Subject:
Date

1)

the expected prediction error at data point (x', y') is:

$$E[(h_D(x') - y')^2] = E[h_D(x')^2 - 2h_D(x')y' + y'^2]$$

According to the linearity of expectation $E[X+Y] = E[X] + E[Y]$ and also the fact that $E[ab] = E[a]E[b]$ if "a" and "b" are independent R.v. we have:

$$E[h_D(x')^2 - 2h_D(x')y' + y'^2] = E[h_D(x')^2] - 2E[h_D(x')]E[y'] + E[y'^2]$$

* if Z is a R.v with probability distribution $p(Z)$ and $\underline{Z} = E_p(Z)$ is the average value of Z we have:

$$E[(Z-\underline{Z})^2] = E[Z^2 - 2Z\underline{Z} + \underline{Z}^2] = E[Z^2] - 2E[Z]E[\underline{Z}] + E[\underline{Z}^2]$$

$$= E[Z^2] - 2\underline{Z}^2 + \underline{Z}^2 = E[Z^2] - \underline{Z}^2$$

So we have: $\boxed{E[Z^2] = E[(Z-\underline{Z})^2] + \underline{Z}^2}$ (1)

using (1) in our last equation we have:

$$E[h_D(x')^2] - 2E[h_D(x')]E[y'] + E[y'^2] \xrightarrow[E[y'] = f(x')]{y' = f(x') + \varepsilon}$$

$$= E[(h_D(x') - E[h_D(x')])^2] + E[h_D(x')]^2 - 2E[h_D(x')]f(x')$$

$$+ E[(y' - f(x'))^2] + f(x')^2$$

PAPCO

Subject:
Date

$$= \underbrace{E\left[\left(h_D(x') - E[h_D(x')]\right)^2\right]}_{\Downarrow \text{ variance}} + \underbrace{\left(E[h_D(x') - f(x')]\right)^2}_{\Downarrow \text{ bias}^2} + \underbrace{E\left[\left(y' - f(x')\right)^2\right]}_{\Downarrow \text{ noise}}$$

thus $E\left[\left(h_D(x') - y'\right)^2\right] = \text{variance} + \text{bias}^2 + \text{noise}$

5)

a)

$$f(x) = \underbrace{-5 \log(x^5)}_{g(x)} \underbrace{\sin(x^2)}_{h(x)}$$

$$\frac{\partial f(x)}{\partial x} = \frac{\partial}{\partial x}\left(g(x)\, h(x)\right) = \frac{\partial g(x)}{\partial x} h(x) + g(x) \frac{\partial h(x)}{\partial x}$$

I) $\frac{\partial g(x)}{\partial x} \Rightarrow$

$$g(x) = -5\log(x^5) \Rightarrow \begin{cases} x^5 = T(x) \\ -5\log(x) = S(x) \end{cases} \Rightarrow g(x) = S(T(x))$$

$$\frac{\partial g(x)}{\partial x} = \frac{\partial S(T(x))}{\partial x} = \frac{\partial S}{\partial T} \frac{\partial T}{\partial x} = S'(T(x))\, T'(x)$$

$T'(x) = 5x^4$

$$S'(T(x)) \xrightarrow[S'(x) = \frac{-5}{x \ln 10}]{} S'(T(x)) = \frac{-5}{x^5 \ln 10}$$

$$\boxed{\frac{\partial g(x)}{\partial x} = \frac{-5}{x^5 \ln 10} \times 5x^4} \quad *$$

PAPCO

Subject:
Date

II) $\frac{\partial h(x)}{\partial x} \Rightarrow$

$$h(x) = \sin(x^2) \Rightarrow \begin{cases} x^2 = T(x) \\ \sin(x) = S(x) \end{cases} \qquad h(x) = S(T(x))$$

$$\frac{\partial h(x)}{\partial x} = \frac{\partial S(T(x))}{\partial x} = \frac{\partial S}{\partial T}\frac{\partial T}{\partial x} = S'(T(x))\,T'(x)$$

$T'(x) = 2x$

$$S'(T(x)) \xrightarrow[S'(x) = \cos(x)]{} S'(T(x)) = \cos(x^2)$$

$$\boxed{\frac{\partial h(x)}{\partial x} = \cos(x^2) \times 2x} \quad (**)$$

$$(*)(**) \Longrightarrow \frac{\partial f(x)}{\partial x} = \frac{\partial g(x)}{\partial x} h(x) + g(x)\frac{\partial h(x)}{\partial x}$$

$$= \frac{25}{x \ln 1.} \sin(x^2) + (-5\log(x^5))\cos(x^2) \times 2x$$

b) $f(x) = 3\exp\left(\frac{-5}{3\delta}(x-\mu)^2\right) \qquad \mu, \delta \in R$

$$\Rightarrow \begin{cases} -\frac{5}{3\delta}(x-\mu)^2 = h(x) \\ 3\exp(x) = T(x) \end{cases} \Rightarrow f(x) = T(h(x))$$

$$\frac{\partial f(x)}{\partial x} = \frac{\partial T}{\partial h}\frac{\partial h}{\partial x} = T'(h(x))\,h'(x)$$

$$h'(x) = -\frac{5}{3\delta} \times 2(x-\mu)$$

PAPCO

Subject:
Date

$$f'(h(x)) \xrightarrow[T'(x) = 3\exp(x)]{} T'(h(x)) = 3\exp\left(-\frac{5}{38}(x-\mu)^2\right)$$

$$\Rightarrow \frac{\partial f(x)}{\partial x} = 3\exp\left(\frac{-5}{38}(x-\mu)^2\right) \times \frac{-5}{38} \times 2 \times (x-\mu)$$

c)

$$f_1(x) = \sin(2x_1)\cos(3x_2) \qquad x \in R^2$$

$$f_2(x,y) = 3x^T y \qquad x, y \in R^n$$

$$f_3(x) = -4xx^T \qquad x \in R^n$$

f_1

ii) $$J = \left[\frac{\partial f_1(x)}{\partial x_1} \quad \frac{\partial f_1(x)}{\partial x_2}\right]$$

$$\frac{\partial f_1(x)}{\partial x_1} = 2\cos(2x_1)\cos(3x_2)$$

$$\frac{\partial f_1(x)}{\partial x_2} = -3\sin(2x_1)\sin(3x_2)$$

$$J = \left[2\cos(2x_1)\cos(3x_2) \quad -3\sin(2x_1)\sin(3x_2)\right]$$

i) $$\frac{\partial f_1(x)}{\partial x} \in R^{1\times 2}$$

f_2

ii) $$J = \left[\frac{\partial f_2(x,y)}{\partial x} \quad \frac{\partial f_2(x,y)}{\partial y}\right]$$

PAPCO

Scanned with CamScanner

Subject:
Date

$$\frac{\partial f_2(x,y)}{\partial x} = \left[\frac{\partial f_2(x,y)}{\partial x_1} \cdots \frac{\partial f_2(x,y)}{\partial x_n} \right] = [3y_1, \ldots, 3y_n] = 3y^T \in R^n$$

$$\frac{\partial f_2(x,y)}{\partial y} = \left[\frac{\partial f_2(x,y)}{\partial y_1} \cdots \frac{\partial f_2(x,y)}{\partial y_n} \right] = [3x_1, \ldots, 3x_n] = 3x^T \in R^n$$

$$J = \left[3y^T \quad 3x^T \right]$$

i) $\frac{\partial f_2(x,y)}{\partial x} \in R^n$

f_3

ii) $J = \left[\frac{\partial f_3(x)}{\partial x_1} \cdots \frac{\partial f_3(x)}{\partial x_n} \right]$

$$f_3(x) = -F \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} [x_1 \cdots x_n] = -F \begin{bmatrix} x_1^2 & x_1x_2 & \cdots & x_1x_n \\ x_1x_2 & x_2^2 & & \\ & & & \\ x_1x_n & \cdots & & x_n^2 \end{bmatrix} = Y$$

$$\frac{\partial f_3(x)}{\partial x_i} = \begin{bmatrix} \frac{\partial Y_{11}}{\partial x_i} & \cdots & \frac{\partial Y_{1n}}{\partial x_i} \\ \vdots & & \vdots \\ \frac{\partial Y_{n1}}{\partial x_i} & \cdots & \frac{\partial Y_{nn}}{\partial x_i} \end{bmatrix} = -F(Z_i + Z_i^T) \in R^{n \times n}$$

Z_i is an all zero matix excep vector x in it's ith column

i) $\frac{\partial f_3(x)}{\partial x} \Rightarrow \in R^{(n \times n) \times n}$

Subject: 4)

Date a)

$$z = wx$$

$$\frac{\partial^2 L(w)}{\partial w^2} \Rightarrow \quad L(w) = -y \log \delta(z) - (1-y)\log(1-\delta(z))$$

$$\frac{\partial L(w)}{\partial w} = \frac{-y}{\delta(z)} \underbrace{\frac{\partial \delta(z)}{\partial w}}_{1} - \frac{(1-y)}{1-\delta(z)} \underbrace{\frac{\partial(1-\delta(z))}{\partial w}}_{2}$$

1 $$\frac{\partial \delta(z)}{\partial w} = \frac{\partial \delta(z)}{\partial z}\frac{\partial z}{\partial w} = \delta(z)(1-\delta(z))x$$

2 $$\frac{\partial(1-\delta(z))}{\partial w} = \frac{\partial(1-\delta(z))}{\partial z}\frac{\partial z}{\partial w} = -\delta(z)(1-\delta(z))x$$

$$\frac{\partial L(w)}{\partial w} = \frac{-y}{\delta(z)}\delta(z)(1-\delta(z))x + \frac{(1-y)}{(1-\delta(z))}\delta(z)(1-\delta(z))x$$

$$= -xy(1-\delta(z)) + x(1-y)\delta(z) = \frac{-xye^{-z}}{1+e^{-z}} + \frac{x(1-y)}{1+e^{-z}}$$

$$= \frac{-xye^{-wx} + x - xy}{1+e^{-wx}}$$

$$\frac{\partial^2 L(w)}{\partial w^2} = \frac{+x^2ye^{-wx}(1+e^{-wx}) - (-xe^{-wx})(-xye^{-wx} + x - xy)}{(1+e^{-wx})^2}$$

$$= \frac{\cancel{x^2ye^{-wx}} + \cancel{x^2ye^{-2wx}} - \cancel{x^2ye^{-2wx}} + x^2e^{-wx} - \cancel{x^2ye^{-wx}}}{(1+e^{-wx})^2}$$

PAPCO

Scanned with CamScanner

Subject:
Date

$$\frac{\partial^2 L(w)}{\partial w^2} = \frac{x^2 e^{wx}}{(1+e^{-wx})^2}$$

$x^2 e^{wx} \geq 0$ since $x^2 \geq 0$ and $e^{wx} \geq 0$

$(1+e^{-wx})^2 \geq 0$

$\frac{\partial^2 L(w)}{\partial w^2} \geq 0 \quad \forall w \Rightarrow L(w)$ is Convex

b) $\nabla_w \delta(z) = \nabla_w \delta(wx) = \nabla_w \frac{1}{1+e^{-wx}} = \frac{-xe^{-wx}}{(1+e^{-wx})^2}$

$\nabla_w \delta(wx)$ is $\in R \Rightarrow$ 1-dimensional

c) $\frac{\partial L(w)}{\partial w} = \frac{-xye^{-wx} + x - xy}{1+e^{-wx}} = 0$

$$\frac{x(1-y-ye^{-wx})}{1+e^{-wx}} = 0 \Rightarrow 1-y-ye^{-wx} = 0$$

$$e^{-wx} = \frac{1-y}{y} \Rightarrow -wx = \log\left(\frac{1-y}{y}\right)$$

$$\boxed{w = \frac{\log\left(\frac{y}{1-y}\right)}{x} \quad \forall x \neq 0}$$

PAPCO

Subject:
Date

d) $$w_{t+1} = w_t - \alpha \frac{\delta}{\delta w_t} L(w_t)$$

$$\frac{\partial}{\partial w} L(w) = \frac{-xye^{-wx} + x - xy}{1+e^{-wx}}$$

$$w_1 = w_0 - \alpha \frac{\delta}{\delta w_0} L(w_0)$$

$$= w_0 - \alpha \frac{xye^{-w_0 x} + x - xy}{1+e^{-w_0 x}}$$

Subject:

Date

d)

i) $f(z) = r\exp\left(-\frac{1}{2} z^2\right)$

$(1\times1) \leftarrow z = g(y) = y^T S^{-1} y$ $\qquad x, \mu \in R^D \quad S \in R^{D\times D}$

$(D\times1) \leftarrow y = h(x) = x - \mu$

$$\frac{df}{dx} = \frac{df}{dz}\,\frac{dz}{dy}\,\frac{dy}{dx}$$

1) $\frac{df}{dz} \Rightarrow \begin{cases} r\exp(z) = S(z) \\ -\frac{1}{2} z^2 = T(z) \end{cases}$

$f(z) = S(T(z)) \Rightarrow$

$$\frac{df}{dz} = \frac{dS}{dT}\,\frac{dT}{dz} = S'(T(z))\,T'(z)$$

$T'(z) = -z$

$S'(T(z)) \xrightarrow[S'(z) = r\exp(z)]{} S'(T(z)) = r\exp\left(-\frac{1}{2} z^2\right)$

$\boxed{\frac{df}{dz} = -rz\exp\left(\frac{-z^2}{2}\right)} \rightarrow 1\times1$

2) $\frac{dz}{dy} \Rightarrow$ according to the equation 5.1.7 of mmL book

$$\frac{\partial\, x^T B x}{\partial x} = x^T (B + B^T)$$

$\boxed{\frac{dz}{dy} = y^T \left(S^{-1} + (S^{-1})^T\right)} \rightarrow 1\times D$

3) $\boxed{\frac{dy}{dx} = I_D} \rightarrow D\times D$

PAPCO

Scanned with CamScanner

Subject:
Date

$$\frac{df}{dx} = -2(x-\mu)^T S^{-1}(x-\mu)\exp\left(-\frac{1}{2}\left[(x-\mu)S^{-1}(x-\mu)^T\right]^2\right)$$
$$\times (x-\mu)^T \left(S^{-1} + S^{-1T}\right)$$

ii)

$$f(x) = tr(xx^T + \delta I) \qquad x \in R^D$$

$$X = xx^T \Rightarrow X_{ij} = x_i x_j \qquad X = \begin{bmatrix} x_1^2 & & x_1 x_n \\ & & \\ x_n x_1 & & x_n^2 \end{bmatrix}$$

$$f(x) = x_1^2 + x_1^2 + \dots + x_n^2 + n\delta$$

$$Tr(xx^T + \delta I) = \sum_{i=1}^{D} (X_{ii} + \delta)$$

$$\frac{\partial Tr(xx^T + \delta I)}{\partial x_j} = \sum_{i=1}^{D} \frac{\partial X_{ii}}{\partial x_j} = 2x_j$$

$$\Rightarrow \frac{\partial Tr(xx^T + \delta I)}{\partial x} = 2x^T \Rightarrow 1 \times D$$

iii)

$$f = Tanh^2(z) \in R^M$$

$$z = Ax + b \in R^M \qquad x \in R^N \qquad A \in R^{M\times N} \qquad b \in R^M$$

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial z}\frac{\partial z}{\partial x} \in M \times N$$

$$\frac{\partial f}{\partial z} = diag(1 - tanh^2(z))\, 2\, Tanh(z) \in R^{M\times M}$$

$$\frac{\partial z}{\partial x} = A \in R^{M\times N}$$

$$\frac{df}{dx} = 2\,Tanh(Ax+b)\, diag(1 - tanh^2(Ax+b)) \times A$$

Subject.
Date.

2)

a- yes

solution1: $x' = (\text{floor}(x)) \bmod 2 \Rightarrow \begin{cases} \text{if } x' = 1 & \text{class } o \\ \text{if } x' = 0 & \text{class } x \end{cases}$

solution2: $x' = \begin{cases} x^2 & \text{if } 2k \leq x < 2k+1 \\ -x^2 & \text{if } 2k+1 \leq x < 2k+2 \end{cases}$

$\Rightarrow \begin{cases} \text{if } x' \geq 0 & \text{class } x \\ \text{if } x' < 0 & \text{class } o \end{cases}$

b- yes

$x' = x_1 x_2 \Rightarrow \begin{cases} \text{if } x' < 0 & \text{blue} \\ \text{if } x' \geq 0 & \text{yellow} \end{cases}$

c- yes

solution1: $r = \sqrt{x_1^2 + x_2^2}$

$r' = (\text{Floor}(r)) \bmod 2 \Rightarrow \begin{cases} \text{if } r' = 1 & \text{class } + \\ \text{if } r' = 0 & \text{class } - \end{cases}$

$\varphi(x) = (r, r')$

$K(x, x') = \varphi(x)^T \varphi(x') =$

$\left(\sqrt{(x_1^2 + x_2^2)(x_1'^2 + x_2'^2)}, [(\text{Floor}(\sqrt{x_1^2 + x_2^2})) \bmod 2] . [(\text{Floor}(\sqrt{x_1'^2 + x_2'^2})) \bmod 2] \right)$ ✓

solution 2: $r = \sqrt{x_1^2 + x_2^2}$

$r' = \begin{cases} r^2 & \text{if } 2k \leq x < 2k+1 \\ -r^2 & \text{if } 2k+1 \leq x < 2k+2 \end{cases}$

$\Rightarrow \text{if} \begin{cases} r' < 0 & \text{class } + \\ r' \geq 0 & \text{class } - \end{cases}$

Subject.

Date.

Q2-cont.:

kernel:

$k(x,x') = \phi(x)^T \phi(x') =$

$\left(\sqrt{(x_1^2+x_2^2)(x_1'^2+x_2'^2)}\ ,\ (x_1^2+x_2^2)(x_1'^2+x_2'^2)\right)$ ✓

3.)

a) ~~risk~~ $risk = \sum_D (y_i - h(x_i))^2$

b) $\underset{D\sim P}{E}[error\_kfold] = E\left[\frac{1}{K}\sum_{i=1}^{K}\frac{1}{n/K}\sum_{j\in ind[i]} L(h_{D\setminus i}(x_j), y_j)\right]$

$= \frac{1}{K}\sum_{i=1}^{K}\frac{1}{n/K}\sum_{j\in ind[i]}\underset{D\sim P}{E}\left[L(h_{D\setminus i}(x_j), y_j)\right]$

$= \frac{1}{K}\sum_{i=1}^{K}\frac{1}{n/K}\sum_{j\in ind[i]}\underset{\substack{D'\sim P,\\(x,y)\sim P}}{E}\left[L(h_{D'}(x), y)\right]$

$= \frac{1}{K}\sum_{i=1}^{K}\cancel{\frac{K}{n}}\times\cancel{\frac{n}{K}}\underset{\substack{D'\sim P,\\(x,y)\sim P}}{E}\left[L(h_{D'}(x), y)\right]$

$= \frac{1}{K}\sum_{i=1}^{K}\underset{\substack{D'\sim P,\\(x,y)\sim P}}{E}\left[L(h_{D'}(x), y)\right] = \underset{\substack{D'\sim P,\\(x,y)\sim P}}{E}\left[(y-h_{D'}(x))^2\right]$

c.)

linear regression : $\theta^* = (X^TX)^{-1}X^Ty$

→ complexity of X^TX ((d×n) matrix × (n×d) matrix) $= O(d^2n)$

→ complexity of inverting X^TX (inverting (d×d) matrix) $= O(d^3)$

Subject.

Date.

Q3- cont.

→ complexity of $X^T y$ ($(d \times n)$ matrix $\times$ $(n \times 1)$ matrix) $= O(dn)$

→ complexity of AB where $A = (X^T X)^{-1}$, a $(d \times d)$ matrix and $B = X^T y$, a $(d \times 1)$ matrix $= O(d^2)$

⇒ Total time complexity $= O(d^2 n + d^3 + dn + d^2) = O(d^2 n + d^3)$ ✓

d)

$$\text{error\_kfold\_reg} = \frac{1}{K} \sum_{i=1}^{K} \frac{1}{n/K} \sum_{j \in ind[i]} (y_j - x_j \theta_{-i}^*)^2$$

$$= \frac{1}{K} \sum_{i=1}^{K} \frac{1}{n/K} \sum_{j \in ind[i]} \left(y_j - x_j \left[(X_{-i}^T X_{-i})^{-1} X_{-i}^T Y_{-i}\right]\right)^2$$ ✓

complexity:

$\Rightarrow$ complexity of θ_{-i}^* based on Previous section is:

$O(d^2(n - \frac{n}{K}) + d^3 + d(n - \frac{n}{K}) + d^2) = O(d^2 n + d^3)$

$\Rightarrow$ complexity of $x_j \theta_{-i}^*$ ((1xd) matrix x (dx1) matrix)

$= O(d)$

$\Rightarrow$ complexity of $\sum_{j \in ind[i]}$ is $O(\frac{n}{K})$

$\Rightarrow$ complexity of $\sum_{i=1}^{K}$ is $O(K)$

$\Rightarrow$ total time complexity $= O(K \cdot \frac{n}{K} (d^2 n + d^3 + d))$

$= O(d^2 n^2 + d^3 n + dn) = \boxed{O(d^2 n^2 + d^3 n)}$ ✓

e)

we should prove that for $k=n$ we have:

$$\frac{1}{K}\sum_{i=1}^{K}\left\| \frac{y_i - X_i w^*}{I - X_i (X^T X)^{-1} X_i^T} \right\|^2 = \frac{1}{K}\sum_{i=1}^{K} \cancel{\frac{1}{n/K}} \sum_{j\in ind[i]} (y_j - x_j[(X_{-i}^T X_{-i})^{-1} X_{-i}^T Y_{-i}])^2$$

$\xrightarrow{n=k} \sum_{j\in ind[i]}$ will loop over all the data we have (each fold will have 1 datapoint $\Rightarrow$ len(ind[i]=1)). $\sum_{j\in ind[i]}$ turns our "j" to "i". So we have our second term above as:

$$(y_i - x_i[(X_{-i}^T X_{-i})^{-1} X_{-i}^T Y_{-i}])^2$$

$$\frac{1}{K}\sum_{i=1}^{k}\left\| \frac{y_i - X_i w^*}{I - X_i (X^T X)^{-1} X_i^T} \right\|^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - X_i w^*}{I - X_i (X^T X)^{-1} X_i^T} \right)^2$$

now we should prove:

$$\frac{y_i - X_i w^*}{I - X_i (X^T X)^{-1} X_i^T} = y_i - x_i (X_{-i}^T X_{-i})^{-1} X_{-i}^T y_{-i}$$

$$\Rightarrow (y_i - x_i(X_{-i}^T X_{-i})^{-1} X_{-i}^T y_{-i})(\underbrace{I}_{=1} - \overbrace{X_i (\underbrace{X^T X}_{d\times d})^{-1} X_i^T}^{(1\times d)\times(d\times d)\times(d\times 1) = 1\times 1 \Rightarrow I = 1})$$

$$= y_i - y_i X_i (X^T X)^{-1} X_i^T - x_i (X_{-i}^T X_{-i})^{-1} X_{-i}^T y_{-i} +$$
$$(x_i (X_{-i}^T X_{-i})^{-1} X_{-i}^T y_{-i})(X_i (X^T X)^{-1} X_i^T)$$

we know that $X^T X = X_{-i}^T X_{-i} + x_i x_i^T \Rightarrow$ so we can rewrite the last term of the equation above as:

3e cont.

$x_i(X_{-i}^T X_{-i})^{-1} X_{-i}^T y_{-i}\ x_i(X^T X)^{-1} x_i^T =$

$x_i(\underbrace{X^T X - x_i x_i^T}) ^{-1} X_{-i}^T y_{-i}\ x_i(X^T X)^{-1} x_i^T$ (*)

using $(A+UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1}+VA^{-1}U)^{-1}VA^{-1}$

where $C=I$ we get: $(A+UV)^{-1} = A^{-1} - A^{-1}U(I+VA^{-1}U)^{-1}VA^{-1}$

so $(\underbrace{X^T X}_{A} - \underbrace{x_i}_{V} \underbrace{x_i^T}_{U})^{-1} = (X^T X)^{-1} - (X^T X)^{-1} x_i^T (I -$

$x_i (X^T X)^{-1} x_i^T)^{-1} (-x_i)(X^T X)^{-1}$

so * is: $x_{-i}(X^T X)^{-1} - x_{-i}(X^T X)^{-1} x_i^T (I - x_i (X^T X)^{-1} x_i^T)^{-1}$

$(-x_i)(X^T X)^{-1} = x_{-i}(X^T X)^{-1} - x_{-i}(X^T X)^{-1} x_i^T (-x_i)$

$(X^T X)^{-1} (I - x_i^T x_i (X^T X)^{-1})^{-1} =$

$x_{-i}(X^T X)^{-1} (I + x_i^T (X^T X)^{-1})(I - x_i^T x_i (X^T X)^{-1})^{-1}$

complexity: for $x_i w^*$ using the complexity of previous sections we have: $O(\frac{nd}{k})$

w^* has complexity of $O(d^3 + d^2 n)$

for $x_i \underbrace{(X^T X)^{-1}}_{O(d^3)} x_i^T$ we have: $O(d^3 + nd^2 + nd^2)$

$= O(d^3 + nd^2)$

for $I = x_i (X^T X)^{-1} x_i$ we have 2 matrixes of size $(\frac{n}{k}, \frac{n}{k})$

$\Rightarrow O(\frac{n^2}{k^2})$

and dividing $y_i - x_i w^*$ by the previous term makes complexity of $O(\frac{n^3}{k^3})$

we also have $\sum_{i=1}^{n}$ so we have $O(n)$

$\Rightarrow$ Total complexity $= O(n \cdot (d^3 + nd^2 + \frac{n^3}{k^3})) =$

$O(nd^3 + n^2 d^2 + \frac{n^4}{k^3})$