HW6 p1

#1 a.) $f(x) = \|x\|$ is convex:

$$f((1-t)x + ty) = \|(1-t)x + ty\| \le (1-t)\|x\| + t\|y\| = (1-t)f(x) + t f(y)$$

triangle and $1-t \ge 0$ and $t \ge 0$ $\quad \Rightarrow$ convex ✓

b.) first note that the sum of two convex function is convex. Also, for $t > 0$ $\; t \cdot f(x)$ is convex if f is convex.

Since $w \mapsto \|Xw - y\|_2^2$ is convex and $\|w\|$ is convex it follows that $f(w)$ is convex.

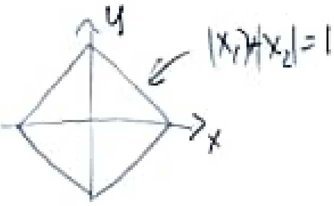
c.) $x = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ $\quad y = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ $\quad x + y = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

$$\left.\begin{array}{l} \|x+y\|_p = 2^{1/p} \\ \|x\|_p = \|y\|_p = 1 \end{array}\right\} \quad \|x\|_p + \|y\|_p = 2 \underset{p<1}{<} 2^{1/p} = \|x+y\|_p$$

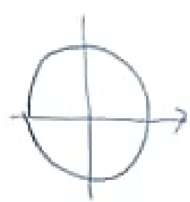
triangle inequality is invalid when $p < 1$

d.)

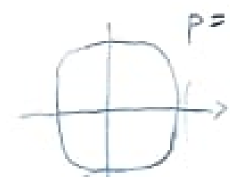
$p = 1$



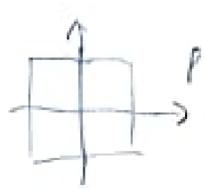
$p = 2$



$p = 4$

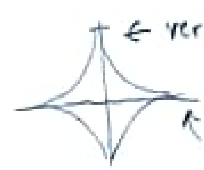


gets flatter $\Rightarrow$



$p = \infty$

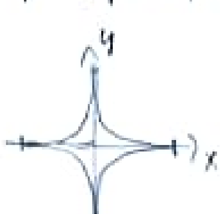
e.) $p = 1/2 \quad x^{1/2} + y^{1/2} = 1 \Rightarrow y = (1 - x^{1/2})^2$



vertical tangent lines

horiz tangent lines

$p = 1/4$



$\Rightarrow p = 0$

$\Leftrightarrow \|x\|_0 = \#$ nonzero elmts in a vector

HW 6 p2

#2| $p(Y=y_k \mid X=x) = \frac{1}{\sqrt{2\pi}\,\sigma_k} \cdot \exp\left(-\frac{(y_k - w_k^T x)^2}{2\sigma_k^2}\right)$

Data

$x_n \in \mathbb{R}^D \Rightarrow X_{nd}$

$y_n \in \mathbb{R}^K \Rightarrow y_{nk}$

parameters

$\theta = w_{kd}$

$$\Rightarrow \mathcal{L}(\theta) = \prod_k \prod_n \frac{1}{\sqrt{2\pi}\,\sigma_k} \cdot \exp\left(-\frac{(y_{kn} - w_k^T x_n)^2}{2\sigma_k^2}\right)$$

$$NLL(\theta) = -\sum_{k,n} \frac{1}{\sqrt{2\pi}\,\sigma_k} + \sum_{k,n} \frac{(y_{kn} - w_k^T x_n)^2}{2\sigma_k^2}$$

thus
$$\min \sum_k \frac{1}{\sigma_k^2} \sum_n (x_n^T w_k - y_{kn})^2$$
$$= \min \sum_k \frac{1}{\sigma_k^2} \| X \underline{w}_k - \underline{y}_k \|^2 \qquad \text{where } X = [x_{nd}] \quad \underline{y}_k = \begin{bmatrix} y_{1k} \\ \vdots \\ y_{nk} \end{bmatrix}$$

When we form the gradient wrt w_k only one term in the sum remains

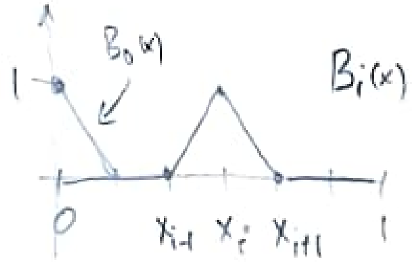
thus the normal equations are

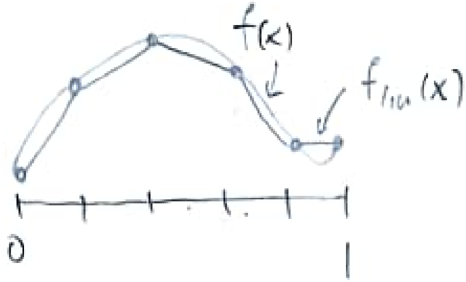
$$X^T X \underline{w}_k = X^T \underline{y}_k \qquad k = 1..K$$

i.e, we can consider each k-component separately.

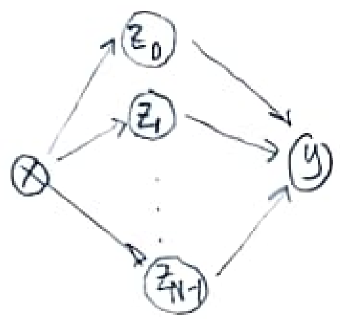
#3| "HW 6 p3"

a.)





b.)



$$z_k = ReLU(x - b_k)$$

$$y = \sum_j w_j z_j + c$$

we want that $y(x_k) = f(x_k)$ for $k = 0..N$

mind that $ReLU(x_k - x_j) = \begin{cases} x_k - x_j & \text{if } k > j \\ 0 & \text{if } k \leq j \end{cases}$

choose $b_k = x_k$ $k = 1..N$ $\Rightarrow$ $z_k = ReLU(x - x_k)$

1.) plug in $x = x_0$: $y(x_0) = \sum_j w_j \underbrace{ReLU(x_0 - x_j)}_{=0} + c \Rightarrow c = f(x_0)$

2.) plug in $x = x_1$: $y(x_1) = w_0 \cdot (x_1 - x_0) + c = f(x_1) \Rightarrow w_0 = - \frac{f(x_0)}{x_1 - x_0}$

3.) plug in $x = x_2$ $y(x_2) = w_0(x_2 - x_0) + w_1(x_1 - x_0) + c = f(x_2)$

$$\Rightarrow w_1 = \frac{1}{x_1 - x_0}\left[f(x_2) - w_0(x_2 - x_0) - c \right]$$

and so on.

4.) plug in $x = x_N$ $y(x_N) = \underbrace{w_0(x_N - x_0) + \ldots + w_{N-2}(x_N - x_{N-2})}_{\text{all these have been computed before}} + \overset{\text{unknown}}{w_{N-1}}(x_N - x_{N-1}) + c$

$$\Rightarrow w_{N-1} = \frac{1}{x_N - x_{N-1}}\left(f(x_N) - c - w_0(x_N - x_0) - \ldots - w_{N-1}(x_N - x_{N-1}) \right)$$

Done.

#4 $\quad X \xrightarrow{f_1} z_1 \xrightarrow{f_2} z_1 \quad \xrightarrow{f_L} z_L = y$ (HW6 p4)

$X \in \mathbb{R}^{m_0}$, $z_1 \in \mathbb{R}^{m_1}$, $z_L \in \mathbb{R}^{m_L}$

$m_0 = n$
$m_L = m$

a.) $V_2 = J_2 \cdot J_1$

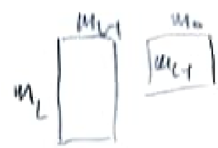
$\Rightarrow$ cost $= 2 m_0 m_1 m_2$

$V_3 = J_3 V_2$

$\Rightarrow$ cost $= 2 m_0 m_2 m_3$

$\vdots$

$V_L = J_L V_{L-1}$



$\Rightarrow$ cost $= 2 m_0 m_{L-1} m_L$

total cost : $2 \underset{= m_0}{n} \cdot (m_1 m_2 + m_2 m_3 + \ldots + m_{L-1} m_L)$

b.) $V_{L-1} = J_L J_{L-1}$

$\Rightarrow 2 \cdot m_{L-2} m_{L-1} \cdot m_L$

$V_{L-2} = V_{L-1} J_{L-2}$

$\Rightarrow 2 \cdot m_{L-3} m_{L-2} m_L$

$V_1 = V_2 J_1$

$\Rightarrow 2 \cdot m_0 m_1 \cdot m_L$

total cost $\quad 2 \underset{= m_L}{m} (m_0 m_1 + m_1 m_2 + \ldots + m_{L-2} \cdot m_{L-1})$

c.) $\underline{v}_2 = J_2 \underline{v}_1$ cost $2 m_2 m_1$

$\underline{v}_3 = J_3 v_2$ cost $2 m_3 m_2$

$\vdots$

$\underline{v}_L = J_L \cdot v_{L-1}$ cost $2 m_L \cdot m_{L-1}$

sum $2(m_1 m_2 + \ldots + m_{L-1} m_L)$

do this for $\underline{v}_1$ = all columns of J_1

$\Rightarrow$ cost $= 2n \cdot (m_1 m_2 + \ldots + m_{L-1} m_L)$

d.) $u_{L-1}^T = u_L^T \cdot J_{L-1}$ cost $2 m_{L-1} m_{L-2}$

$u_{L-2}^T = u_{L-1}^T J_{L-2}$ cost $2 m_{L-2} m_{L-1}$

$\vdots$

$u_1^T = u_2^T J_1$ cost $2 m_1 m_0$

sum: $2(m_0 m_1 + \ldots + m_{L-2} m_{L-1})$

do this for u_L^T = all rows of J_L

$\Rightarrow$ cost $= 2m(m_0 m_1 + \ldots + m_{L-2} m_{L-1})$

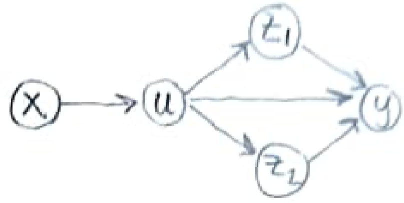
Conclusion: cost of a = cost of c
cost of b = cost of d

cost of a and b can be diffrent, depending on all dimensions.

HW6 p5

#5

$$u = \varphi(ax+b)$$
$$z_1 = \varphi(cu+d)$$
$$z_2 = \varphi(eu+f)$$
$$y = pu + qz_1 + rz_2$$



$$\frac{\partial y}{\partial a} = p\frac{\partial u}{\partial a} + q\frac{\partial z_1}{\partial a} + r\frac{\partial z_2}{\partial a}$$

$$\frac{\partial z_1}{\partial a} = \varphi'(cu+d)\cdot c\frac{\partial u}{\partial a}$$

$$\frac{\partial z_2}{\partial a} = \varphi'(eu+f)\cdot e\frac{\partial u}{\partial a}$$

$$\frac{\partial u}{\partial a} = \varphi'(ax+b)\cdot x$$

combine

$$\frac{\partial y}{\partial a} = \left(p + q\,\varphi'(cu+d)\cdot c + r\cdot\varphi'(eu+f)\cdot e\right)\cdot\varphi'(ax+b)\cdot x$$