CS 189 Fall 2017 — Introduction to Machine Learning — HW2

This homework is due **Friday, September 8 at 10 p.m.**.

1 Getting Started

You may typeset your homework in latex or submit neatly handwritten and scanned solutions. Please make sure to start each question on a new page, as grading (with Gradescope) is much easier that way! Deliverables:

1. Submit a PDF of your writeup to assignment on Gradescope, "HW[n] Write-Up"
2. Submit all code needed to reproduce your results, "HW[n] Code".
3. Submit your test set evaluation results, "HW[n] Test Set".

After you've submitted your homework, be sure to watch out for the self-grade form.

(a) Before you start your homework, write down your team. Who else did you work with on this homework? List names and email addresses. In case of course events, just describe the group. How did you work on this homework? Any comments about the homework?

None
Comments : Homework is very long

(b) Please copy the following statement and sign next to it:
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.

I certify that all solutions are entirely in my words and that I have not looked @ another student's sols. I have credited all external source in this write up.

Problem #2

(a) $$\min \|\vec{y} - X\vec{w}\|_2^2$$

subject to $\|\vec{w}\|_2^2 \le \beta^2 \rightarrow \frac{1}{\beta^2}\|\vec{w}\|_2^2 \le 1$

$$\Rightarrow \min\left(\|\vec{y} - X\vec{w}\|_2^2 + \frac{1}{\beta^2}\|\vec{w}\|_2^2\right) \le \min\|\vec{y} - X\vec{w}\|_2^2 + 1$$

Ridge regression problem

$$\min\left(\|\vec{y} - X\vec{w}\|_2^2 + \underbrace{\frac{1}{\beta^2}}_{\alpha}\|\vec{w}\|_2^2\right)$$

$\beta \uparrow \rightarrow \alpha \downarrow$

(b) Prediction w/o disturbance : $\vec{w}^T\vec{x}$

with disturbance : $\vec{x} + \vec{\varepsilon}$

$\rightarrow$ new prediction : $\vec{w}^T(\vec{x} + \vec{\varepsilon})$

$\rightarrow$ change : $\vec{w}^T\vec{\varepsilon}$

(c) $$\min\left(\|\vec{y} - X\vec{w}\|_2^2 + \lambda\|\vec{w}\|_2^2\right)$$

$$\frac{d\left((\vec{y} - X\vec{w})^T(\vec{y} - X\vec{w}) + \lambda\vec{w}^T\vec{w}\right)}{d\vec{w}}$$

$$= \frac{d\left(\vec{y}^T\vec{y} - 2y^TX\vec{w} + \vec{w}^TX^TX\vec{w} + \lambda\vec{w}^T\vec{w}\right)}{dw}$$

$$= -2y^TX + 2\vec{w}^TX^TX + 2\lambda\vec{w}^T = \vec{0}$$

$$\vec{w}^T(X^TX + \lambda I) = \vec{y}^TX$$

$$(X^TX + \lambda I)\vec{w} = X^T\vec{y}$$

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

$\lambda \to \infty$ leads to $\vec{x} \to \vec{0}$, i.e. shrinking $\vec{x}$

(d) The eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ are the root of:

$$\det(X^T X - \alpha I) = 0$$

i.e $\quad \alpha = \{\lambda_1, \lambda_2, \dots, \lambda_n\}$

Consider the equation:

$$\det(X^T X + \lambda I - \beta I) = 0$$

$$\det(X^T X - (\beta - \lambda) I) = 0$$

$$\Rightarrow \quad \beta - \lambda = \alpha$$

$$\Rightarrow \quad \beta = \alpha + \lambda$$

Thus, the eigenvalues of $(X^T X + \lambda I)$ are $\{\lambda_1 + \lambda, \lambda_2 + \lambda, \dots, \lambda_n + \lambda\}$

$$X^T X = \frac{1}{\underset{\text{singular}}{\det(X^T X)}} (X^T X)^{*T} \quad \text{where } (X^T X)^* \text{ is cofactor of } X^T X$$

$$(X^T X + \lambda I)^{-1} = \frac{1}{\underset{\text{non singular}}{\det(X^T X + \lambda I)}} (X^T X + \lambda I)^{*T}$$

we know that $\det(X^T X) = \prod_{i=1}^{n} \lambda_i = 0$ if any $\lambda_i = 0$.

When we add them up with $\lambda \Rightarrow \det(X^T X + \lambda I) = \prod_{i=0}^{n} (\lambda_i + \lambda)$

so $(X^T X + \lambda I)$ is invertible.

(e) $\lambda_2 = 0.5$ is better since:

$$\det(X^T X) = 1$$

$$\det(X^T X + \lambda_1 I) = 1100 \times 101 \times 100{,}001 \sim O(10^7)$$

$$\det(X^T X + \lambda_2 I) = 1000.5 \times 1.5 \times 0.501 \sim O(10^2)$$

$\Rightarrow$ λ_1 will shrink $\vec{w}$ too much $\rightarrow$ λ_2 is better

(f) Expand $\vec{w} = \vec{w_0} + X^T\vec{a}$ for $\vec{w_0} \in$ null-space (X), $\vec{w_0} \neq \vec{0}$ since X has more columns than rows.

Ridge Regression: $\min \| X\vec{w} - \vec{y} \|_2^2 + \lambda \|\vec{w}\|_2^2$

From question (c), we come up with:

$$(X^T X + \lambda I)\vec{w} = X^T\vec{y}$$

$X \in R^{4\times5} \Rightarrow \text{rank}(X) = \text{rank}(X^T) \leq 4$

$\Rightarrow \text{rank}(X^T X) = \min(\text{rank}(X), \text{rank}(X^T)) \leq 4$, but $X^T X \in R^{5\times5}$

$\Rightarrow$ $X^T X$ singular, or $\det(X^T X) = 0$

$\Rightarrow$ $(X^T X + \lambda I)$ nonsingular

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

OLS: $\min \| X\vec{w} - \vec{y} \|_2^2$

From lecture: $X^T X \vec{w} = X^T \vec{y}$

$$X^T X(\vec{w_0} + X^T\vec{a}) = X^T\vec{y}$$

$$\underbrace{\cancel{X^T X \vec{w_0}}}_{\vec{0}} + X^T X X^T \vec{a} = X^T \vec{y}$$

$$X X^T \vec{a} = \vec{y}$$

$\text{rank}(XX^T) = 4 \Rightarrow \vec{a} = (XX^T)^{-1}\vec{y} \Rightarrow$ infinite number of solutions $\vec{w} = \vec{w_0} + X^T\vec{a}$ because $\vec{w_0} \in$ null-space $(\vec{a})$ infinite

(g) $\lambda \to 0$ leads to OLS, i.e no bound for $\vec{w}$

(h) $$\min \left(\| \vec{y} - X\vec{w} \|_2^2 + \lambda \| \Gamma \vec{w} \|_2^2 \right)$$

Use vector calculus:

$$\frac{d\left(\| \vec{y} - X\vec{w} \|_2^2 + \lambda \| \Gamma \vec{w} \|_2^2 \right)}{d\vec{w}} = -2\vec{y}^T X + 2\vec{w}^T X^T X + 2\lambda \vec{w}^T \Gamma^T \Gamma = \vec{0}$$

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

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

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

Problem #3

(a) $n = 2$ & $D = 1$:

$$F = [\ \vec{p}_1(x_1)\ \ \vec{p}_2(x_2)\ ]^T = \begin{bmatrix} 1 & 1 \\ x_1 & x_2 \end{bmatrix}^T$$

$$\det(F) = x_2 - x_1 \neq 0 \quad \text{iif } x_1 \neq x_2$$

or F has full rank iff $x_1 \neq x_2$

(b)

$$\underbrace{\begin{bmatrix} 1 & & & \\ -1 & 1 & & \\ -1 & & 1 & \\ \vdots & & & \ddots \\ -1 & & & & 1 \end{bmatrix}}_{J} \underbrace{\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & & x_n \\ x_1^2 & x_2^2 & & x_n^2 \\ \vdots & \vdots & & \\ x_1^D & x_2^D & \cdots & x_n^D \end{bmatrix}}_{F^T} = \underbrace{\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 - 1 & x_2 - 1 & & x_n - 1 \\ x_1^2 - 1 & x_2^2 - 1 & & x_n^2 - 1 \\ & \vdots & & \\ x_1^D - 1 & x_2^D - 1 & \cdots & x_n^D - 1 \end{bmatrix}}_{F'^T}$$

$$JF^T = F'^T$$

$$\Rightarrow \det(JF^T) = \det(F'^T)$$

$$\Rightarrow \det(J)\det(F^T) = \det(F'^T)$$

$$\det(J) = 1 \times \begin{vmatrix} 1 & & \\ & 1 & \\ & & \ddots \\ & & & 1 \end{vmatrix} - 0 \times \begin{vmatrix} 1 & & \\ -1 & 1 & \\ & & \ddots \\ & & & 1 \end{vmatrix} + 0 \times \begin{vmatrix} 1 & & \\ 1 & 1 & \\ & & \ddots \\ & & & 1 \end{vmatrix} \cdots$$

$$= 1$$

$$\Rightarrow \det(F^T) = \det(F'^T)$$

$$\text{or } \det(F) = \det(F')$$

(c)

$$F' = \begin{bmatrix} 1 & x_1 - 1 & x_1^2 - 1 & \dots & x_1^D - 1 \\ 1 & x_2 - 1 & x_2^2 - 1 & \dots & x_2^D - 1 \\ & & \dots & & \\ 1 & x_n - 1 & x_n^2 - 1 & \dots & x_n^D - 1 \end{bmatrix}$$

$$F'' = \begin{bmatrix} 1 & -1 & x_1 - 1 & x_1 - 1 & & x_1 - 1 \\ 1 & -1 & x_2(x_2 - x_1) + x_1 - 1 & x_2^2(x_2 - x_1) + x_1 - 1 & \dots & x_2^{D-1}(x_2 - x_1) + x_1 - 1 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 1 & -1 & x_n(x_n - 1) + x_1 - 1 & x_n^2(x_n - x_1) + x_1 - 1 & & x_n^{D-1}(x_n - 1) + x_1 - 1 \end{bmatrix}$$

We prove that if we subtract a column from α * othe column, the determinant does not change.

say $A = [\vec{a}_1 \ \vec{a}_2 \dots \vec{a}_n] \rightarrow B = [\vec{a}_1 \ \vec{a}_2 \dots \vec{a}_i - \alpha \vec{a}_j \dots \vec{a}_n]$

we have

$$\det B = \det([\vec{a}_1^T \ \vec{a}_2^T \dots \vec{a}_n^T]) - \det([\vec{a}_1^T \ \vec{a}_2^T \dots \alpha \vec{a}_j^T \dots \vec{a}_n^T])$$

$$= \det A - \alpha \times \underbrace{\det([\vec{a}_1^T \ \vec{a}_2^T \dots \vec{a}_j^T \dots \vec{a}_j^T \dots \vec{a}_n^T])}_{0 \text{ since there are 2 identical columns}}$$

$$\Rightarrow \det B = \det A$$

Thus, a sequence of transformations acting on F' to make F'' does not change the determinant of F'

$$\Rightarrow \det(F') = \det(F'')$$

Another approach is:

$$F'' = F' \underbrace{\begin{bmatrix} 1 & -x_1 & 0 & & \\ & 1 & -x_1 & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \end{bmatrix}}_{\det(\,) = 1} \Rightarrow \det(F'') = \det(F')$$

(d)

$$B = \begin{bmatrix} 1 & \vec{0}^T \\ \vec{0}^T & A \end{bmatrix}$$

Eigenvalues of B is the roots of equation:

$$\det(B - \lambda I_{d+1}) = 0$$

$$\det \begin{vmatrix} 1-\lambda_0 & \vec{0}^T \\ \vec{0}^T & A - \lambda I_d \end{vmatrix} = 0$$

$$(1-\lambda_0) \times \det(A - \lambda I_d) = 0$$

$$\Rightarrow \lambda_0 = 1 \text{ or } \lambda \in \{\lambda_1(A), \lambda_2(A), \dots, \lambda_d(A)\}$$

$\Rightarrow$ eigenvalue of B are $\{1, \lambda_1(A), \lambda_2(A), \dots, \lambda_d(A)\}$

$$\det(B) = 1 \times \det(A) - 0 \times \det(\dots) + 0 \times \det(\dots) - \dots$$
$$= \det(A)$$

(e) For $D = 1$ $(n = 2)$

$$F_1 = \begin{bmatrix} 1 & 1 \\ x_1 & x_2 \end{bmatrix}^T \Rightarrow \det F = x_2 - x_1 = \prod_{1 \le i \le j \le 2} (x_j - x_i)$$

Assuming that it is correct upto $D = k$ $(n = k+1)$, i.e.

$$F_k = \begin{bmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_k \\ x_1^k & x_2^k & x_k^k \end{bmatrix}^T \text{ and } \det F_k = \prod_{1 \le i \le j \le k+1} (x_j - x_i)$$

We prove that it is correct upto $D = k+1$ $(n = k+2)$

$$F_{k+1} = \begin{bmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_{k+1} \\ \vdots & \vdots & \\ x_1^{k+1} & x_2^{k+1} & x_{k+1}^{k+1} \end{bmatrix}^T$$

We implement a sequence of transformations

Subtract $x_1 \times \text{column}_1$ from column_2

Subtract $x_1 \times \text{column}_2$ from column_3

Subtract $x_1 \times \text{column}_3$ from column_4

...

Subtract $x_1 \times \text{column}_{n-1}$ from column_n

$$\begin{bmatrix} 1 & x_1 & \cdots & x_1^{k+1} \\ 1 & x_2 & & x_2^{k+1} \\ 1 & x_3 & & x_3^{k+1} \\ \vdots & \vdots & & \vdots \\ 1 & x_{k+1} & \cdots & x_{k+1}^{k+1} \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & x_2 - x_1 & x_2(x_2 - x_1) & & x_2^k(x_2 - x_1) \\ 1 & x_3 - x_1 & x_3(x_3 - x_1) & & x_3^k(x_3 - x_1) \\ \vdots & \vdots & & & \vdots \\ 1 & x_{k+1} - x_1 & x_{k+1}(x_{k+1} - x_1) & \cdots & x_{k+1}^k(x_k - x_1) \end{bmatrix}$$

$$\det(\ ) = 1 \times \begin{vmatrix} x_2 - x_1 & x_2(x_2 - x_1) & \cdots & x_2^k(x_2 - x_1) \\ & \cdots & & \\ x_{k+1} - x_1 & x_{k+1}(x_{k+1} - x_1) & \cdots & x_{k+1}^k(x_k - x_1) \end{vmatrix}$$

$$= (x_2 - x_1)(x_3 - x_1) \cdots (x_{k+1} - x_1) \det(p_k(x_2), p_k(x_3) \cdots p_k(x_k))$$

$$= \prod_{i=2}^{n} (x_i - x_1) \det F_k$$

but $\det F_k = \prod_{2 \le i \le j \le k+1} (x_j - x_i)$

$$\Rightarrow \quad \det F_{k+1} = \prod_{1 \le i \le j \le k+1} (x_j - x_i)$$

(f) D - degree , ℓ variables

$\rightarrow$ we have D stars

we have $\ell + 1$ bars (the last bar is 1 because we need adding up $\leq D$)

we have $D + \ell + 1 - 1 = D + \ell$ positions

from which we choose $\ell + 1 - 1$ bar positions

$\Rightarrow$ choose ℓ out of $D + 1$ without order.

$$\Rightarrow \binom{D+\ell}{\ell}$$

(g)

Problem # 4

(a) $f(x) = e^x$

$f'(x) = e^x$

$f''(x) = e^x$

2^{nd}, $m=2$: $f(x) = f(x_0) + f'(x_0)(x-x_0) + \frac{1}{2} f''(x_0)(x-x_0)^2 + \mathcal{E}(x)$

$= e^0 + e^0(x-0) + \frac{1}{2} e^0 (x-0)^2 + \mathcal{E}(x)$

$= 1 + x + \frac{1}{2}x^2 + \mathcal{E}(x) \overset{x=3}{=} 8.5 + \mathcal{E}(3)$

$x=3$: $|f''(3)| = e^3 = T$

Error bound : $\frac{T(x-x_0)^{m+1}}{(m+1)!} = \frac{e^3(3-0)^3}{3!} = \frac{9e^3}{2}$

3^{rd}, $m=3$: $f(x) = 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \mathcal{E}(x) \overset{x=3}{=} 13 + \mathcal{E}(3)$

$x=3$: $|f'''(3)| = e^3 = T$

$\Rightarrow \mathcal{E}$ bound $= \frac{e^3 3^4}{4!}$

4^{th}, $m=4$: $f(x) = 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \frac{1}{24}x^4 + \mathcal{E}(x) \overset{x=3}{=} 16.375 + \mathcal{E}(3)$

$x=3$: $|f''''(3)| = e^3 = T$

$\Rightarrow \mathcal{E}$ bound $= \frac{e^3 3^5}{5!}$

$f(x) = \sin x$

$f'(x) = \cos x$

$f''(x) = -\sin(x)$

$f'''(x) = -\cos(x)$

$f''''(x) = \sin x$

2nd
$m=2$

$$f(x) = \cancel{\sin 0}^{\,0} + \cos(0)(x-0) - \frac{1}{2}\cancel{\sin(0)}^{\,0}(x-x_0)^2 + \varepsilon(x)$$

$$= x + \varepsilon(x) \overset{x=3}{=} 3 + \varepsilon(3)$$

$x=3$: $|f''(3)| = \sin 3 = T$

Error bound : $\dfrac{T(x-x_0)^{m+1}}{(m+1)!} = \dfrac{\sin 3 \times 3^3}{3!} = \dfrac{9\sin 3}{2}$

3rd
$m=3$

$$f(x) = x - \frac{1}{6}x^3 + \varepsilon(x) \overset{x=3}{=} -1.5 + \varepsilon(3)$$

$x=3$: $|f'''(3)| = \cos(3) = T$

Error bound : $\dfrac{T(x-x_0)^{m+1}}{(m+1)!} = \dfrac{\cos 3 \times 3^4}{4!}$

4th
$m=4$

$$f(x) = x - \frac{1}{6}x^3 + 0 + \varepsilon(x) \overset{x=3}{=} -1.5 + \varepsilon(3)$$

$x=3$: $|f''''(3)| = \sin(3) = T$

Error bound : $\dfrac{T(x-x_0)^{m+1}}{(m+1)!} = \dfrac{\sin 3 \times 3^5}{5!}$

(b) $f(x) = e^x$

We have $f^{(D)}(x) = e^x$

$$x \in [0,3] \Rightarrow |f^{(D)}(x)| \le e^3 = T$$

Error bound:

$$\frac{T\,(x-x_0)^{m+1}}{(m+1)!} \le \frac{e^3 (3-0)^{D+1}}{(D+1)!} \le \varepsilon$$

more conservative: $\dfrac{e^3\, 3^{D+1}}{\left(\frac{D+1}{2}\right)^{\frac{D+1}{2}}} \le \varepsilon$

$$\left(\frac{9}{\frac{D+1}{2}}\right)^{\frac{D+1}{2}} \le \frac{\varepsilon}{e^3}$$

$$\frac{D+1}{2} \ln \frac{9}{\frac{D+1}{2}} \le \ln \frac{\varepsilon}{e^3}$$

$$\frac{D+1}{2} \ln \frac{1}{9}\frac{D+1}{2} \ge \ln \frac{e^3}{\varepsilon}$$

$f(x) = \sin x$

$$x \in [0,3] \Rightarrow |f^{(D)}(x)| \le 1$$

Error bound:

$$\frac{1\,(3-0)^{D+1}}{(D+1)!} \le \varepsilon$$

more conservative: $\dfrac{9^{\frac{D+1}{2}}}{\left(\frac{D+1}{2}\right)^{\frac{D+1}{2}}} \le \varepsilon$

$$\frac{D+1}{2} \ln \frac{1}{9}\frac{D+1}{2} \ge \ln \frac{1}{\varepsilon}$$

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

we have $\left(\frac{m+1}{2}\right)^{\frac{m+1}{2}} < (m+1)! < (m+1)^{m+1}$

$$\left(\frac{x-x_0}{m+1}\right)^{m+1} \leq \frac{(x-x_0)^{m+1}}{(m+1)!} < \left(\frac{(x-x_0)^2}{\frac{m+1}{2}}\right)^{\frac{m+1}{2}}$$

$$e^{\ln\left(\frac{x-x_0}{m+1}\right)^{m+1}} \leq \ldots \leq e^{\ln\left(\frac{(x-x_0)^2}{\frac{m+1}{2}}\right)^{\frac{m+1}{2}}}$$

$$e^{(m+1)\ln\frac{x-x_0}{m+1}} \leq \ldots \leq e^{\frac{m+1}{2}\ln\frac{(x-x_0)^2}{\frac{m+1}{2}}}$$

$$1/e^{(m+1)\ln\frac{m+1}{x-x_0}} \leq \ldots \leq 1/e^{\frac{m+1}{2}\ln\frac{\frac{m+1}{2}}{\frac{(x-x_0)^2}{2}}}$$

$m \to \infty$: $(m+1)\ln\frac{m+1}{x-x_0}$ and $\frac{m+1}{2}\ln\frac{\frac{m+1}{2}}{\frac{(x-x_0)^2}{2}} \to \infty$

$\Rightarrow$ LHS & RHS $\to 0$

$\Rightarrow$ middle $\to 0$

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

This means errors bound $\frac{T(x-x_0)^{m+1}}{(m+1)!} \to 0$ if m is large enough

$\rightarrow$ polynomial can approximate any function that is sufficiently smooth.

(d)

$$f(\vec{x}) = f(\vec{x}_0) + \nabla f(\vec{x}_0)^T \cdot (\vec{x} - \vec{x}_0)$$
$$+ \frac{1}{2!}(\vec{x} - \vec{x}_0)^T \cdot \left(H(\vec{x}_0) \cdot (\vec{x} - \vec{x}_0) \right) + \varepsilon(\vec{x})$$

where :

$$\nabla f(\vec{x}_0) = \begin{bmatrix} f_x(x_0, y_0) \\ f_y(x_0, y_0) \end{bmatrix}$$

$$H(\vec{x}_0) = \begin{bmatrix} f_{xx}(x_0, y_0) & f_{xy}(x_0, y_0) \\ f_{yx}(x_0, y_0) & f_{yy}(x_0, y_0) \end{bmatrix}$$

(e)

$$f(\vec{x}) = e^x \sin y \quad , \quad \vec{x}(t) = \left[\frac{1}{\sqrt{2}}t, \frac{1}{\sqrt{2}}t\right]^T \quad \vec{x}_0 \overset{@ t=0}{=} [0, 0]^T$$

$$f_x(\vec{x}_0) = e^{x_0} \sin y_0 = 0$$

$$f_y(\vec{x}_0) = e^{x_0} \cos y_0 = 1$$

$$f_{xx}(\vec{x}_0) = 0 \quad f_{xy}(\vec{x}_0) = 1 \quad f_{yx}(\vec{x}_0) = 1 \quad f_{yy}(\vec{x}_0) = 0$$

$$f_{xxx}(\vec{x}_0) = 0 \quad f_{xxy}(x_0) = 1 \quad f_{xyx}(x_0) = 1 \quad f_{xyy}(x_0) = 0$$

$$f_{yxx}(\vec{x}_0) = 1 \quad f_{yxy}(x_0) = 0 \quad f_{yyx}(x_0) = 0 \quad f_{yyy}(x_0) = -1$$

$$f(\vec{x}_0) = 0$$

$$\Rightarrow f(\vec{x}) = 0 + 0 + y + \frac{1}{2}[0 + xy + xy + 0]$$
$$+ \frac{1}{6}[0 + x^2y + x^2y + 0 + x^2y + 0 + 0 - y^3]$$
$$= y + xy + \frac{1}{2}x^2y - \frac{1}{6}y^3$$

(f) $t \in [0, 3]$

$$\Rightarrow \vec{x}^{T} \leqslant \left[\frac{3}{\sqrt{2}}, \frac{3}{\sqrt{2}} \right]^{T} \qquad \| \vec{x}^{T} \| \leqslant 3$$

$$\Rightarrow | f^{(0)}(\vec{x}) | \leqslant e^{x} = e^{3/\sqrt{2}} = T$$

Error bound:

$$\frac{T \| \vec{x} - \vec{x}_0 \|^{m+1}}{(m+1)!} = \frac{e^{3/\sqrt{2}} \, 3^{m+1}}{(m+1)!} \leq \varepsilon$$

Similar to question (b)

$$\frac{D+1}{2} \ln \frac{1}{g} \frac{D+1}{2} \geqslant \ln \frac{e^{3/\sqrt{2}}}{\varepsilon}$$

Problem # 5

(a) $p_D(x) = \sum_{i=0}^{D} a_i x^i = [\, x^0 \; x^1 \; x^2 \ldots x^D \,] \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_D \end{bmatrix}$

$p_D(x_1) = y_1 + \varepsilon_1$

$p_D(x_2) = y_2 + \varepsilon_2$

...

$p_D(x_n) = y_n + \varepsilon_n$

$$\Rightarrow \underbrace{\begin{bmatrix} x_1^0 & x_1^1 & \cdots & x_1^D \\ x_2^0 & x_2^1 & \cdots & x_2^D \\ & \cdots & & \\ x_n^0 & x_n^1 & \cdots & x_n^D \end{bmatrix}}_{X} \underbrace{\begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_D \end{bmatrix}}_{\vec{a}} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}}_{\vec{y}} + \varepsilon$$

$\Rightarrow \min \| X\vec{a} - \vec{y} \|_2^2$

$\rightarrow$ linear regression problem

(b) See code & next page

(c) See diagram next page, in both 1D_poly.mat and 1D_poly_new.mat: $D \nearrow \rightarrow \varepsilon \searrow$ then $\nearrow$

@ $D = n$. The matrix X^TX singular

(d) This plot has a little bigger error than training plot because it is validation data

See next page for plot & error.

e.) Use (d) to choose hyperparameter $d =$

Use (b) to choose coefficients

See next page

f.) See next page

(g) see next page.