1. C

$$E = \int_0^2 (e^x - wx)^2 dx = \int_0^2 (e^{2x} - 2wxe^x + w^2x^2)\, dx$$

$$= \frac{1}{2}e^{2x} - 2w(xe^x - e^x) + \frac{1}{3}w^2x^3 \Big|_0^2$$

$$= \frac{1}{2}(e^4 - 1) - 2w(2e^2 - e^2 + 1) + \frac{1}{3}w^2 8$$

$$\frac{dE}{dw} = -2(e^2+1) + \frac{16}{3}w = 0$$

$$\Rightarrow w = \frac{3}{8}(e^2+1)$$

magnitude of deterministic noise

$$= |e^x - wx|$$

$$= \left|e^x - \frac{3}{8}(e^2+1)x\right|$$

2. b

$$h^* = \underset{h \in H}{\operatorname{argmin}} \; E_{out}(h)$$

$$A(D) = \underset{h \in H}{\operatorname{argmin}} \; E_{in}(h)$$

$\mathbb{E}_D\left[E_{in}(A(D)) \right] > \mathbb{E}_D\left[E_{out}(A(D)) \right]$ is wrong

3. d

$$X_h = \begin{bmatrix} | & & | & | & & | \\ x_1 & \cdots & x_N & \tilde{x}_1 & \cdots & \tilde{x}_N \\ | & & | & | & & | \end{bmatrix}^T$$

$\sigma^2 = E[\varepsilon^2]$

$= E[\varepsilon\varepsilon^T]$

$$X_h^T X_h = \begin{bmatrix} | & & | & | & & | \\ x_1 & - & x_N & \tilde{x}_1 & - & \tilde{x}_N \\ | & & | & | & & | \end{bmatrix} \begin{bmatrix} - & x_1 & - \\ & \vdots & \\ - & x_N & - \\ - & \tilde{x}_1 & - \\ & \vdots & \\ - & \tilde{x}_N & - \end{bmatrix}$$

$$= \begin{bmatrix} \sum x_{i,0}^2 + \underbrace{\sum \tilde{x}_{i,0}^2}_{\approx N\sigma^2} & \cdots & \sum x_{i,0} x_{i,d} + \underbrace{\sum \tilde{x}_{i,0} \tilde{x}_{i,d}}_{\approx 0} \\ \vdots & & \vdots \\ \sum x_{i,d} x_{i,0} + \underbrace{\sum \tilde{x}_{i,d} \tilde{x}_{i,0}}_{\approx 0} & & \sum x_{i,d}^2 + \underbrace{\sum \tilde{x}_{i,d}^2}_{\approx N\sigma^2} \end{bmatrix}$$

$= 2X^TX + N\sigma^2 I_{d+1}$

4. e

$\varepsilon \in R^{d+1}$

$x_n^T y_n$

$$= \begin{bmatrix} | & & | & | & & | \\ x_1 & \cdots & x_N & \tilde{x}_1 & \cdots & \tilde{x}_N \\ | & & | & | & & | \end{bmatrix} \begin{bmatrix} y \\ \\ y \end{bmatrix}$$

$$= \left[x^T y + \tilde{x}^T y \right]$$

$$= \left[2x^T y + \varepsilon \right]$$

$$E\left[x_n^T y_n\right] = \left[2x^T y \right]$$

5. (d)

$$X^T X = Q \Gamma Q^T$$

$$X_{n \times (d+1)} \qquad x_{(d+1) \cdot 1}$$

$$\underset{n \cdot (d+1)}{Z} = \underset{n \cdot (d+1)}{X} \, Q_{(d+1) \cdot (d+1)}$$

$$\min_{w \in \mathbb{R}^{d+1}} \frac{1}{N} \| Zw - y \|^2 + \frac{\lambda}{N} w^T w$$

$$\frac{2}{N} (Z^T Z w - Z^T y) + \frac{2}{N} \lambda w = 0$$

$$w \leftarrow (Z^T Z + \lambda I)^{-1} Z^T y$$

$$w \leftarrow (Q^T X^T X Q + \lambda I)^{-1} Q^T X^T y$$

$$w \leftarrow (Q^T Q \Gamma Q^T Q + \lambda I)^{-1} Q^T X^T y$$

$$w \leftarrow (\underbrace{\Gamma + \lambda I}_{(d+1) \cdot (d+1)})^{-1} \underbrace{Q^T X^T}_{(d+1) \cdot n} \underbrace{y}_{n \cdot 1}$$

$$w \leftarrow \left(\frac{1}{\Gamma + \lambda I} \right)$$

$$\frac{u_i}{v_i} = \frac{(\Gamma + \lambda I)^{-1}}{(\Gamma)^{-1}}$$

6. a

$$\min_{w \in \mathbb{R}} \frac{1}{N} \sum_{n=1}^{N} (\underbrace{w \cdot x_n - y_n)^2}_{w^2 x_n^2 - 2w x_n y_n + y_n^2} + \frac{\lambda}{N} w^2 = \xi$$

$$\frac{d\xi}{dw} = \frac{1}{N} \sum_{n=1}^{N} (2w x_n^2 - 2x_n y_n) + \frac{2\lambda}{N} w = 0$$

$$w \cdot \sum_{n=1}^{N} 2x_n^2 - \sum_{n=1}^{N} 2x_n y_n + 2\lambda w = 0$$

$$w \cdot \left(\sum_{n=1}^{N} 2x_n^2 + 2\lambda\right) = \sum_{n=1}^{N} 2x_n y_n$$

$$w = \frac{\sum_{n=1}^{N} x_n y_n}{\sum_{n=1}^{N} x_n^2 + \lambda}$$

7. d

$$\min_{y \in \mathbb{R}} \frac{1}{N} \sum_{n=1}^{n} \underbrace{(y - y_n)^2}_{y^2 - 2y y_n + y_n^2} + \frac{2K}{N} \Omega(y) = \varepsilon$$

$$\frac{d\varepsilon}{dy} = \frac{1}{N} \sum_{n=1}^{N} (\cancel{2}y - \cancel{2}y_n) + \frac{\cancel{2}K}{N} \Omega'(y) = 0$$

$$y - \frac{1}{N} \sum_{n=1}^{N} y_n = -\frac{K}{N} \Omega'(y)$$

$$\Omega'(y) = \frac{N}{K} \left(\frac{1}{N} \sum_{n=1}^{N} y_n - y \right)$$

$$= \frac{N}{K} \left(\frac{1}{N}(N+2K)y - \frac{K}{N} - y \right)$$

$$= \frac{N}{K} \left(\frac{2Ky}{N} - \frac{K}{N} \right)$$

$$= \frac{N}{K} \cdot \frac{K}{N} (2y - 1)$$

$$= (2y - 1)$$

$$\Omega(y) = \int (2y-1)\, dy = y^2 - y + C = (y - 0.5)^2 + C$$

8. b

$$\Phi(x) = \Gamma^{-1} x$$

$$\min_{\tilde{w} \in \mathbb{R}^{d+1}} \frac{1}{N} \sum_{n=1}^{N} \left(\tilde{w}^T \Phi(x_n) - y_n\right)^2 + \frac{\lambda}{N} (\tilde{w}^T \tilde{w})$$

$$\min_{\tilde{w} \in \mathbb{R}^{d+1}} \frac{1}{N} \sum_{n=1}^{N} \left(\tilde{w}^T \Gamma^{-1} x_n - y_n\right)^2 + \frac{\lambda}{N} (\tilde{w}^T \tilde{w})$$

$$\tilde{w}^T \Gamma^{-1} \to w^T$$

$$\Omega\left((\Gamma^{-1})^T \tilde{w}\right) = \tilde{w}^T \tilde{w}$$

$$\Omega(w) = w^T \cdot \Gamma \cdot \Gamma^T \cdot w$$

$$= w^T \Gamma^2 w$$

9. (b)

$$\sum_{i=0}^{d} B_i w_i^2 = w^T B w$$

$\boxed{\text{Regularization}}$.

$$\min_{w \in \mathbb{R}^{d+1}} \frac{1}{N} \sum_{n=1}^{N} (w^T x_n - y_n)^2 + \frac{\lambda}{N} w^T B w$$

$$\Rightarrow \frac{1}{N} (Xw - y)^T (Xw - y) + \frac{\lambda}{N} w^T B w$$

$$\frac{\lambda}{N} w^T B w \iff \frac{1}{N} (\tilde{X} w - \tilde{y})^T (\tilde{X} w - \tilde{y})$$

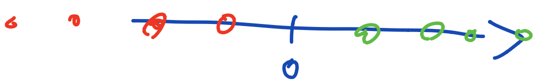
$$\lambda B w^T w \iff (\tilde{X} w - \tilde{y})^T (\tilde{X} w - \tilde{y})$$

10. e

$$E_{loocv} = \frac{1}{N}\sum_{n=1}^{N} e_n = \frac{1}{N}\sum_{n=1}^{N} err(g_n^-(x_n), y_n)$$

$$= \frac{1}{2N}\{N+N\} = 1$$

11. (c)



assume $\frac{N}{2}$ data set <0, $\frac{N}{2}$ data set >0

$$E_{loocv} = \frac{1}{N}\sum_{n=1}^{N} e_n = \frac{1}{N}\sum_{n=1}^{N} err(g_n^-(x_n), y_n)$$

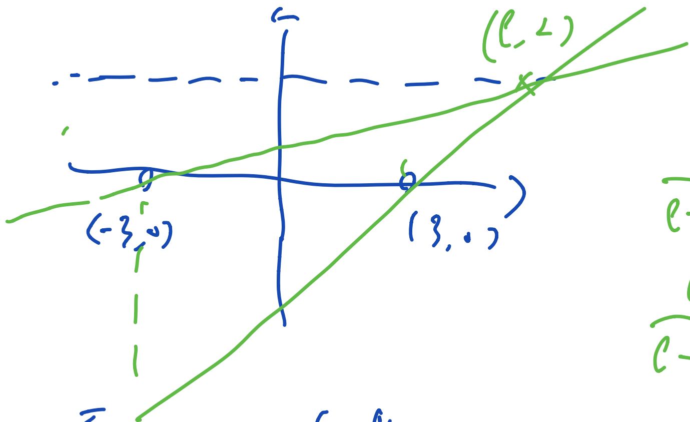
Observation:

Error only happens when the closet point to the origin is removed

$\Rightarrow$ 2 points satisfied

$\Rightarrow E_{loocv} = 2/N$

12. (e)



$$\frac{6}{\rho+3}\cdot 2 = h_1$$

$$\frac{6}{\rho-3}\cdot 2 = h_2$$

$$E_{loocv} = \frac{1}{N}\sum_{n=1}^{N} e_n = \frac{1}{N}\sum_{n=1}^{N} err(g_n^-(x_n), y_n)$$

$$E_{loocv,c} = \frac{1}{3}\left[2^2+1^2+1^2\right] = \frac{1}{3}\cdot 6 = 2$$

$$E_{loocv,lin} = \frac{1}{3}\left\{4+\left(\frac{12}{\rho+3}\right)^2+\left(\frac{12}{\rho-3}\right)^2\right\}$$

$$= \frac{1}{3}\left\{4+\frac{(\rho+3)^2+(\rho-3)^2}{(\rho+3)^2(\rho-3)^2}\cdot 144\right\}$$

$$\Rightarrow \rho = 3\sqrt{9+4\sqrt{6}}$$

13. (d)

For N numbers, variance $\to N\sigma^2$

Since the mean is $1/N$ times the sum

the variance of the sampling distribution of the mean would be $1/N$ times the variance of the sum $\Rightarrow \sigma^2/N$

$$E_{val}(h) = \frac{1}{K}\sum_{i=1}^{K} err(h(x_i)-y)$$

$$V[E_{val}(h)] = V\left[\frac{1}{K}\sum_{i=1}^{K} err(h(x_i)-y)\right]$$

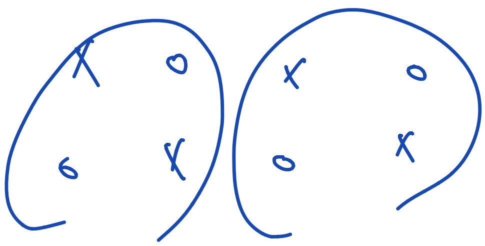
$$= \frac{1}{K^2} V\left[\sum_{i=1}^{K} err(h(x_i)-y)\right]$$

$$= \frac{1}{K^2}\sum_{i=1}^{K} V[err(h(x_i)-y)]$$

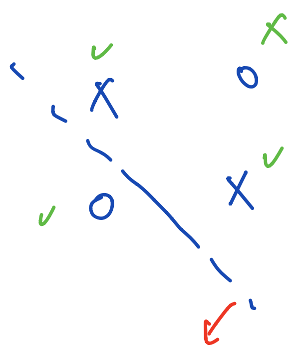
$$= \frac{1}{K} V[err(h(x)-y)]$$

#

14. C



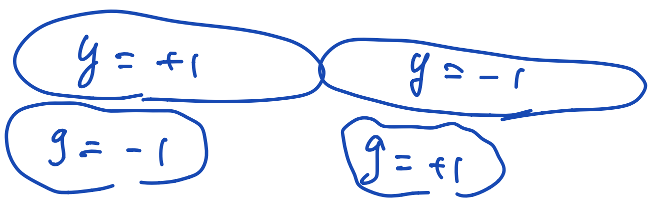
→ 2 cases among 16 can't be fit properly → $2/16$

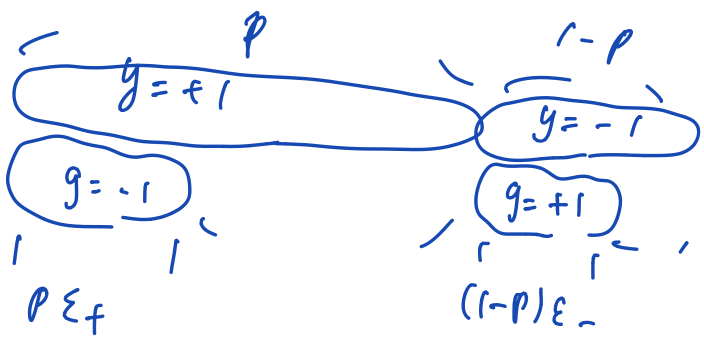


$\Rightarrow \min_{w \in \mathbb{R}^{2+1}} E_{in}(w) = 1/4$

$\Rightarrow \frac{2}{16} \cdot \frac{1}{4} = \frac{2}{64}$

15. (a)





$$E_{out}(g_c) = 1 - p$$

$$E_{out}(g_b) = p\varepsilon_+ + (1-p)\varepsilon_-$$

$$(1-p) = p\varepsilon_+ + (1-p)\varepsilon_-$$

$$1 - p = p\varepsilon_+ + \varepsilon_- - p\varepsilon_-$$

$$p(1 + \varepsilon_+ - \varepsilon_-) = 1 - \varepsilon_-$$

$$p = \frac{1 - \varepsilon_-}{1 + \varepsilon_+ - \varepsilon_-}$$