1) 1. $x^T X Y = \begin{bmatrix} 2 & 2 & 6 \end{bmatrix} \begin{bmatrix} -4 & 7 & 5 \\ 1 & -9 & 2 \\ 8 & 3 & 6 \end{bmatrix} \begin{bmatrix} 2 & 4 & 6 \\ 1 & 0 & 13 \\ -4 & 5 & 5 \end{bmatrix} = \begin{bmatrix} 42 & 14 & 50 \end{bmatrix} \begin{bmatrix} 2 & 4 & 6 \\ 1 & 0 & 13 \\ -4 & 5 & 5 \end{bmatrix}$

$= \begin{bmatrix} -102 & 418 & 684 \end{bmatrix}$

2. $YX = \begin{bmatrix} 2 & 4 & 6 \\ 1 & 0 & 13 \\ -4 & 5 & 5 \end{bmatrix} \begin{bmatrix} -4 & 7 & 5 \\ 1 & -9 & 2 \\ 8 & 3 & 6 \end{bmatrix} = \begin{bmatrix} 44 & -4 & 54 \\ 100 & 46 & 83 \\ 61 & -58 & 20 \end{bmatrix}$

3. $2X + 2Y + X^T Y = \begin{bmatrix} -8 & 14 & 10 \\ 2 & -18 & 4 \\ 16 & 6 & 12 \end{bmatrix} + \begin{bmatrix} 4 & 8 & 12 \\ 2 & 0 & 26 \\ -8 & 10 & 10 \end{bmatrix} + \begin{bmatrix} -4 & 1 & 8 \\ 7 & -9 & 3 \\ 5 & 2 & 6 \end{bmatrix} \begin{bmatrix} 2 & 4 & 6 \\ 1 & 0 & 13 \\ -4 & 5 & 5 \end{bmatrix}$

$= \begin{bmatrix} -4 & 22 & 22 \\ 4 & -18 & 30 \\ 8 & 16 & 22 \end{bmatrix} + \begin{bmatrix} -39 & 24 & 29 \\ -7 & 43 & -60 \\ -12 & 50 & 86 \end{bmatrix} = \begin{bmatrix} -43 & 46 & 51 \\ -3 & 25 & -30 \\ -4 & 66 & 108 \end{bmatrix}$

4. $YZ^T = \begin{bmatrix} 2 & 4 & 6 \\ 1 & 0 & 13 \\ -4 & 5 & 5 \end{bmatrix} \begin{bmatrix} 4 & 1 \\ -1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 12 \\ 4 & 14 \\ -21 & 6 \end{bmatrix}$

5. $Z^T Z x = \begin{bmatrix} 4 & 1 \\ -1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 4 & -1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \\ 6 \end{bmatrix} = \begin{bmatrix} 17 & -3 & 1 \\ -3 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \\ 6 \end{bmatrix} = \begin{bmatrix} 34 \\ 4 \\ 10 \end{bmatrix}$

2) see ipynb file

3) 1. $f(x) = x^T A x - x^T y = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \end{bmatrix}$

$= \begin{bmatrix} x_1 & 2x_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - 2x_1 = x_1^2 + 2x_2^2 - 2x_1$

$f'(x) = \begin{bmatrix} 2x_1 - 2 \\ 4x_2 \end{bmatrix} \Rightarrow f'\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 4 \end{bmatrix}$

2. $f(x) = x^T A \mathbf{1} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} x_1 & 2x_2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = x_1 + 2x_2$

$f'(x) = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \Rightarrow f'\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

3. $f(x) = x^T \mathbb{1} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = x_1 + x_2$

$$f'(x) = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \Rightarrow f'\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

4. $f(w) = \sum_i^2 w^T x_i = \begin{bmatrix} w_1 & w_2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} w_1 & w_2 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = w_1 + 2w_1 + w_2$

$$= 3w_1 + w_2$$

$$f'(w) = \begin{bmatrix} 3 \\ 1 \end{bmatrix} \Rightarrow f'\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

4) linear: $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ $y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$ $f(x) = y^T x = \begin{bmatrix} y_1 & y_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1 y_1 + x_2 y_2$

$$f'(x) = \begin{bmatrix} \frac{df}{dx_1}(x) \\ \frac{df}{dx_2}(x) \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = y$$

quadratic: $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$

$$f(x) = x^T A x = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$= a_{11} x_1^2 + a_{12} x_1 x_2 + a_{21} x_1 x_2 + a_{22} x_2^2$$

$$f'(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2a_{11} x_1 + (a_{12} + a_{21}) x_2 \\ (a_{12} + a_{21}) x_1 + 2a_{22} x_2 \end{bmatrix}$$

$$= \begin{bmatrix} 2a_{11} & (a_{12} + a_{21}) \\ (a_{12} + a_{21}) & 2a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$= \left(\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix} \right) \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = (A + A^T) x$$

trace of quadratic: $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ $f(x) = Tr(x^T A x)$

$$f'(x) = \frac{d}{dx} Tr(x^T A x) = \frac{d}{dx} x^T A x \text{ (prop. 1 from lec. 2)}$$

regression objective: $X \in \mathbb{R}^{N \times d}$ $Y \in \mathbb{R}^N$ $f(w) = \frac{1}{2} \sum_i^N (w^T x_i - y_i)^2$

let $q_i = w^T - x_i \Rightarrow f(w) = \frac{1}{2} \sum_i^N (q_i - y_i)^2$

$$f'(w) = \frac{df}{dq} \cdot \frac{dq}{dw} = \sum_i^N (q_i - y_i) \cdot x_i = \sum_i^N (w^T x_i - y_i) x_i$$

relu: $f(w) = \mathrm{ReLU}(w^T x)$ let $q = w^T x \Rightarrow f'(w) = \frac{df}{dq}\frac{dq}{dw}$

$$f'(w) = 1_{(w^T x > 0)}\, x$$

multivariate Gaussian: $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ $f(x) = e^{-x^T A x}$

$$f'(x) = -e^{-x^T A x}\left[(A + A^T)\right] x \quad \text{(chain rule)}$$

Sigmoid: $x \in \mathbb{R}$ $\sigma(x) = \frac{1}{1+e^{-x}} = (1+e^{-x})^{-1}$

$$\sigma'(x) = -(1+e^{-x})^{-2}(e^{-x})(-1) = (1+e^{-x})^{-2}(e^{x})$$

$$= \frac{e^{-x}}{(1+e^{-x})(1+e^{-x})} = \frac{1}{1+e^{-x}} \frac{e^{-x}}{1+e^{-x}}$$

$$= \frac{1}{1+e^{-x}}\left(\frac{1+e^{-x}}{1+e^{-x}} - \frac{1}{1+e^{-x}}\right) = \frac{1}{1+e^{-x}}\left(1 - \frac{1}{1+e^{-x}}\right) = \sigma(x)(1-\sigma(x))$$

L1 norm: $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ $f(x) = \|x\|_1 = |x_1| + |x_2|$

$$f'(x) = \begin{bmatrix} 1_{(x_1 > 0)} - 1_{(x_1 \leq 0)} \\ 1_{(x_2 > 0)} - 1_{(x_2 \leq 0)} \end{bmatrix} = \mathrm{Sign}(x)$$

L2 norm: $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ $f(x) = \|x\|_2 = (x_1^2 + x_2^2)^{\frac{1}{2}}$

$$f'(x) = \frac{1}{2}(x_1^2 + x_2^2)^{-\frac{1}{2}} \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix} = \frac{1}{(x_1^2+x_2^2)^{1/2}} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{x}{\|x\|_2}$$

PCA: $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ $f(x) = x^T A x - \lambda(x^T I x - 1)$

$$f'(x) = (A + A^T)x - 2\lambda x \quad \text{(using quadratic)}$$

Gaussian MLE $\mu \in \mathbb{R}$ $f(\mu) = \log\left(\prod_i^N \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}\right)$

$$= \sum_i^N \log\left(\frac{1}{\sigma\sqrt{2\pi}}\right) + \sum_i^N \log\left(e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}\right)$$

$$= N\log\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \sum_i^N \frac{(x_i-\mu)^2}{2\sigma^2}$$

$$f'(\mu) = -\left(\frac{1}{2\sigma^2}\sum_i^N 2(x_i-\mu)\cdot(-1)\right) = \frac{1}{\sigma^2}\sum_i^N (x_i - \mu)$$

multivariate Gaussian Tr: $f(x) = e^{-Tr(x^TAx)}$

$$f'(x) = e^{-Tr(x^TAx)}((A+A^T)x)(-1) = -e^{-Tr(x^TAx)}(A+A^T)x$$

Log. Reg. objective: $f(w) = \frac{1}{1+e^{-w^Tx}}$ let $z = w^Tx$

$$f'(w) = \frac{df}{dz}\frac{dz}{dw} = \sigma(z)(1-\sigma(z))x$$

$$= \frac{1}{1+e^{-w^Tx}}\left(1-\frac{1}{1+e^{-w^Tx}}\right)x$$

Uniform MLE: $a, b \in \mathbb{R}$, $a \le b$ $f(a,b) = \log\left(\prod_i^n \frac{1}{b-a}\right)$

$$= \sum_i^n \log\left(\frac{1}{b-a}\right)$$

$$= \sum_i^n \log((b-a)^{-1})$$

$$= -n\log(b-a)$$

$$f'(a,b) = \begin{bmatrix} \frac{n}{b-a} \\ -\frac{n}{b-a} \end{bmatrix}$$

Exponential MLE: $\lambda \in \mathbb{R}$ $f(\lambda) = \log(\prod_i^n \lambda e^{-\lambda x_i})$

$$= \sum_i^n \log(\lambda e^{-\lambda x_i})$$

$$= \sum_i^n \log(\lambda) + \log(e^{-\lambda x_i})$$

$$= \sum_i^n \log(\lambda) - \lambda x_i = n\log(\lambda) - \sum_i^n \lambda x_i$$

$$f'(\lambda) = \frac{n}{\lambda} - \sum_i^n x_i$$

Bernoulli MLE: $\alpha \in \{0,1\}$ $p \in \mathbb{R}$

$$f(p) = \log\left(\prod_i^n p^{\alpha_i}(1-p)^{1-\alpha_i}\right) = \sum_i^n \log(p^{\alpha_i}(1-p)^{1-\alpha_i})$$

$$= \sum_i^n \log(p^{\alpha_i}) + \log((1-p)^{1-\alpha_i}) = \log p \sum_i^n \alpha_i + \log(1-p)\sum_i^n (1-\alpha_i)$$

$$f'(p) = \frac{1}{p}\left(\sum_i^n \alpha_i\right) - \frac{1}{1-p}\left(\sum_i^n 1-\alpha_i\right)$$

SVM objective: $f(w) = \sum_i^n \mathrm{ReLU}(1-y_i\langle x_i, w\rangle)$

$$= \sum_i^n \mathrm{ReLU}(1-y_i(x_i^Tw))$$

$$f'(w) = \sum_i^n 1_{(1-y_i(x_i^Tw)>0)}(-y_i x_i^T)$$