HW1

N1

$g(x)$ is convex if

$$f(\lambda x+(1-\lambda)y) \le \lambda f(x)+(1-\lambda)f(y) \quad \forall x,y \in f$$

$f(x)$ is convex then $g(x)=f(Ax+b)$ is convex.

Proof:

$$g(\lambda x+(1-\lambda)y) = f(A(\lambda x+(1-\lambda)y)+b)$$

$$= f(\lambda Ax+(1-\lambda)Ay+b) =$$

$$f(\lambda Ax+b\lambda-b\lambda+(1-\lambda)Ay+b) =$$

$$= f(\lambda(Ax+b)+Ay-\lambda Ay+b-b\lambda) =$$

$$= f(\lambda(Ax+b)+(1-\lambda)(Ay+b))$$

$$\le \lambda f(Ax+b)+(1-\lambda)f(Ay+b)$$ by convexity of f.

$\blacksquare$

b) $\cdot$ $f_1(x), \dots f_n(x)$ are conv.

$h = \max\{f_1(x), \dots f_n(x)\}$

Proof: $h(\lambda x + (1-\lambda)y) =$

$\max\{f_1(\lambda x + (1-\lambda)y), \dots f(\lambda x, (1-\lambda)y\} \leq$

$\leq \max\{\lambda f_1(x) + (1-y) f_1(y), \dots \lambda f_n(x) + (1-\lambda) f_n(y)\}$

$\leq \max\{\lambda f_1(x), \dots \lambda f_n(x)\} + \max\{(1-\lambda) f_1(y), \dots (1-\lambda) f_n(y)\}.$

$= \lambda \max\{f_1(x) \dots f_n(x)\} + (1-\lambda) \max\{f_1(y), \dots f_n(y)\}$

$= \lambda h + (1-\lambda) h$ ■

c) $\lambda > 0$, $v, u \in \arg\min f(w)$

$f(x)$ is λ-strong convex if

$\forall x, y \quad f(y) \leq f(x) + \nabla f(x)^T (y-x) + \frac{\lambda}{2} \|y-x\|^2$

Let $u \neq v$. Then, there will be two points

$x_1 \neq x_2$ where $\nabla f(x_1) = f(x_2) = 0$ (1)

Then, $f(x_1) \leq f(x_2) + \nabla f(x_2)^T (x_1 - x_2) + \frac{\lambda}{2} \|x_1 - x_2\|^2$

which is equivalent to

$\langle \nabla f(x_1) - \nabla f(x_2), x - y \rangle \geq \lambda \|x_1 - x_2\|^2$

Since $\nabla f(x_1) - \nabla f(x_2) = 0$, LHS $= 0$

Also, $\lambda > 0$ and $\|x_1 - x_2\|_2^2 > 0$ by (1)

Therefore, $0 \geq \lambda \|x_1 - x_2\| > 0$

which is a contradiction $\blacksquare$

d) $\Omega = [0, 1]$ $f(x) = x$ The function is differentiable everywhere except at $\{0, 1\}$.

Then, ∂f becomes $(-\infty, 1], x = 0, \{1\}, x \in (0, 1)$

So, as we see, by any means, $0 \in \partial f(x^*)$, where x^* is obviously 0.

N2.

$\frac{1}{p} + \frac{1}{q} = 1$ $\quad \|z\|_c = \left(\sum |z_i|^c\right)^{\frac{1}{c}}$, $1 < p < 2$

$2 \leq q < \infty$

a) $a, b \geq 0$, $ab \leq \frac{a^p}{p} + \frac{b^q}{q}$

Let $\frac{1}{p} = t$ $\frac{1}{q} = 1 - t$

$ab \leq a^{\frac{1}{t}} t + b^{\frac{1}{(1-t)}} (1-t)$ Lets keep p and q in expon.

ln is concave

$\ln a + \ln b \leq \ln(a^p t + (1-t) b^q)$

$\geq \ln(a^p) t + (1-t) \ln p = \ln a \cancel{\frac{1}{t}} \cancel{t} + \ln b \cancel{\frac{1-t}{1-t}}$

$\ln(a^p) t + (1-t) \ln(b^q) \geq \ln a + \ln b = \ln ab$

The inequality strictly holds if $a^p = b^q$

b) $$\langle x,y\rangle \leq \frac{\|x\|_p^p}{p} + \frac{\|y\|_p^p}{q} = \frac{\sum^d |x_i|^p}{p} + \frac{\sum^d |y_i|^q}{q}$$

$$= \sum_{i=1}^{d}\left(\frac{|x_i|^p}{p} + \frac{|y_i|^p}{q}\right) \geq \sum_{i=1}^{d} x_i y_i$$ which

holds because of Young's inequality. = from a) directly

c) $\langle x, y \rangle \leq \|x\|_p + \|y\|_p$

Proof: Let's commit substitution,

$$t_i = \frac{x_i}{\left(\sum_{i=1}^{d} |x_i|^p\right)^{\frac{1}{p}}} \qquad z_i = \frac{y_i}{\left(\sum_{i=1}^{d} |y_i|^q\right)^{\frac{1}{q}}}$$

Now, having b)

$$\sum_{i=1}^{d} |t_i z_i| \leq \sum_{k=1}^{d} \left(\frac{|t_i|^p}{p} + \frac{|z_i|^q}{q} \right)$$

Then, $\sum_{k=1}^{d} \left(\frac{|t_i|^p}{p} + \frac{|z_i|^q}{q} \right) = \sum_{i=1}^{d} \left(\frac{|x_i|^p}{p\left(\sum^{d} |x_i|^p\right)} + \frac{|y_i|^p}{q\left(\sum_{i=1}^{d} |y_i|^q\right)} \right) \leq$

because $\sum |t_i|^p = \frac{\sum_{i=1}^{d} |x_i|^p}{\left|\left(\sum_{i=1}^{d} |x_i|^p\right)^{\frac{1}{p}}\right|^p} = 1$ Same for z_i

Hence, $\sum \frac{|t_i|^p}{p} + \frac{|z_i|^q}{q} = \frac{1}{p} + \frac{1}{q} = 1$, so $\sum |t_i z_i| \leq 1$

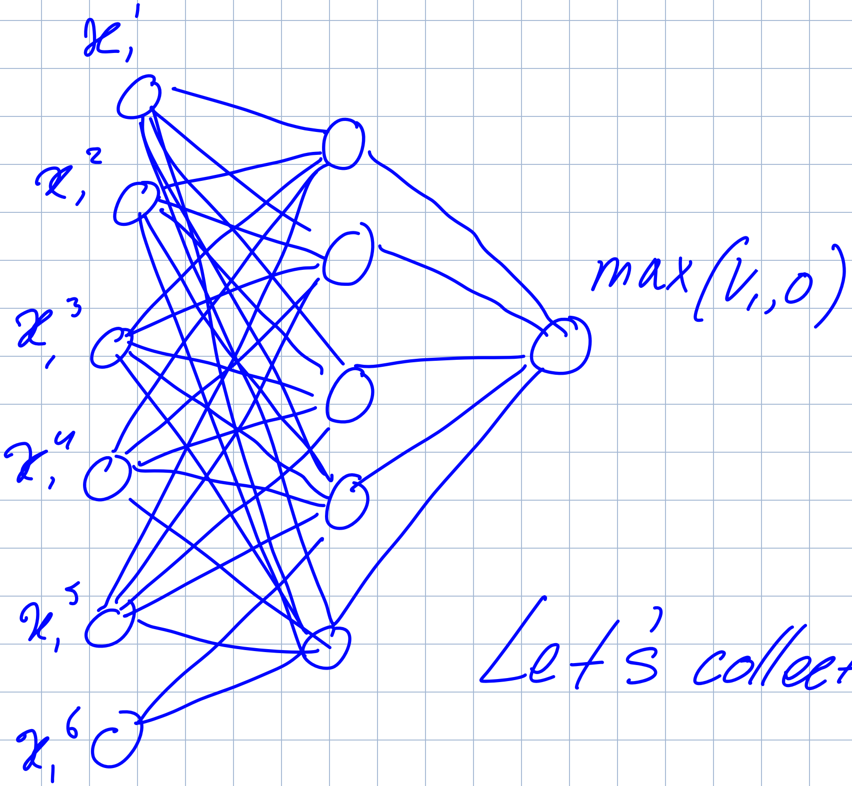
After multiplying by $\left(\sum_{i=1} |x_i|^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^{d} |y_i|^q\right)^{\frac{1}{q}}$ so the proof is done.

№3.

$$\mathcal{F}_n = \{ h_{W_1 \dots W_n} : \forall i, W_i \in \mathbb{R}_+^{N_i \times N_{i-1}}, \forall i \|W_i^j\|_1 = B, \}$$

$N_0 = d \quad N_n = 1.$

Let $n=1$, then $W_1 \in \mathbb{R}_+^{d \times d-1}$



Let's collect some facts:

1) First, notice that we can have no more than d layers. Moreover, we can have only $\sum_{i=1}^{d-n} i$ nodes. Therefore, the amount of weights is $\sum_{i=1}^{d-n} i \cdot (i-1)$

2) ReLu is a 1-Lipschitz function.

Therefore, by contraction inequality,

$$Rad(\sigma(Wx)) := Rad(\sigma(\langle W, x\rangle)) \le Rad(\langle W, x\rangle)$$

The Rad of linear predictor of

ℓ_1-norm is $B_1 X_\infty \sqrt{\frac{2\ln(2d)}{m}}$

So, $Rad(\mathcal{F}_1) \le B_1 X_\infty \sqrt{\frac{2\ln(2d)}{m}}$

Now, consider $Rad(\mathcal{F}_2)$.

$$Rad(\mathcal{F}_2) = \mathbb{E}_\sigma\left[\sup_{h \in \mathcal{F}_1} \frac{1}{m}\sum_{i=1}^m \sigma_i h(x_i)\right)$$

$$= \mathbb{E}_\sigma\left[\sup_{h \in \mathcal{F}_1} \frac{1}{m}\sum \sigma_i \sigma(W_2 \langle W_1, x_i\rangle)\right]$$

$$\le \mathbb{E}_\sigma\left[\sup \frac{1}{m}\sum \sigma_i W_2 \langle W_1, x_i\rangle\right. \text{ by contraction.}$$

Now, what is $W_2 \langle W_1, x_i\rangle$?

This is $g(x) := WX$ – affine transform.

Affine transfom is bounded by its norm. Therefore we can see it as

$\mathcal{R}ad_S(\mathcal{F}_\ell) \leq \mathcal{R}ad(W \circ \tilde{\mathcal{F}}_{\ell-1})$ where

W is B_ℓ- lipschitz. Therefore,

$\mathcal{R}ad_S(\mathcal{F}_\ell) \leq \mathcal{R}ad(W \circ \tilde{\mathcal{F}}_{\ell-1}) =$

$B_\ell \mathcal{R}ad(\tilde{\mathcal{F}}_{\ell-1})$ Having that $\mathcal{R}ad(\mathcal{F}_1) \leq X_\infty B_1 \sqrt{\frac{2\ln(2d)}{m}}$

Therefore, $\boxed{\mathcal{R}ad(\tilde{\mathcal{F}}_n) \leq \left(\prod_{i=1}^{n} B_i\right) X_\infty \sqrt{\frac{2\ln(2d)}{m}}}$

Since $\sigma(z) = \frac{1}{1+e^{z}}$ is also

1- lipcshitz the result shoud be the

same.

N4. a)

$\|W_2\| \leq 20$ So, $B = 20$.

Then log loss: $\ln(1 - \exp(-y\langle w, x\rangle))$

$$\nabla l = \frac{-y(x)\exp(-y\langle w, x\rangle)}{(1 + \exp(-y\langle w, x\rangle))}$$

$\|\nabla l\| \leq \max_i \|x_i\|_2$ So, $\rho = \max_i \|x_i\|_2$

Therefore log loss is $\max \|x_i\|_2$ lipschitz.

Then, having $\|W\| = 20$, we have

$$\eta = \sqrt{\frac{B^2}{\max \|x\|^2 T}} = \sqrt{\frac{400}{\max \|x_i\|^2 T}}$$

got from Shai SGD page.

N 4 b, c

in the CODE-HW.

NS ~ 3 days