function is convex, inequality constraint is also convex.

The Conclusion is the global optimal lies either at the boundary or the extremum.

If it is at the boundary, correspond to "$\lambda > 0$, $\|w\|_2^2 = \beta^2$"



In this case, according to Optimal Condition, there exists $\lambda^* > 0$ s.t. $\frac{\partial L(\vec{w}, \lambda^*)}{\partial \vec{w}} = 0$ (as $\frac{\partial \beta^2}{\partial \vec{w}} = 0$)

And this is exactly what Ridge Regression trys to solve.

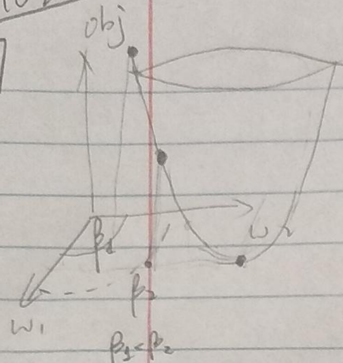
Until Now the conclusion is "Ridge Regression is trying to solve

$\underset{\vec{w}}{\text{minimize}} \|\vec{y} - X\vec{w}\|_2$
s.t. $\|\vec{w}\|_2 \leq \beta^2$ w.r.t. the λ^* (corresponding to β)

(Omit the part where $\beta > \beta_0$ which is $\lambda = 0$ $\|w\|_2^2 \leq \beta^2$ section.)

$\boxed{\text{Explanation 2:}}$ when $\beta \uparrow$ $\lambda \downarrow$ (within $\beta < \beta_0$)

$\boxed{\text{HO2-2}}$



$\boxed{a)}$ As $\beta\uparrow$, the minimum objective function can achieve will be smaller.
Remember λ is to penalize a larger $\|w\|_2$,
In fact, if a better Minimum can be reached within the constraint not much penalty should be used.
That is why λ will become smaller.

$\boxed{b)}$

$$\|\vec{\omega}^T(\vec{x}+\vec{\epsilon}) - \vec{\omega}^T x\|_2 = \|\vec{\omega}^T\vec{\epsilon}\|_2 \le \|\vec{\omega}\|_2\|\vec{\epsilon}\|_2 = \epsilon\|\vec{\omega}\|_2$$

which means when $\vec{\omega}$ small, less disturbance will appear in the result.

$\boxed{c)}$ $\vec{\omega}=\vec{0}$ when $\lambda\to\infty$ Because $(X^TX+\lambda I)^{-1}\to O_{n\times n}$

$\boxed{d)}$ The eigenvalues are $\lambda_1+\lambda, \lambda_2+\lambda \cdots \lambda_d+\lambda$.

SVD: $X^TX = U_x D_x U_x^T$ U_x is orthonormal
Since, $X^TXU_x = U_xD_x$, diagonal elements in D_x are eigenvalues of X^TX.

Then, $X^TX+\lambda I = U_xD_xU_x^T + \lambda U_xU_x^T = U_x(D_x+\lambda I)U_x^T$

When X^TX nearly invertible, means D_x has rather small values, λI can make the matrix invertible.

$\boxed{e)}$ 0.5 is better. Because 0.5 gives us a smaller training error.

Just using the conclusion of a). when λ becomes larger, β can be small

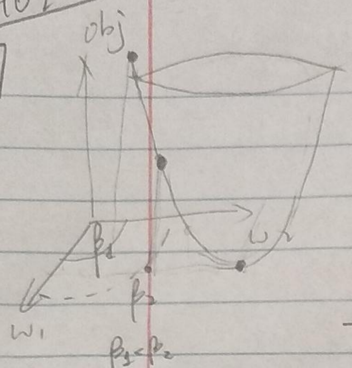
Thus the minimum of the training data is also small.

2.1 a 0 / 0

✓ - 0 pts -

HO2-2

a)



As $\beta \uparrow$, the minimum objective function can achieve will be smaller.
Remember λ is to penalize a larger $\|w\|_2$,
In fact, if a better Minimum can be reached within the constraint not much penalty should be used.
That is why λ will become smaller.

b)

$$\|\vec{w}^T(\vec{x}+\vec{\varepsilon}) - \vec{w}^T x\|_2 = \|\vec{w}^T \vec{\varepsilon}\|_2 \leq \|\vec{w}\|_2 \|\vec{\varepsilon}\|_2 = \varepsilon \|\vec{w}\|_2$$

which means when $\vec{w}$ small, less disturbance will appear in the result.

c) $\vec{w} = \vec{0}$ when $\lambda \to \infty$ Because $(X^T X + \lambda I)^{-1} \to 0_{n\times n}$

d) The eigenvalues are $\lambda_1+\lambda, \lambda_2+\lambda \cdots \lambda_d+\lambda$.

SVD: $X^T X = U_x D_x U_x^T$ U_x is orthonormal

Since, $X^T X U_x = U_x D_x$, diagonal elements in D_x are eigenvalues of $X^T X$.

Then, $X^T X + \lambda I = U_x D_x U_x^T + \lambda U_x U_x^T = U_x (D_x + \lambda I) U_x^T$

When $X^T X$ nearly invertible, means D_x has rather small values, λI can make the matrix invertible.

e) 0.5 is better. Because 0.5 gives us a smaller training error.

Just using the conclusion of a). when λ becomes larger, β can be small

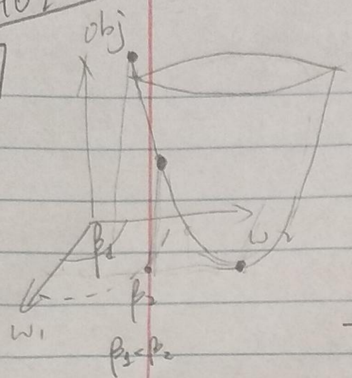
Thus the minimum of the training data is also small.

2.2 b 0 / 0

✓ - 0 pts -

HO2-2

a)



As $\beta \uparrow$, the minimum objective function can achieve will be smaller.
Remember λ is to penalize a larger $\|w\|_2$,
In fact, if a better Minimum can be reached within the constraint not much penalty should be used.
That is why λ will become smaller.

b)

$$\|\vec{w}^T(\vec{x}+\vec{\varepsilon}) - \vec{w}^T x\|_2 = \|\vec{w}^T\vec{\varepsilon}\|_2 \le \|\vec{w}\|_2\|\vec{\varepsilon}\|_2 = \varepsilon\|\vec{w}\|_2$$

which means when $\vec{w}$ small, less disturbance will appear in the result.

c) $\vec{w} = \vec{0}$ when $\lambda \to \infty$ Because $(X^TX + \lambda I)^{-1} \to O_{n \times n}$

d) The eigenvalues are $\lambda_1+\lambda, \lambda_2+\lambda \cdots \lambda_d+\lambda$.

SVD: $X^TX = U_x D_x U_x^T$ U_x is orthonormal
Since, $X^TXU_x = U_xD_x$, diagonal elements in D_x are eigenvalues of X^TX.

Then, $X^TX + \lambda I = U_xD_xU_x^T + \lambda U_xU_x^T = U_x(D_x+\lambda I)U_x^T$

When X^TX nearly invertible, means D_x has rather small values, λI can make the matrix invertible.

e) 0.5 is better. Because 0.5 gives us a smaller training error.

Just using the conclusion of a). when λ becomes larger, β can be small

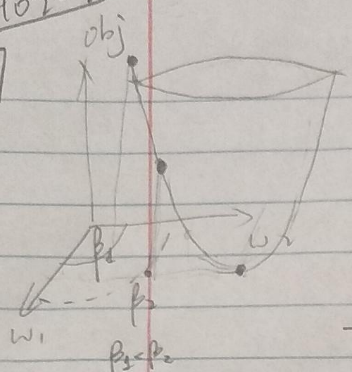
Thus the minimum of the training data is also small.

2.3 C 3 / 10

+ **10 pts** Correct
+ **5 pts** Correct Derivation
✓ + **3 pts** **Correct behavior for lambda to infinity**
+ **2 pts** Second order condition verified (Hessian)
+ **3 pts** Incomplete or minor error in derivation
+ **1 pts** Incomplete or minor error in lambda to infinity behavior
+ **0 pts** Incorrect or Blank

H02-2

a)



As $\beta \uparrow$, the minimum objective function can achieve will be smaller.
Remember λ is to penalize a larger $\|w\|_2$,
In fact, if a better Minimum can be reached within the constraint not much penalty should be used.
That is why λ will become smaller.

b)

$$\|\vec{\omega}^T(\vec{x}+\vec{\varepsilon}) - \vec{\omega}^T x\|_2 = \|\vec{\omega}^T \vec{\varepsilon}\|_2 \le \|\vec{\omega}\|_2 \|\vec{\varepsilon}\|_2 = \varepsilon \|\vec{\omega}\|_2$$

which means when $\vec{\omega}$ small, less disturbance will appear in the result.

c) $\vec{\omega} = \vec{0}$ when $\lambda \to \infty$ Because $(X^T X + \lambda I)^{-1} \to O_{n \times n}$

d) The eigenvalues are $\lambda_1 + \lambda, \lambda_2 + \lambda \cdots \lambda_d + \lambda$.

SVD: $X^T X = U_x D_x U_x^T$ U_x is orthonormal
Since, $X^T X U_x = U_x D_x$, diagonal elements in D_x are eigenvalues of $X^T X$.

Then, $X^T X + \lambda I = U_x D_x U_x^T + \lambda U_x U_x^T = U_x (D_x + \lambda I) U_x^T$

When $X^T X$ nearly invertible, means D_x has rather small values, λI can make the matrix invertible.

e) 0.5 is better. Because 0.5 gives us a smaller training error.

Just using the conclusion of a). when λ becomes larger, β can be small

Thus the minimum of the training data is also small.

2.4 d 10 / 10

✓ + 6 pts Correct eigenvalues

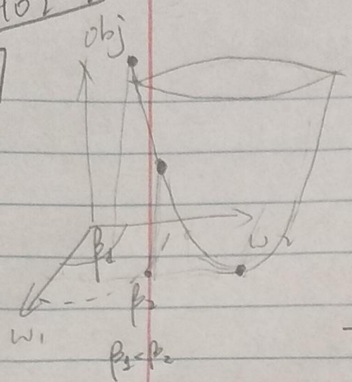
✓ + 4 pts Correct comment on improving the inversion operation

+ 4 pts Minor mistake in computing eigenvalues

+ 0 pts Incorrect or blank

HW2-2

a)



As $\beta \uparrow$, the minimum objective function can achieve will be smaller.
Remember λ is to penalize a larger $\|w\|_2$,
In fact, if a better Minimum can be reached within the constraint not much penalty should be used.
That is why λ will become smaller.

b)

$$\|\vec{\omega}^T(\vec{x}+\vec{\varepsilon}) - \vec{\omega}^T x\|_2 = \|\vec{\omega}^T\vec{\varepsilon}\|_2 \le \|\vec{\omega}\|_2\|\vec{\varepsilon}\|_2 = \varepsilon\|\vec{\omega}\|_2$$

which means when $\vec{\omega}$ small, less disturbance will appear in the result.

c) $\vec{\omega} = \vec{0}$ when $\lambda \to \infty$ Because $(X^TX + \lambda I)^{-1} \to O_{n\times n}$

d) The eigenvalues are $\lambda_1+\lambda, \lambda_2+\lambda \cdots \lambda_d+\lambda$.

SVD: $X^TX = U_x D_x U_x^T$ U_x is orthonormal
Since, $X^TXU_x = U_xD_x$, diagonal elements in D_x are eigenvalues of X^TX.

Then, $X^TX + \lambda I = U_xD_xU_x^T + \lambda U_xU_x^T = U_x(D_x+\lambda I)U_x^T$

When X^TX nearly invertible, means D_x has rather small values, λI can make the matrix invertible.

e) 0.5 is better. Because 0.5 gives us a smaller training error.

Just using the conclusion of a). when λ becomes larger, β can be small

Thus the minimum of the training data is also small.

2.5 e 10 / 10

✓ + **10 pts** **Set lambda = 0.5 with correct justification**

+ **0 pts** Incorrect lambda

+ **5 pts** Set lambda = 0.5 with incorrect justification

+ **0 pts** Missing problem

HO2-T2

$U_X D_X V_X^T \quad U_X D_X \cancel{V_X^T} V_X D_X U_X^T$

minimize

$\boxed{f)}$ $\vec{w}$ $\|\vec{y} - X\vec{w}\|_2$ $\quad R(X) = R(X^T X)$

$X^T X \vec{w} = X^T \vec{y} \quad X^T X$ is not full rank.

The ~~ans~~ solution of this equation is Homogeneous + Particular;

$\vec{w}$ lies in the null space of $X^T X$, Homogeneous solutions are infinite.

$\boxed{g)}$ $\lim \lambda \to 0$, $X^T X + \lambda I$ is invertible, we have one solution.

However $\|\vec{w}\|_2^2$ will be really large.

This is because ~~both~~ ridge regression is to solve a problem: $\min_{\vec{w}} \|\vec{y} - X\vec{w}\|_2$ where β is really large. And the solution $\|\vec{w}\|_2^2 \le \beta^2$

will be at $\|w\|_2^2 = \beta^2$ (Since the solutions are infinite). So add a $\lim \lambda \to 0$ will cause a big $\|w\|_2$ ~~such that~~ which is not what we want.

$\boxed{h)}$ $\min_{\vec{w}} \|y - X\vec{w}\|_2^2 + \lambda \|\Gamma \vec{w}\|_2^2$

$$\frac{\partial \; y^T y + \vec{w}^T X^T X \vec{w} - 2\vec{y}^T X \vec{w} + \lambda \vec{w}^T \Gamma^T \Gamma \vec{w}}{\partial \vec{w}^T}$$

$$= 2X^T X \vec{w} - 2X^T \vec{y} + 2\lambda \Gamma^T \Gamma \vec{w} = 0.$$

$$\vec{w} = (X^T X + \lambda \Gamma^T \Gamma)^{-1} X^T \vec{y}$$

2.6 f 4 / 10

+ **10 pts** Correct
+ **9 pts** very much on the right track, but minor mistake/but missing some logical steps
+ **6 pts** either justification of ridge regression or OLS is wrong
- ✓ + **4 pts** show effort, but mostly incorrect
+ **0 pts** Incorrect

H02-T2

$U_X D_X V_X^T \quad U_X D_X \cancel{V_X^T V_X} D_X U_X^T$

minimize

$\boxed{f)}$ $\vec{w} \quad \|\vec{y} - X\vec{w}\|_2 \qquad R(X) = R(X^T X)$

$X^T X \vec{w} = X^T\vec{y} \qquad X^T X$ is not full rank.

The ~~ans~~ solution of this equation is Homogeneous + Particular;

$\vec{w}$ lies in the null space of $X^T X$, Homogeneous solutions are infinite.

$\boxed{g)}$ $\lim \lambda \to 0$, $X^T X + \lambda I$ is invertible, we have one solution.
However $\|\vec{w}\|_2^2$ will be really large.
This is because ~~tech~~ ridge regression is to solve a
problem: $\min_{\vec{w}} \|\vec{y} - X\vec{w}\|_2$ where β is really large. And the solution
$\|\vec{w}\|_2^2 \le \beta^2$

will be at $\|w\|_2^2 = \beta^2$ (Since the solutions are infinite). So add a $\lim \lambda \to 0$
will cause a big $\|w\|_2$ ~~such that~~ which is not what we want.

$\boxed{h)}$ $\min_{\vec{w}} \|y - X\vec{w}\|_2^2 + \lambda \|\Gamma \vec{w}\|_2^2$

$$\frac{\partial \; y^T y + \vec{w}^T X^T X \vec{w} - 2\vec{y}^T X \vec{w} + \lambda \vec{w}^T \Gamma^T \Gamma \vec{w}}{\partial \vec{w}^T}$$

$$= 2X^T X \vec{w} - 2X^T \vec{y} + 2\lambda \Gamma^T \Gamma \vec{w} = 0.$$

$$\vec{w} = (X^T X + \lambda \Gamma^T \Gamma)^{-1} X^T \vec{y}$$

2.7 g 10 / 0

✓ + **10 pts** Correct

+ **8 pts** Mostly Correct

+ **5 pts** Made Progress

+ **2 pts** Trivial Progress

+ **0 pts** No Progress

HO2-T2

$U_x D_x V_x^T \quad U_x D_x \cancel{V_x^T} V_x D_x U_x^T$

minimise

$\boxed{f)}$ $\underset{\vec{w}}{\text{minimise}} \; \|\vec{y} - X\vec{w}\|_2 \qquad R(X) = R(X^T X)$

$X^T X \vec{w} = X^T\vec{y} \qquad X^T X$ is not full rank.

The ~~ans~~ solution of this equation is Homogeneous + Particular;

$\vec{w}$ lies in the null space of $X^T X$, Homogeneous solutions are infinite.

$\boxed{g)}$ $\lim \lambda \to 0$, $X^T X + \lambda I$ is invertible, we have one solution. However $\|\vec{w}\|_2^2$ will be really large.

This is because ~~both~~ ridge regression is to solve a problem: $\min_{\vec{w}} \|\vec{y} - X\vec{w}\|_2$ where β is really large. And the solution
$\|\vec{w}\|_2^2 \le \beta^2$

will be at $\|w\|_2^2 = \beta^2$ (Since the solutions are infinite). So add a $\lim \lambda \to 0$ will cause a big $\|w\|_2$ ~~such that~~ which is not what we want.

$\boxed{h)}$ $\min_{\vec{w}} \|y - X\vec{w}\|_2^2 + \lambda \|\Gamma \vec{w}\|_2^2$

$$\frac{\partial \; y^T y + \vec{w}^T X^T X \vec{w} - 2\vec{y}^T X \vec{w} + \lambda \vec{w}^T \Gamma^T \Gamma \vec{w}}{\partial \vec{w}^T}$$

$$= 2X^T X\vec{w} - 2X^T\vec{y} + 2\lambda \Gamma^T \Gamma \vec{w} = 0.$$

$$\vec{w} = (X^T X + \lambda \Gamma^T \Gamma)^{-1} X^T \vec{y}$$

2.8 h 0 / 0

✓ - 0 pts -

HO2-3

(a)

$$F=\begin{bmatrix}1 & x_1\\ 1 & x_2\end{bmatrix}$$

If $x_1 \neq x_2 \Rightarrow \det F \neq 0 \quad \operatorname{rank}(F)=2$

If $\operatorname{rank}(F)=2 \Rightarrow \det F \neq 0 \quad x_1 \neq x_2$

(b)

$$F=\begin{bmatrix}1 & x_1 & \cdots & x_1^d\\ 1 & x_2 & \cdots & x_2^d\\ \vdots & \vdots & & \vdots\\ 1 & x_n & & x_n^d\end{bmatrix} \quad F'=\begin{bmatrix}1 & x_1 & \cdots & x_1^d\\ 0 & x_2-x_1 & & x_2^d-x_1^d\\ 0 & x_3-x_1 & & \vdots\\ \vdots & \vdots & & \\ 0 & x_n-x_1 & & x_n^d-x_1^d\end{bmatrix}$$

$$F'=\begin{bmatrix}1 & 0 & 0 & \cdots & 0\\ -1 & 1 & & 0 & \\ \vdots & & \ddots & & \\ -1 & & & & 1\end{bmatrix}F=\Lambda F \quad \det(F')=\det(\Lambda F)=\det(\Lambda)\det(F)$$

$$\det(\Lambda)=1$$

$$1\cdot F'_{(n-1)\times(n-1)}=1\cdot 1\cdot F'_{(n-2)\times(n-2)}=1\cdots 1\cdot F_1 =1$$

$\det(\lambda I-F')=(\lambda-1)^n \quad \lambda_1=\cdots=\lambda_n=1$

The eigenvalue of a triangle matrix are its diagonal elements.

(c)

$$F''=\begin{bmatrix}1 & 0 & 0 & \cdots & 0\\ 0 & x_2-x_1 & x_1(x_2-x_1) & \cdots & x_1^{d-1}(x_2-x_1)\\ 0 & x_3-x_1 & x_1(x_3-x_1) & \cdots & x_1^{d-1}(x_3-x_1)\\ \vdots & & & & \vdots\\ 0 & x_n-x_1 & x_1(x_n-x_1) & \cdots & x_1^{d-1}(x_n-x_1)\end{bmatrix}=F'\underbrace{\begin{bmatrix}1 & 0 & \cdots & & 0\\ 0 & 1 & & & 0\\ 0 & & \ddots & & \vdots\\ & 0 & & 1 & -x_1\\ 0 & & & 0 & 1\end{bmatrix}}_{\Sigma_n}\cdots\underbrace{\begin{bmatrix}1 & 0 & \cdots & & 0\\ 0 & 1 & -x_1 & 0 & \\ & & 1 & & \\ & & 0 & & \\ & & & & 1\end{bmatrix}}_{\Sigma_2}$$

$\det(F'')=\det(F')\det(\Sigma_n\cdots\Sigma_1)=\det(F')$

(d)

$$B=\begin{bmatrix}1 & \vec{0}^T\\ \vec{0} & A\end{bmatrix} \quad \det(\lambda I-B)=0 \quad \det[(\lambda-1)(\lambda I-A)]=0$$

$\therefore$ B eigenvalue $\{1, \lambda_1(A), \lambda_2(A) \cdots \lambda_d(A)\}$

$$\det(A)=\prod_{i=1}^{n}\lambda_i(A) = 1\cdot\prod_{i=1}^{n}\lambda_i(A)=\det(B)$$

3.1 a 0 / 0

✓ - 0 pts -

$\boxed{\text{HO2-3}}$

$\boxed{a)}$

$$F = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \end{bmatrix}$$

If $x_1 \neq x_2 \Rightarrow \det F \neq 0 \quad \mathrm{rank}(F) = 2$

If $\mathrm{rank}(F) = 2 \Rightarrow \det F \neq 0 \quad x_1 \neq x_2$

$\boxed{b)}$

$$F = \begin{bmatrix} 1 & x_1 & \cdots & x_1^d \\ 1 & x_2 & \cdots & x_2^d \\ \vdots & \vdots & & \vdots \\ 1 & x_n & & x_n^d \end{bmatrix} \quad F' = \begin{bmatrix} 1 & x_1 & \cdots & x_1^d \\ 0 & x_2 - x_1 & & x_2^d - x_1^d \\ 0 & x_3 - x_1 & & \vdots \\ \vdots & \vdots & & \\ 0 & x_n - x_1 & & x_n^d - x_1^d \end{bmatrix}$$

$$F' = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & & 0 & \\ \vdots & & \ddots & & \\ -1 & & & & 1 \end{bmatrix} F = \Lambda F \quad \det(F') = \det(\Lambda F) = \det(\Lambda)\det(F)$$

$$\det(\Lambda) = 1 \cdot F'_{(n-1)\times(n-1)} = 1 \cdot 1 \cdot F'_{(n-2)\times(n-2)} = 1 \cdots 1 \cdot F_1 = 1$$

$\det(\lambda I - F') = (\lambda - 1)^n \quad \lambda_1 = \cdots = \lambda_n = 1$

The eigenvalue of a triangle matrix are its diagonal elements.

$\boxed{c)}$

$$F'' = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & x_2 - x_1 & x_1(x_2 - x_1) & \cdots & x_1^{d-1}(x_2 - x_1) \\ 0 & x_3 - x_1 & x_1(x_3 - x_1) & \cdots & x_1^{d-1}(x_3 - x_1) \\ \vdots & & & & \vdots \\ 0 & x_n - x_1 & x_1(x_n - x_1) & \cdots & x_1^{d-1}(x_n - x_1) \end{bmatrix} = F' \underbrace{\begin{bmatrix} 1 & 0 & \cdots & & 0 \\ 0 & 1 & & & 0 \\ 0 & & \ddots & 0 & \vdots \\ & 0 & & 1 & -x_1 \\ 0 & & & 0 & 1 \end{bmatrix}}_{\Sigma_n} \cdots \underbrace{\begin{bmatrix} 1 & 0 & \cdots & & 0 \\ 0 & 1 & -x_1 & 0 & \\ & & 1 & & \\ & & 0 & & \\ & & & & 1 \end{bmatrix}}_{\Sigma_2}$$

$\det(F'') = \det(F')\det(\Sigma_n \cdots \Sigma_1) = \det(F')$

$\boxed{d)}$

$$B = \begin{bmatrix} 1 & \vec{0}^T \\ \vec{0} & A \end{bmatrix} \quad \det(\lambda I - B) = 0 \quad \det[(\lambda - 1)(\lambda I - A)] = 0$$

$\therefore$ B eigenvalue $\{1, \lambda_1(A), \lambda_2(A) \cdots \lambda_d(A)\}$

$$\det(A) = \prod_{i=1}^{n} \lambda_i(A) = 1 \cdot \prod_{i=1}^{n} \lambda_i(A) = \det(B)$$

3.2 b **0 / 0**

✓ **- 0 pts** -

HO2-3

(a)

$$F=\begin{bmatrix}1 & x_1\\ 1 & x_2\end{bmatrix}$$

If $x_1 \neq x_2 \Rightarrow \det F \neq 0 \quad \mathrm{rank}(F)=2$

If $\mathrm{rank}(F)=2 \Rightarrow \det F \neq 0 \quad x_1 \neq x_2$

(b)

$$F=\begin{bmatrix}1 & x_1 & \cdots & x_1^d\\ 1 & x_2 & \cdots & x_2^d\\ \vdots & \vdots & & \vdots\\ 1 & x_n & & x_n^d\end{bmatrix} \quad F'=\begin{bmatrix}1 & x_1 & \cdots & x_1^d\\ 0 & x_2-x_1 & & x_2^d-x_1^d\\ 0 & x_3-x_1 & & \vdots\\ \vdots & \vdots & & \\ 0 & x_n-x_1 & & x_n^d-x_1^d\end{bmatrix}$$

$$F'=\begin{bmatrix}1 & 0 & 0 & \cdots & 0\\ -1 & 1 & & 0 & \\ \vdots & & \ddots & & \\ -1 & & & & 1\end{bmatrix}F=\Lambda F \quad \det(F')=\det(\Lambda F)=\det(\Lambda)\det(F)$$

$$\det(\Lambda) = 1$$

$$1\cdot F'_{(n-1)\times(n-1)}=1\cdot 1\cdot F'_{(n-2)\times(n-2)}=1\cdots 1\cdot F_1 =1$$

$\det(\lambda I-F')=(\lambda-1)^n \quad \lambda_1=\cdots=\lambda_n=1$

The eigenvalue of a triangle matrix are its diagonal elements.

(c)

$$F''=\begin{bmatrix}1 & 0 & 0 & \cdots & 0\\ 0 & x_2-x_1 & x_1(x_2-x_1) & \cdots & x_1^{d-1}(x_2-x_1)\\ 0 & x_3-x_1 & x_1(x_3-x_1) & \cdots & x_1^{d-1}(x_3-x_1)\\ \vdots & & & & \vdots\\ 0 & x_n-x_1 & x_1(x_n-x_1) & \cdots & x_1^{d-1}(x_n-x_1)\end{bmatrix}=F'\underbrace{\begin{bmatrix}1 & 0 & \cdots & & 0\\ 0 & 1 & & & 0\\ 0 & & \ddots & & \vdots\\ & 0 & & 1 & -x_1\\ 0 & & & 0 & 1\end{bmatrix}}_{\Sigma_n}\cdots\underbrace{\begin{bmatrix}1 & 0 & \cdots & & 0\\ 0 & 1 & -x_1 & 0 & \\ & & 1 & & \\ & & 0 & & \\ & & & & 1\end{bmatrix}}_{\Sigma_2}$$

$\det(F'')=\det(F')\det(\Sigma_n\cdots\Sigma_2)=\det(F')$

(d)

$$B=\begin{bmatrix}1 & \vec{0}^T\\ \vec{0} & A\end{bmatrix} \quad \det(\lambda I-B)=0 \quad \det[(\lambda-1)(\lambda I-A)]=0$$

$\therefore$ B eigenvalue $\{1, \lambda_1(A), \lambda_2(A) \cdots \lambda_d(A)\}$

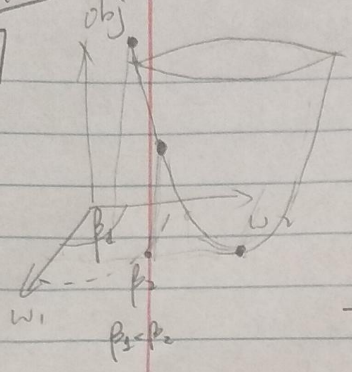
$$\det(A)=\prod_{i=1}^{n}\lambda_i(A) = 1\cdot\prod_{i=1}^{n}\lambda_i(A)=\det(B)$$

3.3 C 0 / 0

✓ - 0 pts -

HO2-2

a)



As $\beta \uparrow$, the minimum objective function can achieve will be smaller.
Remember λ is to penalize a larger $\|w\|_2$,
In fact, if a better Minimum can be reached within the constraint not much penalty should be used.
That is why λ will become smaller.

b)

$$\|\vec{\omega}^T(\vec{x}+\vec{\varepsilon}) - \omega^T x\|_2 = \|\vec{\omega}^T \vec{\varepsilon}\|_2 \le \|\vec{\omega}\|_2 \|\vec{\varepsilon}\|_2 = \varepsilon \|\vec{\omega}\|_2$$

which means when $\vec{\omega}$ small, less disturbance will appear in the result.

c) $\vec{\omega} = \vec{0}$ when $\lambda \to \infty$ Because $(X^T X + \lambda I)^{-1} \to O_{n \times n}$

d) The eigenvalues are $\lambda_1 + \lambda, \lambda_2 + \lambda \cdots \lambda_d + \lambda$.

SVD: $X^T X = U_x D_x U_x^T$ U_x is orthonormal
Since, $X^T X U_x = U_x D_x$, diagonal elements in D_x are eigenvalues of $X^T X$.

Then, $X^T X + \lambda I = U_x D_x U_x^T + \lambda U_x U_x^T = U_x (D_x + \lambda I) U_x^T$

When $X^T X$ nearly invertible, means D_x has rather small values, λI can make the matrix invertible.

e) 0.5 is better. Because 0.5 gives us a smaller training error.

Just using the conclusion of a). when λ becomes larger, β can be small

Thus the minimum of the training data is also small.

3.4 d **0 / 0**

✓ **- 0 pts** -

e) We already know that row/column subtraction doesn't change the determinant.

$$F = \begin{bmatrix} 1 & x_1 & \cdots & x_1^d \\ \vdots & & & \\ 1 & x_n & \cdots & x_n^d \end{bmatrix} \xrightarrow{-R_1} \overset{F'}{\begin{bmatrix} 1 & x_1 & \cdots & x_1^d \\ 0 & & & x_2^d - x_1^d \\ \vdots & & & \\ 0 & x_n - x_1 & & x_n^d - x_1^d \end{bmatrix}} \xrightarrow{C_i - x_1 C_{i-1}} \overset{F''}{\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & x_2 - x_1 & (x_2-x_1)x_2 \cdots & (x_2-x_1)x_2^{d-1} \\ \vdots & & & \\ 0 & x_n - x_1 & & (x_n-x_1)x_n^{d-1} \end{bmatrix}} = \prod_{i=2}^{n}(x_i - x_1) \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & x_2 \cdots & x_2^{d-1} \\ \vdots & & & \\ 0 & 1 & x_n \cdots & x_n^{d-1} \end{bmatrix}$$

$n = d+1$

$$\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & x_2^{d-1} \\ \vdots & & & \\ 0 & 1 & \cdots & x_n^{d-1} \end{bmatrix} \xrightarrow{-R_2} \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & x_2^{d-1} \\ \vdots & 0 & & \\ 0 & 0 & \cdots & x_n^{d-1} - x_2^{d-1} \end{bmatrix} \xrightarrow{C_i - x_2 C_{i-1}} \begin{bmatrix} 1 & 0 & \cdots & 0 \\ & 1 & 0 & 0 \\ & \vdots & & \\ & 0 & & x_n^{d-2}(x_n - x_2) \end{bmatrix} = \prod_{i=2}^{n}(x_i - x_1) \prod_{j=3}^{n}(x_j - x_2) \cdot \begin{bmatrix} 1 & 0 & \cdots \\ 0 & 1 & \cdots \\ & & 1 \; x_3 \cdots \end{bmatrix}$$

$\vdots$

$$\det(F) = \prod_{1 \le i < j \le n} (x_j - x_i)$$

f) $x_1 \; x_2 \cdots x_\ell \; 1$ assign d to $\ell+1$ Thus $\binom{\ell+d}{\ell}$.

g) $$F_\ell = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} \cdots & x_{1,\ell} & \underbrace{x_{1,1}^2 \; x_{1,2}^2 \cdots x_{1,\ell-1} x_{1,\ell}}_{C_{d+\ell-1}^{\ell-1}} & \cdots & \underbrace{x_{1,1}^d \cdots x_{1,1} \cdots x_{1,d}}_{C_{d+\ell-1}^{\ell-1}} \\ 1 & x_{2,1} & & & & & \\ \vdots & \vdots & & & & & \\ 1 & x_{n,1} & & & & & \end{bmatrix}$$

$F_\ell|_{x_{i,1} = x_{i,2} = \cdots = x_{i,\ell} = \alpha_i} = \begin{bmatrix} 1 & \alpha_1 & \alpha_1 \cdots & \alpha_1 & \alpha_1^2 \cdots & \alpha_1^2 & \alpha_1^3 \cdots & \alpha_1^d \\ 1 & \alpha_2 & \alpha_2 \cdots & \alpha_2 & & & & \\ \vdots & & & & & & & \\ 1 & \alpha_n & \alpha_n & & & & & \end{bmatrix}$

the rank is always $d+1$, because all columns with same degree are linear dependent.

3.5 e 0 / 0

✓ - 0 pts -

$\boxed{e)}$ We already know that row/column subtraction doesn't change the determinant.

$$F = \begin{bmatrix} 1 & x_1 & \cdots & x_1^d \\ \vdots & & & \\ 1 & x_n & \cdots & x_n^d \end{bmatrix} \xRightarrow{-R_1} \underbrace{\begin{bmatrix} 1 & x_1 & \cdots & x_1^d \\ 0 & & & x_2^d - x_1^d \\ \vdots & & & \\ 0 & x_n - x_1 & & x_n^d - x_1^d \end{bmatrix}}_{F'} \xRightarrow{C_i - x_1 C_{i-1}} \underbrace{\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & x_2 - x_1 & (x_2-x_1)x_2 \cdots & (x_2-x_1)x_2^{d-1} \\ \vdots & & & \\ 0 & x_n - x_1 & & (x_n - x_1)x_n^{d-1} \end{bmatrix}}_{F''} = \prod_{i=2}^{n}(x_i - x_1) \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & x_2 \cdots & x_2^{d-1} \\ \vdots & & & \\ 0 & 1 & x_n \cdots & x_n^{d-1} \end{bmatrix}$$

$n = d+1$

$$\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & x_2^{d-1} \\ \vdots & & & \\ 0 & 1 & \cdots & x_n^{d-1} \end{bmatrix} \xRightarrow{-R_2} \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & x_2^{d-1} \\ \vdots & 0 & & \\ 0 & 0 & \cdots & x_n^{d-1} - x_2^{d-1} \end{bmatrix} \xRightarrow{C_i - x_2 C_{i-1}} \begin{bmatrix} 1 & 0 & \cdots & 0 \\ & 1 & 0 & 0 \\ & \vdots & & \\ & 0 & & x_n^{d-2}(x_n - x_2) \end{bmatrix} = \prod_{i=2}^{n}(x_i - x_1) \prod_{j=3}^{n}(x_j - x_2) \cdot \begin{bmatrix} 1 & 0 & \cdots \\ 0 & 1 & \cdots \\ \vdots & & 1\ x_3 \cdots \end{bmatrix}$$

$\vdots$

$$\det(F) = \prod_{1 \le i < j \le n} (x_j - x_i)$$

$\boxed{f)}$ $x_1\ x_2 \cdots x_\ell\ 1$ assign d to $\ell+1$ / Thus $\binom{\ell+d}{\ell}$.

$\boxed{g)}$

$$F_\ell = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} \cdots & x_{1,\ell} & \overbrace{x_{1,1}^2 \ x_{1,2}^2 \cdots x_{1,\ell-1} x_{1,\ell}}^{C^{\ell-1}_{d+\ell-1}} & \cdots & \overbrace{x_{1,1}^d \cdots x_{1,1} \cdots x_{1,d}}^{C^{\ell-1}_{d+\ell-1}} \\ 1 & x_{2,1} & & & & & \\ \vdots & \vdots & & & & & \\ 1 & x_{n,1} & & & & & \end{bmatrix}$$

$$F_\ell \Big|_{x_{i,1} = x_{i,2} = \cdots = x_{i,\ell} = x_i} = \begin{bmatrix} 1 & \alpha_1 & \alpha_1 \cdots & \alpha_1 & \alpha_{11}^2 & \cdots & \alpha_{1,1}^2 & \alpha_{11}^3 \cdots & \alpha_{11}^d \\ 1 & \alpha_2 & \alpha_2 \cdots & \alpha_2 & & & & & \\ \vdots & & & & & & & & \\ 1 & \alpha_n & \alpha_n & & & & & & \end{bmatrix}$$

the rank is always $d+1$, because all columns with same degree are linear dependent.

3.6 f 0 / 0

✓ - 0 pts -

$\boxed{e)}$ We already know that row/column subtraction doesn't change the determinant.

$$F = \begin{bmatrix} 1 & x_1 & \cdots & x_1^d \\ \vdots & & & \\ 1 & x_n & \cdots & x_n^d \end{bmatrix}_{n=d+1} \xRightarrow{-R_1} F' = \begin{bmatrix} 1 & x_1 & \cdots & x_1^d \\ 0 & & & x_2^d - x_1^d \\ \vdots & & & \\ 0 & x_n - x_1 & & x_n^d - x_1^d \end{bmatrix} \xRightarrow{C_i - x_1 C_{i-1}} F'' = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & x_2 - x_1 & (x_2-x_1)x_2 \cdots & (x_2-x_1)x_2^{d-1} \\ \vdots & & & \\ 0 & x_n - x_1 & & (x_n-x_1)x_n^{d-1} \end{bmatrix} = \prod_{i=2}^{n}(x_i - x_1) \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & x_2 \cdots & x_2^{d-1} \\ \vdots & & & \\ 0 & 1 & x_n \cdots & x_n^{d-1} \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & x_2^{d-1} \\ \vdots & & & \\ 0 & 1 & \cdots & x_n^{d-1} \end{bmatrix} \xRightarrow{-R_2} \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & x_2^{d-1} \\ \vdots & 0 & & \\ 0 & 0 & & x_n^{d-1} - x_2^{d-1} \end{bmatrix} \xRightarrow{C_i - x_2 C_{i-1}} \begin{bmatrix} 1 & 0 & \cdots & 0 \\ & 1 & 0 & 0 \\ & \vdots & & \\ & 0 & & x_n^{d-2}(x_n - x_2) \end{bmatrix} = \prod_{i=2}^{n}(x_i - x_1)\prod_{j=3}^{n}(x_j - x_2) \cdot \begin{bmatrix} 1 & 0 & \cdots \\ 0 & 1 & \cdots \\ \vdots & & 1 \; x_3 \cdots \end{bmatrix}$$

$\vdots$

$$\det(F) = \prod_{1 \le i < j \le n} (x_j - x_i)$$

$\boxed{f)}$ $x_1 \; x_2 \cdots x_\ell \; 1$ assign d to $\ell+1$ | Thus $\binom{\ell+d}{\ell}$.

$\boxed{g)}$

$$F_\ell = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} \cdots & x_{1,\ell} & \underbrace{x_{1,1}^2 \; x_{1,2}^2 \cdots x_{1,\ell-1}x_{1,\ell}}_{C^{\ell-1}_{d+\ell-1}} & \cdots & \underbrace{x_{1,1}^d \cdots x_{1,1} \cdots x_{1,d}}_{C^{\ell-1}_{d+\ell-1}} \\ 1 & x_{2,1} & & & & & \\ \vdots & \vdots & & & & & \\ 1 & x_{n,1} & & & & & \end{bmatrix}$$

$$F_\ell \Big|_{x_{i,1} = x_{i,2} = \cdots = x_{i,\ell} = \alpha_i} = \begin{bmatrix} 1 & \alpha_1 & \alpha_1 \cdots & \alpha_1 & \alpha_{1,1}^2 \cdots & \alpha_{1,1}^2 & \alpha_{1,1}^3 \cdots & \alpha_{1,1}^d \\ 1 & \alpha_2 & \alpha_2 \cdots & \alpha_2 & & & & \\ \vdots & & & & & & & \\ 1 & \alpha_n & \alpha_n & & & & & \end{bmatrix}$$

the rank is always $d+1$, because all columns with same degree are linear dependent.

3.7 g 10 / 10

✓ + **10 pts** **Correctly identifies that columns produced by same degree will be identical to each other and therefore lin dependent.**

+ **7 pts** Showed linear dependence but incorrect notion of feature matrix (Saying all columns are same. Only columns of same degree are identical).

+ **4 pts** Tries to show linear independence but argument or interpretation is not quite sound

+ **0 pts** Completely Wrong/Answer Missing

e) We already know that row/column subtraction doesn't change the determinant.

$$F = \begin{bmatrix} 1 & x_1 & \cdots & x_1^d \\ \vdots & & & \\ 1 & x_n & \cdots & x_n^d \end{bmatrix} \underset{n=d+1}{} \xrightarrow{-R_1} \overset{F'}{\begin{bmatrix} 1 & x_1 & \cdots & x_1^d \\ 0 & & & x_2^d - x_1^d \\ \vdots & & & \\ 0 & x_n - x_1 & & x_n^d - x_1^d \end{bmatrix}} \xrightarrow{C_i - x_1 C_{i-1}} \overset{F''}{\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & x_2 - x_1 & (x_2-x_1)x_2 \cdots & (x_2-x_1)x_2^{d-1} \\ \vdots & & & \\ 0 & x_n - x_1 & & (x_n-x_1)x_n^{d-1} \end{bmatrix}} = \prod_{i=2}^{n}(x_i - x_1) \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & x_2 \cdots & x_2^{d-1} \\ \vdots & & & \\ 0 & 1 & x_n \cdots & x_n^{d-1} \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & x_2^{d-1} \\ \vdots & & & \\ 0 & 1 & \cdots & x_n^{d-1} \end{bmatrix} \xrightarrow{-R_2} \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & x_2^{d-1} \\ \vdots & 0 & & \\ 0 & 0 & \cdots & x_n^{d-1} - x_2^{d-1} \end{bmatrix} \xrightarrow{C_i - x_2 C_{i-1}} \begin{bmatrix} 1 & 0 & \cdots & 0 \\ & 1 & 0 & 0 \\ & \vdots & & \\ & 0 & & x_n^{d-2}(x_n - x_2) \end{bmatrix} = \prod_{i=2}^{n}(x_i - x_1) \prod_{j=3}^{n}(x_j - x_2) \cdot \begin{bmatrix} 1 & 0 & \cdots \\ 0 & 1 & \\ & & 1 \; x_3 \cdots \end{bmatrix}$$

$\vdots$

$$\det(F) = \prod_{1 \le i \le j \le n} (x_j - x_i)$$

f) $x_1 \; x_2 \cdots x_\ell \; 1$ assign d to $\ell+1$ / Thus $\binom{\ell+d}{\ell}$.

g)
$$F_\ell = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} \cdots & x_{1,\ell} & \overbrace{x_{1,1}^2 \; x_{1,2}^2 \cdots x_{1,\ell-1} x_{1,\ell}}^{C^{\ell-1}_{2+\ell-1}} & \cdots & \overbrace{x_{1,1}^d \cdots x_{1,1} \cdots x_{1,d}}^{C^{\ell-1}_{d+\ell-1}} \\ 1 & x_{2,1} & & & & & \\ \vdots & \vdots & & & & & \\ 1 & x_{n,1} & & & & & \end{bmatrix}$$

$$F_\ell\Big|_{x_{i,1} = x_{i,2} = \cdots = x_{i,\ell} = \alpha_i} = \begin{bmatrix} 1 & \alpha_1 & \alpha_1 \cdots & \alpha_1 & \alpha_{11}^2 \cdots & \alpha_{11}^2 & \alpha_{11}^3 \cdots & \alpha_{11}^d \\ 1 & \alpha_2 & \alpha_2 \cdots & \alpha_2 & & & & \\ \vdots & & & & & & & \\ 1 & \alpha_n & \alpha_n & & & & & \end{bmatrix}$$

the rank is always $d+1$, because all columns with same degree are linear dependent.

3.8 h **0 / 10**

+ **10 pts** Fully correct

+ **6 pts** Relates solution to picking univariate monomials from the previous parts but doesn't have correct vectors

+ **2 pts** Hits some related points in explanation but doesn't choose correct vectors

✓ + **0 pts** **Blank or incorrect**

Hb2-4

$\boxed{a)}$ Assume $R_n(x)$ is the Remainder of Taylor Approximation, n is the order of Approximation.

$$f(x) = P_n(x) + R_n(x) \qquad R_n(x) = \frac{f^{(n+1)}(z)}{(n+1)!}(x-c)^{n+1}$$

z: is a value between c, x
C: is the point of Taylor Appro

(Error at $x=3$)

1st $\quad e^x = 1 + x + R_1(x) \qquad R_1(x) = \frac{f^{(2)}(z)}{2!}(x-0)^2\Big|_{x=3} \le \frac{e^3}{2}\times 9$

2nd $\quad e^x = 1 + x + \frac{1}{2!}x^2 + R_2(x) \qquad R_2(x) = \frac{f^{(3)}(z)}{3!}(x-0)^3\Big|_{x=3} \le \frac{e^3}{6}\times 27$

3rd $\quad e^x = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + R_3(x) \qquad R_3(x) = \frac{f^{(3)}(z)}{4!}(x-0)^4 \le \frac{e^3}{24}\times 81$

4th $\quad e^x = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \frac{1}{4!}x^4 + R_4(x) \qquad R_4(x) = \frac{f^{(5)}(z)}{5!}(x-0)^5 \le \frac{e^3}{120}\times 243$

2nd $\quad \sin x = x + R_2(x) \qquad R_2(x) = \frac{f^{(3)}(z)}{3!}(x-0)^3\Big|_{x=3} \le \frac{1}{6}\times 27$

3rd $\quad \sin x = x - \frac{x^3}{3!} + R_3(x) \qquad R_3(x) = \frac{f^{(4)}(z)}{4!}(x-0)^4\Big|_{x=3} \le \frac{1}{24}\times 81$

4th $\quad \sin x = x - \frac{x^3}{3!} + R_4(x) \qquad R_4(x) = \frac{f^{(5)}(z)}{5!}(x-0)^5\Big|_{x=3} \le \frac{1}{120}\times 243$

$\boxed{b)}$ $\quad |f(x) - \phi_D(x)| \le \varepsilon$

D=n $\quad |R_D(x)| \le \varepsilon \qquad \left|\frac{f^{(D+1)}(z)}{(D+1)!}(x-c)^{D+1}\right| \le \varepsilon$

for $e^x \quad \left|\frac{e^3}{(D+1)!}3^{D+1}\right| \le \varepsilon \qquad \frac{3^{D+1}}{(D+1)!} \le \frac{\varepsilon}{e^3}$

$\sin x \quad \left|\frac{1}{(D+1)!}3^{D+1}\right| \le \varepsilon \qquad \frac{3^{D+1}}{(D+1)!} \le \varepsilon$

D	$\frac{3^{D+1}}{(D+1)!}$
1	4.5
2	4.5
⋮	3.375
10	0.00444

$\boxed{c)}$

$$\lim_{m\to\infty} \frac{(x-x_0)^{m+1}}{(m+1)!} = 0$$

for part $x > x_0$ $\quad \frac{(x-x_0)^{m+1}}{(m+1)!} \le \frac{(x-x_0)}{1}\cdot\frac{(x-x_0)}{2}\cdots\frac{(x-x_0)}{m+1}$

Assume a_0 is the smallest integer where $|x-x_0| < a_0$

$$\lim_{m\to 0} \frac{(x-x_0)}{1}\cdots\frac{x-x_0}{a_0-1}\cdot\frac{x-x_0}{a_0}\cdots\frac{x-x_0}{m+1} \le \lim_{m\to\infty}\frac{x-x_0}{1}\cdots\frac{x-x_0}{a_0-1}\cdot\left(\frac{x-x_0}{a_0}\right)^{m-a_0} = 0$$

4.1 a 10 / 10

✓ + **10 pts** part i and part ii correct

+ **5 pts** part i correct
+ **5 pts** part ii correct
+ **5 pts** expansions correct
+ **3 pts** partially correct part i
+ **3 pts** partially correct part ii
+ **0 pts** Incorrect or blank

HW2-4

$\boxed{a)}$ Assume $R_n(x)$ is the Remainder of Taylor Approximation, n is the order of Approximation.

$f(x) = P_n(x) + R_n(x)$ $\quad R_n(x) = \dfrac{f^{(n+1)}(z)}{(n+1)!}(x-c)^{n+1}$ $\quad$ z: is a value between c, x

c: is the point of Taylor Appro

(Error at $x=3$)

1st $\quad e^x = 1 + x + R_1(x)$ $\quad R_1(x) = \dfrac{f^{(2)}(z)}{2!}(x-0)^2\Big|_{x=3} \le \dfrac{e^3}{2} \times 9$

2nd $\quad e^x = 1 + x + \dfrac{1}{2!}x^2 + R_2(x)$ $\quad R_2(x) = \dfrac{f^{(3)}(z)}{3!}(x-0)^3\Big|_{x=3} \le \dfrac{e^3}{6} \times 27$

3rd $\quad e^x = 1 + x + \dfrac{1}{2!}x^2 + \dfrac{1}{3!}x^3 + R_3(x)$ $\quad R_3(x) = \dfrac{f^{(4)}(z)}{4!}(x-0)^4 \le \dfrac{e^3}{24} \times 81$

4th $\quad e^x = 1 + x + \dfrac{1}{2!}x^2 + \dfrac{1}{3!}x^3 + \dfrac{1}{4!}x^4 + R_4(x)$ $\quad R_4(x) = \dfrac{f^{(5)}(z)}{5!}(x-0)^5 \le \dfrac{e^3}{120} \times 243$

2nd $\quad \sin x = x + R_2(x)$ $\quad R_2(x) = \dfrac{f^{(3)}(z)}{3!}(x-0)^3\Big|_{x=3} \le \dfrac{1}{6} \times 27$

3rd $\quad \sin x = x - \dfrac{x^3}{3!} + R_3(x)$ $\quad R_3(x) = \dfrac{f^{(4)}(z)}{4!}(x-0)^4\Big|_{x=3} \le \dfrac{1}{24} \times 81$

4th $\quad \sin x = x - \dfrac{x^3}{3!} + R_4(x)$ $\quad R_4(x) = \dfrac{f^{(5)}(z)}{5!}(x-0)^5\Big|_{x=3} \le \dfrac{1}{120} \times 243$

$\boxed{b)}$ $\quad |f(x) - \phi_D(x)| \le \varepsilon$

$D=n$ $\quad |R_D(x)| \le \varepsilon$ $\quad \left| \dfrac{f^{(D+1)}(z)}{(D+1)!}(x-c)^{D+1} \right| \le \varepsilon$

for e^x $\quad \left| \dfrac{e^3}{(D+1)!} 3^{D+1} \right| \le \varepsilon$ $\quad \dfrac{3^{D+1}}{(D+1)!} \le \dfrac{\varepsilon}{e^3}$

$\sin x$ $\quad \left| \dfrac{1}{(D+1)!} 3^{D+1} \right| \le \varepsilon$ $\quad \dfrac{3^{D+1}}{(D+1)!} \le \varepsilon$

D	$\dfrac{3^{D+1}}{(D+1)!}$
1	4.5
2	4.5
⋮	3.375
10	0.00444

$\boxed{c)}$

$$\lim_{m\to\infty} \frac{(x-x_0)^{m+1}}{(m+1)!} = 0$$

for part $x > x_0$ $\quad \dfrac{(x-x_0)^{m+1}}{(m+1)!} \le \dfrac{(x-x_0)}{1} \cdot \dfrac{(x-x_0)}{2} \cdots \dfrac{(x-x_0)}{m+1}$

Assume a_0 is the smallest integer where $|x-x_0| < a_0$

$$\lim_{m\to\infty} \frac{(x-x_0)}{1} \cdots \frac{x-x_0}{a_0-1} \cdot \frac{x-x_0}{a_0} \cdots \frac{x-x_0}{m+1} \le \lim_{m\to\infty} \frac{x-x_0}{1} \cdots \frac{x-x_0}{a_0-1} \cdot \left(\frac{x-x_0}{a_0}\right)^{m-a_0} = 0$$

4.2 b 0 / 0

✓ - 0 pts -

HW2-4

$\boxed{a)}$ Assume $R_n(x)$ is the Remainder of Taylor Approximation, n is the order of Approximation.

$f(x) = P_n(x) + R_n(x)$ $\quad R_n(x) = \frac{f^{(n+1)}(z)}{(n+1)!}(x-c)^{n+1}$ $\quad$ z: is a value between c, x
C: is the point of Taylor Appro

(Error at $x=3$)

1st $\quad e^x = 1 + x + R_1(x)$ $\quad R_1(x) = \frac{f^{(2)}(z)}{2!}(x-0)^2 \Big|_{x=3} \le \frac{e^3}{2} \times 9$

2nd $\quad e^x = 1 + x + \frac{1}{2!}x^2 + R_2(x)$ $\quad R_2(x) = \frac{f^{(3)}(z)}{3!}(x-0)^3 \Big|_{x=3} \le \frac{e^3}{6} \times 27$

3rd $\quad e^x = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + R_3(x)$ $\quad R_3(x) = \frac{f^{(3)}(z)}{4!}(x-0)^4 \le \frac{e^3}{24} \times 81$

4th $\quad e^x = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \frac{1}{4!}x^4 + R_4(x)$ $\quad R_4(x) = \frac{f^{(5)}(z)}{5!}(x-0)^5 \le \frac{e^3}{120} \times 243$

2nd $\quad \sin x = x + R_2(x)$ $\quad R_2(x) = \frac{f^{(3)}(z)}{3!}(x-0)^3 \Big|_{x=3} \le \frac{1}{6} \times 27$

3rd $\quad \sin x = x - \frac{x^3}{3!} + R_3(x)$ $\quad R_3(x) = \frac{f^{(4)}(z)}{4!}(x-0)^4 \Big|_{x=3} \le \frac{1}{24} \times 81$

4th $\quad \sin x = x - \frac{x^3}{3!} + R_4(x)$ $\quad R_4(x) = \frac{f^{(5)}(z)}{5!}(x-0)^5 \Big|_{x=3} \le \frac{1}{120} \times 243$

$\boxed{b)}$ $\quad |f(x) - P_D(x)| \le \varepsilon$
$D=n$ $\quad |R_D(x)| \le \varepsilon$ $\quad \left| \frac{f^{(D+1)}(z)}{(D+1)!}(x-c)^{D+1} \right| \le \varepsilon$

for e^x $\quad \left| \frac{e^3}{(D+1)!} 3^{D+1} \right| \le \varepsilon$ $\quad \frac{3^{D+1}}{(D+1)!} \le \frac{\varepsilon}{e^3}$

$\sin x$ $\quad \left| \frac{1}{(D+1)!} 3^{D+1} \right| \le \varepsilon$ $\quad \frac{3^{D+1}}{(D+1)!} \le \varepsilon$

D	$\frac{3^{D+1}}{(D+1)!}$
1	4.5
2	4.5
$\vdots$	3.375
10	0.00444

$\boxed{c)}$

$$\lim_{m \to \infty} \frac{(x-x_0)^{m+1}}{(m+1)!} = 0$$

~~for part $x > x_0$ $\frac{(x-x_0)^{m+1}}{(m+1)!} \le \frac{(x-x_0)}{1} \cdot \frac{(x-x_0)}{2} \cdots \frac{(x-x_0)}{m+1}$~~

~~Assume a_0 is the smallest integer where $|x-x_0| < a_0$~~

$$\lim_{m \to \infty} \frac{(x-x_0)}{1} \cdots \frac{x-x_0}{a_0-1} \cdot \frac{x-x_0}{a_0} \cdots \frac{x-x_0}{m+1} \le \lim_{m \to \infty} \frac{x-x_0}{1} \cdots \frac{x-x_0}{a_0-1} \cdot \left(\frac{x-x_0}{a_0}\right)^{m-a_0} = 0$$

4.3 C **0 / 0**

✓ **- 0 pts** -

$\boxed{d)}$ $\vec{x} = [x, y]^T$ $\vec{x}_0 = [x_0, y_0]^T$ $f(\vec{x}) = f(x, y)$

$$f(x,y) = f(\vec{x}_0) + [f_x \; f_y](\vec{x} - \vec{x}_0) + \frac{1}{2!}(\vec{x} - \vec{x}_0)^T \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix} (x - \vec{x}_0) + R_2(\vec{x})$$

$\boxed{e)}$ $f([x(t)\; y(t)]^T) = e^{x(t)} \sin y(t)$

$$f'_t = \frac{\partial f([x(t)\; y(t)]^T)}{\partial t} = [f_x \; f_y] \begin{bmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{bmatrix} = \frac{1}{\sqrt{2}}\left(e^{x(t)} \sin y(t) + e^{x(t)} \cos y(t)\right)$$

$$f''_t = f_x \frac{\partial^2 x}{\partial t^2} + f_y \frac{\partial^2 y}{\partial t^2} + f_{xx}\left(\frac{\partial x}{\partial t}\right)^2 + 2 f_{xy} \frac{\partial x}{\partial t}\frac{\partial y}{\partial t} + f_{yy}\left(\frac{\partial y}{\partial t}\right)^2 = 0 + 0 + \cancel{\sin y(t) e^{x(t)} \cdot \left(\frac{1}{\sqrt{2}}\right)^2} + 2 \cdot e^{x(t)} \cos y(t) \left(\frac{1}{\sqrt{2}}\right)^2 + \cancel{e^{x(t)}(-\sin y(t))\frac{1}{\sqrt{2}}}$$

$$= e^{x(t)} \cos y(t)$$

4

$$f'''_t = \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t} = e^{x(t)}\left(\cos y(t) - \sin y(t)\right)\frac{1}{\sqrt{2}}$$

$$f(\vec{x}(t)) = f([x(t_0), y(t_0)]^T) + f'([x(t_0), y(t_0)])(t - t_0) + f''([x(t_0), y(t_0)]^T)(t - t_0)^2 \frac{1}{2!} + f'''([x(t_0), y(t_0)]^T)\frac{1}{3!} + R_3(t)$$

$$= 0 + \frac{1}{\sqrt{2}} t + \frac{1}{2!} t^2 + \frac{1}{3!}\frac{1}{\sqrt{2}} t^3 + R_3(t)$$

$\boxed{f)}$ $|R_D(t)| \le \varepsilon$ $\quad R_3(t) = \frac{f^{(D)}(z)}{D!}(t - 0)^D \le \varepsilon$

$$f^{(4)}(z)\Big|_{z \in (0,3)} \le \left| e^{\frac{3}{\sqrt{2}}} \right| \le$$

$$f^{(4)}(t) = e^{x(t)}(\cos y(t) - \sin y(t))\frac{1}{2} + e^{x(t)}(-\sin y(t) - \cos y(t))\frac{1}{2}$$

$$= -e^{x(t)} \sin y(t)$$

$$\frac{e^{\frac{3}{\sqrt{2}}}}{D!} 3^D \le \varepsilon$$

4.4 d 10 / 10

✓ + 10 pts Correct

+ 9 pts Mostly correct/minor mistake

+ 6 pts A significant mistake

+ 4 pts show effort, but incorrect

+ 0 pts Incorrect

$\boxed{d)}$ $\vec{x} = [x, y]^T$ $\vec{x_0} = [x_0, y_0]^T$ $f(\vec{x}) = f(x, y)$

$$f(x,y) = f(\vec{x_0}) + [f_x \; f_y](\vec{x} - \vec{x_0}) + \frac{1}{2!}(\vec{x} - \vec{x_0})^T \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix} (x - \vec{x_0}) + R_2(\vec{x})$$

$\boxed{e)}$ $f([x(t)\; y(t)]^T) = e^{x(t)} \sin y(t)$

$$f_t' = \frac{\partial f([x(t)\; y(t)]^T)}{\partial t} = [f_x \; f_y] \begin{bmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{bmatrix} = \frac{1}{\sqrt{2}}\left(e^{x(t)} \sin y(t) + e^{x(t)} \cos y(t)\right)$$

$$f_t'' = f_x \frac{\partial^2 x}{\partial t^2} + f_y \frac{\partial^2 y}{\partial t^2} + f_{xx}\left(\frac{\partial x}{\partial t}\right)^2 + 2 f_{xy} \frac{\partial x}{\partial t}\frac{\partial y}{\partial t} + f_{yy}\left(\frac{\partial y}{\partial t}\right)^2 = 0 + 0 + \cancel{\sin y(t)\, e^{x(t)}} \cdot \left(\frac{1}{\sqrt{2}}\right)^2 + 2 \cdot e^{x(t)} \cos y(t) \left(\frac{1}{\sqrt{2}}\right)^2 + \cancel{e^{x(t)}(-\sin y(t))}\left(\frac{1}{\sqrt{2}}\right)^2$$

$$= e^{x(t)} \cos y(t)$$

$$f_t''' = \frac{\partial \ddot{f}}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial \ddot{f}}{\partial y}\frac{\partial y}{\partial t} = e^{x(t)}\left(\cos y(t) - \sin y(t)\right)\frac{1}{\sqrt{2}}$$

$$f(\vec{x}(t)) = f([x(t_0), y(t_0)]^T) + f'([x(t_0), y(t_0)])(t - t_0) + f''([x(t_0), y(t_0)]^T)(t - t_0)^2 \frac{1}{2!} + f'''([x(t_0), y(t_0)]^T)^3 \frac{1}{3!} + R_3(t)$$

$$= 0 + \frac{1}{\sqrt{2}} t + \frac{1}{2!} t^2 + \frac{1}{3!}\frac{1}{\sqrt{2}} t^3 + R_3(t)$$

$\boxed{f)}$ $|R_D(t)| \leq \varepsilon$ $\quad R_3(t) = \frac{f^{(D)}(z)}{D!}(t - 0)^D \leq \varepsilon$

$$f^{(4)}(z)\Big|_{z \in (0,3)} \leq \left| e^{\frac{3}{\sqrt{2}}} \right| \leq$$

$$f^{(4)}(t) = e^{x(t)}(\cancel{\cos y(t)} - \sin y(t))\frac{1}{2} + e^{x(t)}(-\sin y(t) - \cancel{\cos y(t)})\frac{1}{2}$$

$$= -e^{x(t)} \sin y(t)$$

$$\frac{e^{\frac{3}{\sqrt{2}}}}{D!} 3^D \leq \varepsilon$$

4.5 e **10 / 10**

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

$\boxed{d)}$ $\vec{x} = [x, y]^T$ $\vec{x}_0 = [x_0, y_0]^T$ $f(\vec{x}) = f(x, y)$

$$f(x,y) = f(\vec{x}_0) + [f_x \ f_y](\vec{x} - \vec{x}_0) + \frac{1}{2!}(\vec{x} - \vec{x}_0)^T \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix} (x - \vec{x}_0) + R_2(\vec{x})$$

$\boxed{e)}$ $f([x(t)\ y(t)]^T) = e^{x(t)} \sin y(t)$

$$f'_t = \frac{\partial f([x(t)\ y(t)]^T)}{\partial t} = [f_x \ f_y] \begin{bmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{bmatrix} = \frac{1}{\sqrt{2}}\left(e^{x(t)} \sin y(t) + e^{x(t)} \cos y(t)\right)$$

$$f''_t = f_x \frac{\partial^2 x}{\partial t^2} + f_y \frac{\partial^2 y}{\partial t^2} + f_{xx}\left(\frac{\partial x}{\partial t}\right)^2 + 2 f_{xy} \frac{\partial x}{\partial t}\frac{\partial y}{\partial t} + f_{yy}\left(\frac{\partial y}{\partial t}\right)^2 = 0 + 0 + \cancel{\sin y(t) e^{x(t)}} \cdot \left(\frac{1}{\sqrt{2}}\right)^2 + 2 \cdot e^{x(t)} \cos y(t) \left(\frac{1}{\sqrt{2}}\right)^2 + \cancel{e^{x(t)}(-\sin y(t))}\frac{1}{\sqrt{2}}$$

$$= e^{x(t)} \cos y(t)$$

4

$$f'''_t = \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t} = e^{x(t)}\left(\cos y(t) - \sin y(t)\right)\frac{1}{\sqrt{2}}$$

$$f(\vec{x}(t)) = f([x(t_0), y(t_0)]^T) + f'([x(t_0), y(t_0)])(t - t_0) + f''([x(t_0), y(t_0)]^T)(t - t_0)^2 \frac{1}{2!} + f'''([x(t_0), y(t_0)]^T)\frac{1}{3!} + R_3(t)$$

$$= 0 + \frac{1}{\sqrt{2}} t + \frac{1}{2!} t^2 + \frac{1}{3!}\frac{1}{\sqrt{2}} t^3 + R_3(t)$$

$\boxed{f)}$ $|R_D(t)| \le \varepsilon$ $\quad R_3(t) = \frac{f^{(D)}(z)}{D!}(t - 0)^D \le \varepsilon$

$$f^{(4)}(z)\Big|_{z \in (0,3)} \le \left| e^{\frac{3}{\sqrt{2}}} \right| \le$$

$$f^{(4)}(t) = e^{x(t)}(\cos y(t) - \sin y(t))\frac{1}{2} + e^{x(t)}(-\sin y(t) - \cos y(t))\frac{1}{2}$$

$$= -e^{x(t)} \sin y(t)$$

$$\frac{e^{\frac{3}{\sqrt{2}}}}{D!} 3^D \le \varepsilon$$

4.6 f **0 / 0**

✓ **- 0 pts** Correct

$\boxed{d)}$ $\vec{x} = [x, y]^T$ $\vec{x_0} = [x_0, y_0]^T$ $f(\vec{x}) = f(x,y)$

$$f(x,y) = f(\vec{x_0}) + [f_x \ f_y](\vec{x} - \vec{x_0}) + \frac{1}{2!}(\vec{x} - \vec{x_0})^T \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix} (x - \vec{x_0}) + R_2(\vec{x})$$

$\boxed{e)}$ $f([x(t)\ y(t)]^T) = e^{x(t)} \sin y(t)$

$$f_t' = \frac{\partial f([x(t)\ y(t)]^T)}{\partial t} = [f_x \ f_y] \begin{bmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{bmatrix} = \frac{1}{\sqrt{2}}\left(e^{x(t)} \sin y(t) + e^{x(t)} \cos y(t)\right)$$

$$f_t'' = f_x \frac{\partial^2 x}{\partial t^2} + f_y \frac{\partial^2 y}{\partial t^2} + f_{xx}\left(\frac{\partial x}{\partial t}\right)^2 + 2 f_{xy} \frac{\partial x}{\partial t}\frac{\partial y}{\partial t} + f_{yy}\left(\frac{\partial y}{\partial t}\right)^2 = 0 + 0 + \cancel{\sin y(t) e^{x(t)}} \cdot \left(\frac{1}{\sqrt{2}}\right)^2 + 2 \cdot e^{x(t)} \cos y(t) \left(\frac{1}{\sqrt{2}}\right)^2 + \cancel{e^{x(t)}(-\sin y(t))} \left(\frac{1}{\sqrt{2}}\right)^2$$

$$= e^{x(t)} \cos y(t)$$

4

$$f_t''' = \frac{\partial f''}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f''}{\partial y}\frac{\partial y}{\partial t} = e^{x(t)}\left(\cos y(t) - \sin y(t)\right)\frac{1}{\sqrt{2}}$$

$$f(\vec{x}(t)) = f([x(t_0), y(t_0)]^T) + f'([x(t_0), y(t_0)])(t - t_0) + f''([x(t_0), y(t_0)]^T)(t - t_0)^2 \frac{1}{2!} + f'''([x(t_0), y(t_0)]^T)^3 \frac{1}{3!} + R_3(t)$$

$$= 0 + \frac{1}{\sqrt{2}} t + \frac{1}{2!} t^2 + \frac{1}{3!}\frac{1}{\sqrt{2}} t^3 + R_3(t)$$

$\boxed{f)}$ $|R_D(t)| \le \varepsilon$ $\quad R_3(t) = \frac{f^{(D)}(z)}{D!}(t - 0)^D \le \varepsilon$

$$f^{(4)}(z)\Big|_{z \in (0,3)} \le \left| e^{\frac{3}{\sqrt{2}}} \right| \le$$

$$f^{(4)}(t) = e^{x(t)}(\cancel{\cos y(t)} - \sin y(t))\frac{1}{2} + e^{x(t)}(-\sin y(t) - \cancel{\cos y(t)})\frac{1}{2}$$

$$= -e^{x(t)} \sin y(t)$$

$$\frac{e^{\frac{3}{\sqrt{2}}}}{D!} 3^D \le \varepsilon$$

4.7 g **0 / 0**

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

HW02-5

a)

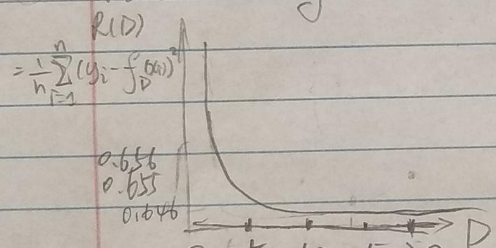
$$\min_{\vec{w}} \left\| \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & \cdots & x_1^D \\ 1 & x_2 & x_2^2 & x_2^3 & \cdots & x_2^D \\ \vdots & & \vdots & & & \\ 1 & x_n & x_n^2 & x_n^3 & \cdots & x_n^D \end{bmatrix} \vec{w} - \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \right\|_2$$

where $\vec{w} = [w_1\ w_2 \cdots w_n]^T$

b)

?

Plot training error



c) As D increase the training error becomes small.

When using n degree as model, the $(X^T X)$ is not invertable.

Thus it has infinite solutions.

d) The error in fresh data is larger than training error

e) I would chose 4, because that is where our model start overfiting.

5.1 a 0 / 0

✓ - 0 pts -

HW02-5

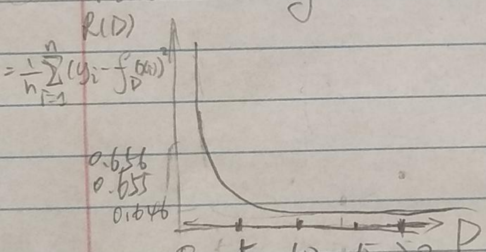
a)

$$\min_{\vec{w}} \left\| \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & \cdots & x_1^D \\ 1 & x_2 & x_2^2 & x_2^3 & \cdots & x_2^D \\ \vdots & & \vdots & & & \\ 1 & x_n & x_n^2 & x_n^3 & \cdots & x_n^D \end{bmatrix} \vec{w} - \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \right\|_2$$

where $\vec{w} = [w_1\ w_2 \cdots w_n]^T$

b) ?

Plot training error



c) As D increase the training error becomes small.

When using n degree as model, the (X^TX) is not invertable.

Thus it has infinite solutions.

d) The error in fresh data is larger than training error

e) I would choose 4, because that is where our model start overfitting.

5.2 b 0 / 0

✓ - 0 pts -

HW02 -5

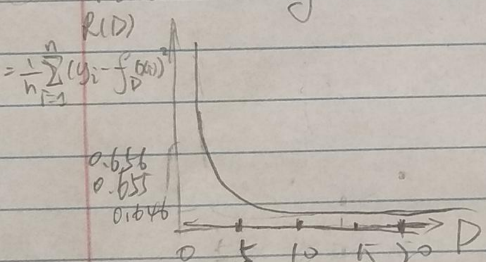
a)

$$\min_{\vec{w}} \left\| \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & \cdots & x_1^D \\ 1 & x_2 & x_2^2 & x_2^3 & \cdots & x_2^D \\ \vdots & & \vdots & & & \\ 1 & x_n & x_n^2 & x_n^3 & \cdots & x_n^D \end{bmatrix} \vec{w} - \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \right\|_2$$

where $\vec{w} = [w_1 \; w_2 \cdots w_n]^T$

b) ?

Plot training error



c) As D increase the training error becomes small.

When using n degree as model, the (X^TX) is not invertable.

Thus it has infinite solutions.

d) The error in fresh data is larger than training error

e) I would chose 4, because that is where our model start overfitting.

5.3 C **0 / 0**

✓ **- 0 pts** -

HW02-5

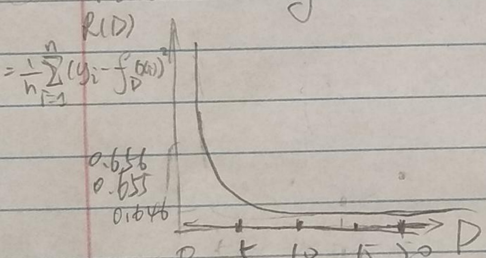
a)

$$\min_{\vec{w}} \left\| \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & \cdots & x_1^D \\ 1 & x_2 & x_2^2 & x_2^3 & \cdots & x_2^D \\ \vdots & & \vdots & & & \\ 1 & x_n & x_n^2 & x_n^3 & \cdots & x_n^D \end{bmatrix} \vec{w} - \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \right\|_2$$

where $\vec{w} = [w_1 \ w_2 \cdots w_n]^T$

b) ?

Plot training error



c) As D increase the training error becomes small.

When using n degree as model, the (X^TX) is not invertable.

Thus it has infinite solutions.

d) The error in fresh data is larger than training error

e) I would chose 4, because that is where our model start overfitting.

5.4 d 0 / 10

+ **10 pts** Correct
+ **5 pts** Partially Correct
+ **3 pts** Penalty

✓ + **0 pts** Incorrect

No plot for d?

HW02-5

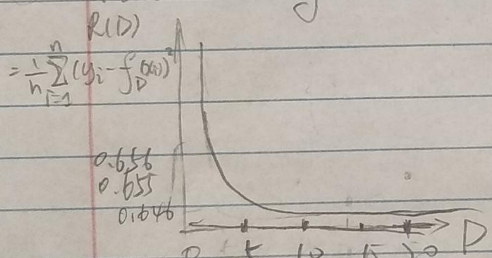
a)

$$\min_{\vec{w}} \left\| \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & \cdots & x_1^D \\ 1 & x_2 & x_2^2 & x_2^3 & \cdots & x_2^D \\ \vdots & & \vdots & & & \\ 1 & x_n & x_n^2 & x_n^3 & \cdots & x_n^D \end{bmatrix} \vec{w} - \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \right\|_2$$

where $\vec{w} = [w_1 \; w_2 \cdots w_n]^T$

b) ?

Plot training error



c) As D increase the training error becomes small.

When using n degree as model, the $(X^T X)$ is not invertable.

Thus it has infinite solutions.

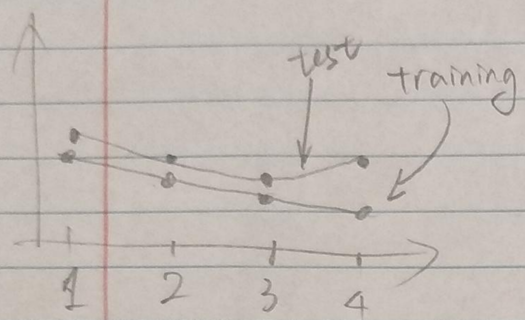
d) The error in fresh data is larger than training error

e) I would chose 4, because that is where our model start overfitting.

5.5 e 0 / 0

✓ - 0 pts -

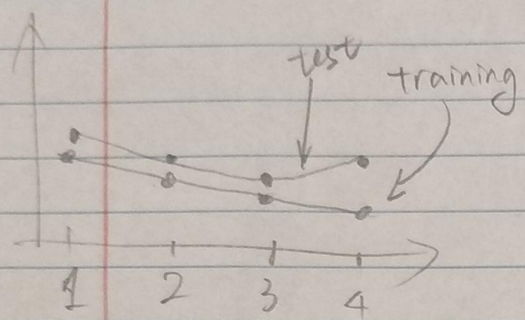
$\boxed{f)}$ I will choose ~~three~~ 3.



5.6 f 0 / 0

✓ - 0 pts -

f) I will choose ~~three~~ 3.



5.7 g **0 / 10**

- \+ **10 pts** Correct
- \+ **8 pts** Small Mistake
- \+ **5 pts** Partial
- \+ **3 pts** Penalty
- ✓ + **0 pts** **Incorrect/Empty**

As well as

HO2-2 / HO2-6 My own Question: What is the relationship between Ridge & Lagrange?

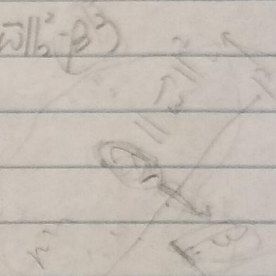
$\boxed{a)}$ The Conclusion is as β becomes large, λ becomes small.
If $\beta \geq \beta_0$, λ becomes 0.
Where $\beta_0 = \left\| \underset{\vec{w}}{\operatorname{argmin}} \|\vec{y} - X\vec{w}\|_2^2 \right\|_2$

(Relation between Lagrange & Ridge Regression)

$\boxed{\text{Explanation 1:}}$ Optimization Problem

Note: minimize $f(\vec{w})$ s.t. $g(\vec{w}) \leq 0$ — Standard form.

$$\underset{\vec{w}}{\text{minimize}} \|\vec{y} - X\vec{w}\|_2^2 \quad \text{s.t. } \|\vec{w}\|_2^2 \leq \beta^2 \xrightarrow[\text{Multiplier}]{\text{Using Lagrange}} L(\vec{w}, \lambda) = \|\vec{y} - X\vec{w}\|_2^2 + \lambda(\|\vec{w}\|_2^2 - \beta^2)$$



The Optimal Conditions: ① $\nabla L(\vec{w}, \lambda) = 0$

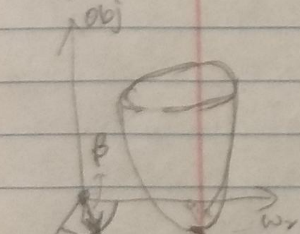
$$\frac{\partial L(\vec{w}, \lambda)}{\partial \vec{w}} = 0$$

② $\|w\|_2^2 \leq \beta^2$

③ $\lambda(\|\vec{w}\|_2^2 - \beta^2) = 0$

If $\lambda = 0$, $\|w\|_2^2 \leq \beta^2$
$\lambda > 0$, $\|\vec{w}\|_2^2 = \beta^2$ > That's where we need to check.

Instead of check it, think of this problem reversely.
First, note objective function is convex, inequality constraint is also convex.
The Conclusion is the global optimal lies either at the boundary or the extremum.
If it is at the boundary, correspond to "$\lambda > 0$, $\|w\|_2^2 = \beta^2$"



In this case, according to Optimal Condition,
there exists $\lambda^* > 0$ s.t. $\frac{\partial L(\vec{w}, \lambda^*)}{\partial \vec{w}} = 0$ (as $\frac{\partial \beta^2}{\partial \vec{w}} = 0$)

And this is exactly what Ridge Regression trys to solve.

Until Now the conclusion is "Ridge Regression is trying to solve
$\underset{\vec{w}}{\text{minimize}} \|\vec{y} - X\vec{w}\|_2$
s.t. $\|\vec{w}\|_2 \leq \beta^2$ w.r.t. the λ^* (corresponding to β)
(Omit the part where $\beta > \beta_0$ which is $\lambda = 0$ $\|w\|_2^2 \leq \beta^2$ section.)

$\boxed{\text{Explanation 2:}}$ when $\beta \uparrow$ $\lambda \downarrow$ (within $\beta < \beta_0$)

6 Your Own Question 10 / 10

✓ + **10 pts** Correct

+ **0 pts** No Answer