1.

(a) $d_{vc}(H) = D+1 = 57+1 = 58$

(b) $E_{out}(h_g) \le E_{in}(h_g) + \sqrt{\frac{8}{N}\ln\frac{4[(2N)^{d_{vc}}+1]}{\delta}}$ $\qquad \delta = 0.1$

$$E_{out}(h_g) \le E_{in}(h_g) + \sqrt{\frac{8}{3000}\ln\frac{4[(2\cdot 3000)^{58}+1]}{0.1}}$$

$$E_{out}(h_g) \le E_{in}(h_g) + 1.1642$$

(c) $D=10 \qquad d_{vc}(H)=11 \qquad N=10000$

$$\varepsilon = \sqrt{\frac{8}{N}\ln\frac{4[(2N)^{d_{vc}}+1]}{\delta}} = \sqrt{\frac{8}{10000}\ln\frac{4[(2\cdot 10000)^{11}+1]}{0.1}} = 0.3002$$

(d) $\varepsilon = \sqrt{\frac{8}{N}\ln\frac{4[(2N)^{58}+1]}{\delta}} = 0.1$

$$\frac{8}{N}\ln\frac{4[(2N)^{58}+1]}{0.1} = 0.01$$

$$\ln\frac{4[(2N)^{58}+1]}{0.1} = \frac{0.01N}{8}$$

$$[(2N)^{58}+1] = \frac{e^{\frac{0.01N}{8}}\cdot 0.1}{4}$$

$$(2N)^{58}+1 = 0.025\, e^{\frac{0.01N}{8}}$$

$$N = 622555$$

若 $D=10$ $d_{vc}(H)=11$

$$\varepsilon = \sqrt{\frac{8}{N} \ln \frac{4[(2N)^{11}+1]}{0.1}} = 0.1$$

$$\frac{8}{N} \ln \frac{4[(2N)^{11}+1]}{0.1} = 0.01$$

$$\ln \frac{4[(2N)^{11}+1]}{0.1} = \frac{0.01 \cdot N}{8}$$

$$[(2N)^{11}+1] = e^{\frac{0.01 \cdot N}{8}} \cdot \frac{0.1}{4}$$

$$N = 111307$$

e. The best model selected from trainning dataset models cardinality is 1.

f. $\varepsilon = \sqrt{\frac{1}{2N} \ln \frac{2M}{\delta}} = \sqrt{\frac{1}{2 \cdot 1601} \cdot \ln \frac{2}{0.1}} = 0.03$

2.4.

(a) set $x_i = \begin{bmatrix} 1 \\ a_1 \\ a_2 \\ \vdots \\ a_d \end{bmatrix}$

$$\underline{\underline{X}} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & \cdots \\ a_1 & b_1 & & & & \\ a_2 & b_2 & & & & \\ \vdots & \vdots & & & & \\ a_d & b_d. & & & & \end{bmatrix} \quad (d+1) \times (d+1) \text{ matrix}$$

$$h(\underline{\underline{X}}) = \underline{W}^T \underline{\underline{X}} = \underbrace{\begin{bmatrix} w_{a0} & w_{b0} & \\ w_{a_1} & w_{b1} & \\ w_{a_2} & \vdots & \cdots \\ \vdots & \vdots & \\ w_{ad}. & w_{bd} & \end{bmatrix}^T}_{(d+1) \times (d+1)} \times \underbrace{\begin{bmatrix} 1 & 1 & 1 & 1 & \cdots \\ a_1 & \ddots & & & \\ a_2 & & \ddots & & \\ \vdots & & & \ddots & \\ a_d & & & & \ddots \end{bmatrix}}_{(d+1) \times (d+1)}$$

$\underline{\underline{X}}$ is non singular, you can find W^T that can map.

$\underline{\underline{X}}$ so that $h(\underline{\underline{X}}) = \text{sgn}(\underline{\underline{W}}^T \underline{\underline{X}}) = \text{diag}(1)$

$W^T = X^{-1}$ since X is non singular.

so there are at least $d+1$ possible dichotomies.

$d_{vc} \geq d+1$

(b) $\underline{\underline{X}}$ is linearly-independent. so X_{d+2} term $= X_k$ must be a linear combination of other points.

$$X_k = X_{d+2} = \sum_{i=0}^{d} a_i x_i \qquad (d+1) \times (d+2) \qquad (d+1) \times (d+1)$$

assume $h(X_k) = \operatorname{sgn}(W_k^T X_k) = 0$ $\quad W_k^T$ can be any row of W^T

$h(X_k) = \operatorname{sgn}(W_i^T X_k) = \operatorname{sgn}(W_k^T \sum_{i=0}^{d} a_i X_i)$ has to be 1

since W_k^T is chosen from W^T and there is alway some $a_i X_i$ such that $\operatorname{sgn}(W_i^T a_i X_i) = 1$

There there is at most $d+1$ points.

$$d_{VC} \le d+1$$

so $\quad d_{VC} = d+1$

2.24.

a.1 $g^D(x) = ax+b$ $\qquad a = \frac{f(x)}{x} \qquad b=0$

$g(x) = \frac{y_1}{x_1} x \qquad$ set $y_i = f(x_i)$

a.2 $\qquad$ set $y_i = f(x_i)$

$y = ax+b \qquad (x_1, y_1) \qquad (x_2, y_2)$

$$a = \frac{y_1 - y_2}{x_1 - x_2} \qquad y_1 = \frac{y_1 - y_2}{x_1 - x_2} x_1 + b$$

$$b = y_1 - \frac{y_1 - y_2}{x_1 - x_2} x_1$$

$$b = \frac{\cancel{y_1 x_1} - y_1 x_2 - \cancel{y_1 x_1} + y_2 x_1}{x_1 - x_2}$$

$$g^D(x) = \frac{y_1 - y_2}{x_1 - x_2} x + \frac{y_2 x_1 - y_1 x_2}{x_1 - x_2}$$

$$\bar{g}(x) = E_D[g^D(x)] = E_D\left[\frac{y_1 - y_2}{x_1 - x_2} x + \frac{y_2 x_1 - y_1 x_2}{x_1 - x_2}\right]$$

$$= E_D\left[\frac{y_1 - y_2}{x_1 - x_2} x\right] + E_D\left[\frac{y_2 x_1 - y_1 x_2}{x_1 - x_2}\right]$$

since x is uniformably distributed between $[-1, 1]$ $p(x_i) = \frac{1}{2}$

$$= \int_{-1}^{1}\int_{-1}^{1} \frac{y_1 - y_2}{x_1 - x_2} \frac{1}{2} \cdot \frac{1}{2} \, dx_2 \, dx_1 + \int_{-1}^{1}\int_{-1}^{1} \frac{y_2 x_1 - y_1 x_2}{x_1 - x_2} \frac{1}{2} \cdot \frac{1}{2} \, dx_2 \, dx_1$$

$$= \frac{1}{4}\int_{-1}^{1}\int_{-1}^{1} \underbrace{\frac{x_1^2 - x_2^2}{x_1 - x_2}}_{\Downarrow \; x_1 + x_2} dx_2\, dx_1 + \frac{1}{4}\int_{-1}^{1}\int_{-1}^{1} \frac{\overset{x_1x_2(x_2 - x_1)}{x_2^2 x_1 - x_1^2 x_2}}{x_1 - x_2} dx_2\, dx_1$$

$$\Downarrow$$

$$(-x_1x_2)\, dx_2\, dx_1$$

$$= \frac{1}{4}\int_{-1}^{1} x_1x_2 + \frac{1}{2}x_2^2 \Big|_{-1}^{1} dx_1 + \frac{1}{4}\int_{-1}^{1} -\frac{1}{2}x_1x_2^2 \Big|_{-1}^{1} dx_1$$

$$x_1 + \frac{1}{2} - (-x_1 + \frac{1}{2})$$

$$\Downarrow$$

$$2x_1$$

$$\Downarrow$$

$$-\frac{1}{2}x_1 + \frac{1}{2}x_1 = 0.$$

$$= \frac{1}{4}\int_{-1}^{1} 2x_1\, dx_1 + 0$$

$$= \frac{1}{4}\, x_1^2 \Big|_{-1}^{1} + 0 = \frac{1}{4}\cdot 0 + 0 = 0.$$

(b) $\bar{g}(\underline{x})$ creat n datapoints that satisfy $y=x^2$

from that dataset choose m datasets. each dataset has two datapoints. Acording to these two datapoints predict a and b and take mean over all D.

$\bar{g}(x)=\bar{a}x+\bar{b}$

Ed(Eout) According to $\bar{a}$ and $\bar{b}$, construting $\bar{g}(x)$

evaluate $\overline{g(x)^2} - 2\overline{g(x)}f + f^2$ over all n datapoints and take the mean. value.

bias evaluate $(\bar{g}(x)-f)^2$ over all n datapoints

variance evaluate $g^D(x)-\bar{g}(x)$ over all m D

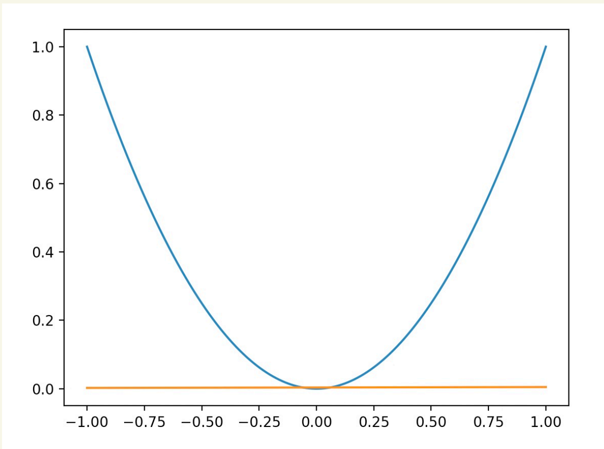
then evaluate over all x. take the mean value

(c)

```
mean of g(x): 0.0012758160000001643 * x + 0.003622166864800001
Ed(Eout): 0.5313779652362515
bias: 0.2024364889312275
variance: 0.33462168972675377

Process finished with exit code 0
```

$$E_d[E_{out}] = \text{bias} + \text{variance}$$



d.

$$E_{out} = E_x[(g(x) - f(x))^2] = E_x[((ax+b) - f(x))^2]$$

$$= E_x[(ax+b-x^2)^2]$$

$$= E_x[-2ax^3 + a^2x^2 + 2abx + b^2 - 2bx^2 + x^4]$$

$$= -2a\,E_x(x^3) + a^2 E_x(x^2) + 2ab\,E_x(x) + b^2 - 2b\,E(x^2) + E(x^4)$$

$$= -2a \cdot \frac{1}{2}\int_{-1}^{1} x^3 dx + a^2 \frac{1}{2}\int_{-1}^{1} x^2 dx + 2ab\frac{1}{2}\int_{-1}^{1} x\,dx + b^2 - 2b\frac{1}{2}\int_{-1}^{1} x^2 dx$$
$$+ \frac{1}{2}\int_{-1}^{1} x^4 dx$$

$$= -2a \cdot \frac{1}{2}\left(\frac{1}{4}x^4\Big|_{-1}^{1}\right) + \frac{a^2}{2}\left(\frac{1}{3}x^3\Big|_{-1}^{1}\right) + ab\left(\frac{1}{2}x^2\Big|_{-1}^{1}\right) + b^2$$
$$- b\left(\frac{1}{3}x^3\Big|_{-1}^{1}\right) + \frac{1}{2}\left(\frac{1}{5}x^5\Big|_{-1}^{1}\right)$$

$$= -a \times 0 + \frac{a^2}{2}\left(\frac{2}{3}\right) + ab(0) + b^2 - b\frac{2}{3} + \frac{1}{2}\cdot\frac{2}{5}$$

$$= \frac{a^2}{3} + b^2 - \frac{2b}{3} + \frac{1}{5}$$

$$= \frac{a^2 - 2b}{3} + b^2 + \frac{1}{5}$$

$$a = x_1 + x_2 \qquad b = -x_1 x_2$$

$$E_D(E_{out}) = E_D\left[\frac{a^2-2b}{3} + b^2 + \frac{1}{5}\right]$$

$$= E_D\left[\frac{(x_1+x_2)^2 + 2x_1x_2}{3}\right] + E_D\left[x_1^2 x_2^2\right] + \frac{1}{5}$$

$$= \frac{1}{4}\int_{-1}^{1}\int_{-1}^{1} \frac{x_1^2+x_2^2+4x_1x_2}{3}\,dx_1dx_2 + \frac{1}{4}\int_{-1}^{1}\int_{-1}^{1} x_1^2x_2^2\,dx_1dx_2 + 1/5$$

$$= \frac{1}{4}\int_{-1}^{1} \frac{\frac{1}{3}x_1^3 + x_1x_2^2 + 2x_1^2x_2}{3}\Bigg|_{-1}^{1} dx_2 + \frac{1}{4}\int_{-1}^{1} \frac{1}{3}x^3x_2^2\Big|_{-1}^{1} dx_2 + 1/5$$

$$= \frac{1}{4}\int_{-1}^{1} \frac{\frac{2}{3}+2x_2^2}{3}\,dx_2 + \frac{1}{4}\int_{-1}^{1}\frac{2}{3}x_2^2\,dx_2 + 1/5$$

$$= \frac{1}{4}\cdot\frac{1}{3}\left[\frac{2}{3}x_2 + \frac{2}{3}x^3\Big|_{-1}^{1}\right] + 1/4\left[\frac{2}{9}x_2^3\Big|_{-1}^{1}\right] + 1/5$$

$$= \frac{1}{12}\left[\frac{2}{3}+\frac{2}{3}-\left(-\frac{2}{3}-\frac{2}{3}\right)\right] + 1/4\left[\frac{2}{9}+\frac{2}{9}\right] + 1/5$$

$$= \frac{1}{12}\cdot\frac{8}{3} + \frac{1}{\cancel{4}}\cdot\frac{4}{9} + \frac{1}{5}$$

$$= \frac{2}{9} + \frac{1}{9} + 1/5 = \frac{8}{15}$$

$$\text{bias} = E_x[(\bar{g}(\underline{x}) - f(x))^2]$$

$$= E_x[(0 - x^2)^2]$$

$$= E_x[x^4] = \frac{1}{2}\int_{-1}^{1} x^4 \, dx = \frac{1}{2} \quad \frac{1}{5}x^5\Big|_{-1}^{1}$$

$$= \frac{1}{2}\left(\frac{1}{5} + \frac{1}{5}\right) = \frac{1}{5} = 0.2$$

$$\text{variance} = E_x[E_D[(g(x) - \bar{g}(x))^2]]$$

$$= E_x[E_D[(ax+b - 0)^2]] = E_x[E_D[a^2x^2 + 2abx + b^2]]$$

$$= E_x E_D(x_1+x_2)^2 x^2 - 2(x_1+x_2)x_1x_2 x + x_1^2x_2^2]$$

$$= E_x \frac{1}{4}\int_{-1}^{1}\int_{-1}^{1} x^2(x_1^2 + 2x_1x_2 + x_2^2) - 2(x_1^2x_2 + x_2^2x_1)x + x_1^2x_2^2 \, dx_1 dx_2$$

$$= E_x \frac{1}{4}\int_{-1}^{1} x^2\left(\frac{1}{3}x_1^3 + x_1^2x_2 + x_2^2x_1\right) - 2\left(\frac{1}{3}x_1^3x_2 + \frac{1}{2}x_1^2x_2^2\right)x + \frac{1}{3}x_1^3x_2^2\Big|_{-1}^{1} dx_2$$

$$x^2\left(\frac{1}{3} + x_2 + x_2^2\right) - 2\left(\frac{1}{3}x_2 + \frac{1}{2}x_2^2\right)x + \frac{1}{3}x_2^2$$

$$-\left(x^2\left(-\frac{1}{3} + x_2 - x_2^2\right) - 2\left(-\frac{1}{3}x_2 + \frac{1}{2}x_2^2\right)x - \frac{1}{3}x_2^2\right))$$

$$\underline{x^2\left(\frac{2}{3} + 2x_2^2\right) - 2\left(\frac{2}{3}x_2\right)x + \frac{2}{3}x_2^2}$$

$$= E_x \frac{1}{4} \left(x^2 \left(\frac{2}{3} x_2 + \frac{2}{3} x_2^3 \right) - \frac{4}{3} \frac{1}{2} x_2^2 x + \frac{2}{3} \cdot \frac{1}{3} x_2^3 \Big|_{-1}^{1} \right)$$

$$x^2 \left(\frac{2}{3} + \frac{2}{3} \right) - \frac{2}{3} x + \frac{2}{9} - \left[x^2 \left(-\frac{2}{3} - \frac{2}{3} \right) - \frac{2}{3} x - \frac{2}{9} \right]$$

$$= \frac{4}{3} x^2 - \frac{2}{3} x + \frac{2}{9} - \left[-\frac{4}{3} x^2 - \frac{2}{3} x - \frac{2}{9} \right]$$

$$= \left(\frac{8}{3} x^2 + \frac{4}{9} \right) \frac{1}{4}$$

$$= E_x \left(\frac{2}{3} x^2 + \frac{1}{9} \right)$$

$$= \frac{1}{2} \int_{-1}^{1} \frac{2}{3} x^2 + \frac{1}{9} \, dx$$

$$= \frac{1}{2} \left(\frac{2}{9} x^3 + \frac{1}{9} x \Big|_{-1}^{1} \right)$$

$$= \frac{1}{2} \left(\frac{2}{9} + \frac{1}{9} + \frac{2}{9} + \frac{1}{9} \right) = \frac{6}{9} \cdot \frac{1}{2} = \frac{1}{3}$$

4.3

(a) deterministic noise will decrease.

(b) deterministic noise will increase.

4.5

(a) $$W^T\Gamma^T\Gamma W = \sum_{q=0}^{Q} w_q^2$$

$$\Gamma^T\Gamma = I \qquad \Gamma = I$$

(b) $$W^T\Gamma^T\Gamma^T W = \left(\sum_{q=0}^{Q} w_q\right)^2$$

$$W^T\Gamma^T = \sum_{q=0}^{Q} w_q$$

$$\Gamma^T = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$