1. [b]

$$\mathcal{E}_D\left[E_{in}(w_{lin})\right] = \sigma^2\left(1 - \frac{d+1}{N}\right) \geq 0.006$$

$\sigma = 0.1$, $d = 11$

$$0.01\left(1 - \frac{12}{N}\right) \geq 0.006$$

$$\left(1 - \frac{12}{N}\right) \geq 0.6$$

$$\frac{12}{N} \leq 0.4$$

$$N \geq 30$$

② [a]

if we want $X^TXw = X^Ty$

unique solution

$\Rightarrow w = (X^TX)^{-1}X^Ty$ (if X^TX invertible)

$= X^{+}y$ (if X^TX singular)

may have many optimal solution

③ [C]

span of X are linear combination of columns of X

=> if we modify proportion of elements within a column

=> the span of X will be modified

=> H will change!

=> [C]

4. [e]

$$E_{in}(\hat{y}) = \frac{1}{N} \sum_{n=1}^{N} (\hat{y} - y_n)^2$$

$$\nabla E_{in}(\hat{y}) = \left(\frac{d}{d\hat{y}}\right) \frac{1}{N} \sum_{n=1}^{N} (\hat{y}^2 - 2\hat{y} y_n + y_n^2)$$

$$= \frac{1}{N} \sum_{n=1}^{N} (2\hat{y} - 2y_n)$$

$$= 2v$$

$$\Rightarrow -\nabla E_{in}(\hat{y}) = -2v$$

5. [a]

probability of generating one trial $= \frac{1}{\theta}$

For N values, $\left(\frac{1}{\theta}\right)^N$

6. [b]

When all correct, $-yw^Tx = -1$ & err $= 0$

When all wrong, $-yw^Tx = 1$ & err $= 1$

According to the direction, we pick $err(w, x, y)$

$$= \max(0, -yw^Tx)$$

7. [a]

$$err_{exp}(w, x, y) = \exp(-y w^T x)$$

$$-\nabla err_{exp}(w, x, y) = \exp(-y w^T x) \cdot (-y x)$$

$$-\nabla err_{exp}(w, x_n, y_n) = (-y_n x_n) \exp(-y_n w^T x_n)$$

8. [b]

$$\nabla E(w) = b_E(u) + A_E(u)(w-u) = 0$$

$$b_E(u) + A_E(u)\, v = 0$$

$$v = -(A_E(u))^{-1} b_E(u)$$

$(d+1)\cdot n \cdot n \cdot (d+1) \cdot d+1 \cdot 1$

9. [b]

$$E_{in}(w) = \frac{1}{N} \| Xw - y \|^2$$

$$\nabla E_{in}(w) = \frac{2}{N} (X^T X w - X^T y)$$

$$\nabla^2 E_{in}(w) = \frac{2}{N} (X^T X)$$

10. [b]

$$h_y(x) = \frac{\exp(w_y^T x)}{\sum_{i=1}^{k} \exp(w_i^T x)}$$

$$err = -\ln(h_y(x)) = -\sum [y=k] \ln(h_k(x))$$

$$\frac{d\, err(w,x,y)}{d\, w_{ik}} = \frac{-1}{h_k(x)} \frac{\partial h_k(x)}{\partial w_{ik}} \cdot [y=k]$$

$$= \frac{-1}{h_k(x)} \cdot \frac{1}{(\sum \exp(w_i^T x))^2} \Big(x_i \exp(w_k^T x) \sum \exp(w_i^T x) - \exp(w_k^T x) \cdot x_i \exp(w_k^T x) \Big) \cdot [y=k]$$

$$= \frac{-1}{h_k(x)} \frac{x_i \exp(w_k^T x)}{\sum \exp(w_i^T x)} + \frac{1}{h_k(x)} \frac{x_i \exp(w_k^T x)^2}{(\sum \exp(w_i^T x))^2} \cdot [y=k]$$

$$= \frac{-1}{h_k(x)} x_i h_k(x) + \frac{1}{h_k(x)} \frac{x_i \exp(w_k^T x)}{\sum \exp(w_i^T x)} h_k(x) \cdot [y=k]$$

$$= -x_i + x_i \frac{\exp(w_k^T x)}{\sum \exp(w_i^T x)} \cdot [y=k]$$

$$= (-[y=k] + h_k(x))\, x_i$$

11. [e]

$y_n = 2$, $y_n' = 1$

$$h_2(x) = \frac{\exp(w_2^T x)}{\exp(w_1^T x) + \exp(w_2^T x)} = \frac{1}{1 + \exp(w_1^T x - w_2^T x)}$$

$$= \frac{1}{1 + \exp(-(w_2^T - w_1^T) x)}$$

$$= \theta(s)$$

$y_n = 1$ $y_n' = -1$.

$$h_1(x) = \frac{\exp(w_1^T x)}{\exp(w_1^T x) + \exp(w_2^T x)} = \frac{1}{1 + \exp(w_2^T x - w_1^T x)}$$

$$= \frac{1}{1 + \exp(-(w_1^T - w_2^T) x)}$$

$$= 1 - \theta(s)$$

$s = (w_2^T - w_1^T) x$

When (w_1^*, w_2^*) has optimal solution in MLR

$w_2^* - w_1^*$ will have optimal solution in logistic regression

12. {e}

13. [b]

We can divide the problem into d positive & negative ray

$$d(2n) = 2^n$$

$$\log_2(2nd) = n$$

$$1 + \log_2 n + \log_2 d = n$$

$$1 + n/2 + \log_2 d = n$$

$$n = 2(\log_2 d + 1)$$