CS 189 Introduction to Machine Learning
Fall 2017 HW13

This homework is due **Friday, December 1 at 10pm.**

1 Getting Started

You may typeset your homework in latex or submit neatly handwritten and scanned solutions. Please make sure to start each question on a new page, as grading (with Gradescope) is much easier that way! Deliverables:

1. Submit a PDF of your writeup to assignment on Gradescope, "HW[n] Write-Up"
2. Submit all code needed to reproduce your results, "HW[n] Code".
3. Submit your test set evaluation results, "HW[n] Test Set".

After you've submitted your homework, be sure to watch out for the self-grade form.

(a) Before you start your homework, write down your team. Who else did you work with on this homework? List names and email addresses. In case of course events, just describe the group. How did you work on this homework? Any comments about the homework?

None. Alone
Comments: check spelling of code. e.g. "accuracy" not "accurracy"

(b) Please copy the following statement and sign next to it:

I certify that all solutions are entirely in my words and that I have not looked at another student's solutions. I have credited all external sources in this write up.

I certify that all solutions are entirely in my words and that I have not looked at another student's solutions. I have credited all external sources in this write up.

Hamel

ATTENTION : THE WHOLE SOLUTION TO HW13 PROB 2 IS BASED ON THE OLD HW13 VERSION, ACCORDING TO THE POLICY, I DO NOT NEED TO ADAPT MY SOLUTION TO THE NEW HW13 VERSION TO GET FULL CREDITS.

Problem #2 :

(a) $\sigma(t) = \dfrac{1}{1+e^{-t}}$

$t \to \infty$: $e^{-t} \to 0 \Rightarrow \lim\limits_{t\to\infty} \sigma(t) = \dfrac{1}{1+e^{-t}} = \dfrac{1}{1+0} = 1$ (1)

$t \to -\infty$: $e^{-t} \to +\infty \Rightarrow \lim\limits_{t\to-\infty} \sigma(t) = \dfrac{1}{1+e^{-t}} = \dfrac{1}{1+\infty} = 0$ (2)

$$t_1 > t_2 \Rightarrow -t_1 < -t_2$$
$$\Rightarrow e^{-t_1} < e^{-t_2}$$
$$\Rightarrow 1+e^{-t_1} < 1+e^{-t_2}$$
$$\Rightarrow \frac{1}{1+e^{-t_1}} > \frac{1}{1+e^{-t_2}}$$
$$\sigma(t_1) > \sigma(t_2)$$

$\Rightarrow$ $\sigma(t)$ is monotonically increasing (3)

From (1), (2) & (3) $\Rightarrow$ $\sigma(t)$ is bounded by 0 & 1

$\Rightarrow$ $\sigma(t)$ is thresholding function.

$\theta(t) = \text{ReLu}(t) - \text{ReLu}(t-1) = \max(0,t) - \max(0,t-1)$

3 cases :

1. If $t < 0$: $\theta(t) = 0 - 0 = 0$
2. If $0 \leq t < 1$: $\theta(t) = t - 0 = t$
3. If $t \geq 1$: $\theta(t) = t - (t-1) = 1$

$\Rightarrow$ $t \to \infty$: $\theta(t) \to 1$ $\quad (\theta(t) = 1)$

$t \to -\infty$: $\theta(t) \to 0$

From 3 cases we see that $\theta(t)$ is increasing (not strictly) and bounded by 0 & 1

$\Rightarrow$ $\theta(t)$ is thresholding function

(b) See code attached

(c) We define the set of function:

$$f_k(x) = \tau(\langle w_k, x\rangle + b_k)$$

corresponding to different pairs (w_k, b_k)

Consider the function

$$f^*(x) = \lim_{k\to\infty} f_k(x) = \lim_{k\to\infty} \tau(\langle w_k, x\rangle + b_k)$$

if $k\to\infty$ leads to $\langle w_k, x\rangle + b_k \to \infty$ when $\langle w', x\rangle + b' \geq 0$

and $\langle w_k, x\rangle + b_k \to -\infty$ when $\langle w', x\rangle + b' < 0$

then

$$\lim_{k\to\infty} \tau(\langle w_k, x\rangle + b_k) = 1 \text{ for } x \text{ s.t: } \langle w', x\rangle + b' \geq 0$$

$$\lim_{k\to\infty} \tau(\langle w_k, x\rangle + b_k) = 0 \text{ for } x \text{ s.t: } \langle w', x\rangle + b' < 0$$

since $\tau(t)\to 1$ for $t\to\infty$ & $\tau(t)\to 0$ for $t\to -\infty$ (definition)

In such case, $f^*(x)$ becomes the step function

Now, our job is to find such w_k, b_k so that it satisfies: $k\to\infty$ leads to $\langle w_k, x\rangle + b_k \to \infty$ if $\langle w', x\rangle + b \geq 0$ and $\langle w_k, x\rangle + b_k \to -\infty$ if $\langle w', x\rangle + b < 0$

We can choose w_k and b_k the scaling-up of w' & b with factor k, i.e $w_k = kw'$, $b_k = kb'$

thus, the closesure $cl(\{\tau(\langle w,t\rangle + b)$ for some $w, b\})$

include $\lim_{k\to\infty} \tau(\langle w_k, x\rangle + b_k)$ if we choose

$w = kw'$, $b = kb'$ and set k as large as

possible

This is actually the step function $S(\langle w', x\rangle + b)$, i.e

$$f^*(x) = \lim_{k\to\infty} f_k(x) = \lim_{k\to\infty} \tau(\langle w_k, x\rangle + b_k) = S \ldots$$

Since τ can approach to S as much as we want by scaling up w & b (through k)

$\Rightarrow$ τ is bounded by S

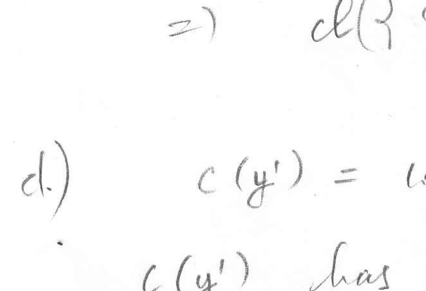
$\Rightarrow$ $cl(\{\tau$ for some $w, b\})$ contains S

d.) $c(y') = \cos(\|w\|_* y')$ where $y' = y/\|w\|$,

$c(y')$ has domain $y' \in [-1, 1]$ and range $c(y') \in [-1, 1]$

i.e $c(y')$ is bounded by $[-1, 1]$

Since $c(y')$ is continuous, we can divide the interval $[-1, 1]$ into $\lceil 2/\delta \rceil$ elements, so each element has size $\frac{(1-(-1))}{\lceil 2/\delta \rceil} < \delta$ (see figure)



It does mean, within each element, the jump of $c(y')$ is less than δ.

Thus if we set the step function at the ends of those elements (as in figure), we can decompose $c(y')$ into the combination of step function w/ jump $\leq \delta$.

e.) we have $c' = \lim_{\Delta z \to 0} \dfrac{c(z_i) - c(z_{i-1})}{z_i - z_{i-1}}$ $\quad \Delta z = z_i - z_{i-1}$

$$\Rightarrow c' \simeq \frac{c(z_i) - c(z_{i-1})}{\Delta z} = \frac{\Delta c(z)}{\Delta z}$$

$$\Rightarrow \Delta c(z) = c' \Delta z$$

$$\int f(z)\,dz = \lim_{\Delta z \to 0} \sum f(z) \Delta z$$

$$\Rightarrow \int f(z)\,dz \simeq \sum f(z) \Delta z$$

$$\Rightarrow \sum f(z) \simeq \frac{1}{\Delta z} \int f(z)\,dz$$

now substitute $f(z)$ by $\Delta c(z)$:

$$\sum f(z) = \sum \Delta c(z) \simeq \frac{1}{\Delta z} \int \Delta c(z)\,dz$$

$$= \frac{1}{\cancel{\Delta z}} \int c' \cancel{\Delta z}\,dz$$

$$= \int c'\,dz$$

z take the range from $-1 \to 1$

$$\Rightarrow \sum_i |c(z_i) - c(z_{i-1})| \simeq \int_{-1}^{1} |c|'\,dz = |c| \Big|_{-1}^{1}$$

c is periodic function w/ period $2\pi / \|w\|_1$

$$\Rightarrow |c| \Big|_{-1}^{1} = \text{range of } c * \#\text{ of periods over } y' \in [-1,1]$$

$$= \underbrace{(1-(-1))}_{\text{range of } c} * \frac{\overbrace{1-(-1)}^{\text{range of } y'}}{2\pi / \|w\|_1}$$

$$= 2 * \frac{\|w\|_1}{\pi} = \frac{\|w\|_1}{\pi/2} < \|w\|_1$$

explanation:
$c(y') = \cos(\|w\|_1 y')$
$\cos()$ has period 2π
$\Rightarrow \|w\|_1 y' = 2\pi$
$\Rightarrow y' = 2\pi / \|w\|_1$ is the period of $c(y')$

$$\Rightarrow \sum_i |c(z_i) - c(z_{i-1})| < \|w\|_1$$

(f)

$$F_{cos} = \{ f : \mathbb{R}^d \to \mathbb{R} \text{ s.t. } f(x) = \frac{1}{2\|w\|_1} \cos(\langle w, x \rangle) \text{ for } w \neq 0 \}$$

each $f(x) \in F_{cos}$ can be decomposed into a combination of step functions scaled up with factor $\frac{1}{2\|w\|_1}$

From 2(c) we have:

$$S \subseteq cl(\{\tau\})$$

$$\Rightarrow c_i S \subseteq cl(\{c_i \tau\})$$

$$\Rightarrow \sum c_i S \subseteq cl(\{\sum c_i \tau\})$$

if we choose $c_i \geq 0$ w/ $\sum_i c_i = 1$

then $\sum c_i \tau$ is the convex hull of $\{\tau\}$, and $cl(\{\sum c_i \tau\})$ is the closed convex hull of $\{\tau\}$, i.e.

$$\sum c_i S \subseteq \overline{conv}(\{\tau\})$$

now we show that $f(x) \in F_{cos}$ can be represented by $\sum c_i S_i$ w/ $c_i \geq 0$ $\sum_i c_i = 1$ and S_i is step function

$$f(x) = \frac{1}{2\|w\|_1} \cos(\langle w, x \rangle)$$

From 2(d) we can write:

$$\cos(\langle w, x \rangle) = \sum_i |c(z_i) - c(z_{i-1})| \, S_i(\langle w, x \rangle + b)$$

$$\Rightarrow f(x) = \sum_i \frac{|c(z_i) - c(z_{i-1})|}{2\|w\|_1} S_i(\ldots)$$

from 2(e) we know that $\sum_i |c(z_i) - c(z_{i-1})|$ is bounded by $\|w\|_1$,

$$\Rightarrow \sum_i \frac{|c(z_i) - c(z_{i-1})|}{2\|w\|_1} \text{ is bounded by } \frac{1}{2}.$$

WHY $\frac{1}{2}$ BUT NOT 1 ? BECAUSE THIS SOLUTION IS BASED ON THE OLD WRONG HW #13 VERSION AND ACCORDING TO THE POLICY, I STILL GET FULL CREDITS IF MY REASONING MAKES SENSE WITHOUT ADAPTING TO THE NEW HW VERSION.

(g) $$E[\|f_P - f\|^2] = E\left[\int_{x \in [0,1]^d} (f_P - f)^2 \, dx\right]$$

$$= \int E[(f_P - f)^2] \, dx$$

$$E[(f_P - f)^2] = Var[f_P - f] + E[f_P - f]^2.$$

$$= Var[f_P] + (E[f_P] - \underbrace{E[f]}_{0})^2$$

$$= Var[f_P] + E[f_P]^2$$

$$= E[f_P^2] = E\left[\frac{1}{P^2}\left(\sum_{i=1}^{P} G_i\right)^2\right]$$

$$= \frac{1}{P^2} E\left[\left(\sum_{i=1}^{P} G_i\right)^2\right] \leq \frac{1}{P^2} \cdot \frac{1}{P} \sum \underbrace{(1 + \dots + 1)}_{P \text{ terms}}^2$$

$$= \frac{1}{P^2} \times \frac{1}{P} \times P^2 = \frac{1}{P}$$

$$\Rightarrow E[(f_P - f)^2] \leq \frac{1}{P}$$

$$\int_{x \in [0,1]^d} E[(f_P - f)]^2 \, dx \leq \int_{x \in [0,1]^d} \frac{1}{P} \, dx = \frac{1}{P} \Rightarrow E[\|f_P - f\|^2] \leq \frac{1}{P}$$

(h) Since

$$E[\| f_p - f \|^2] \leq \frac{1}{p} \qquad (1)$$

is true for a convex combination of p randomly choosen threshosh functions, if there is no deterministic choice of f_p so that (1) is true $\rightarrow$ this contradicts with 2(g), i.e. the equation (1) is impossible:

thus consider

$$E(f, p) = \inf \int_{x \in [0,1]^d} (f(x) - h(x))^2 \, dx$$

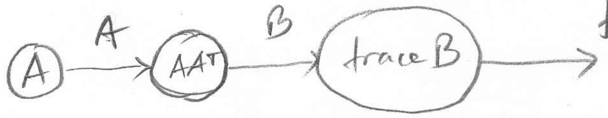
where $h(x) = \sum_1^p c_k \tau(\langle w_k, x \rangle + b_k)$

is such the deterministic choice, and $f(x)$ is any function $f \in F \subseteq \overline{\mathrm{conv}}(\{\tau\})$ that can be represented by $\sum_1^m c_i \tau_i$, we have

$$E(f, p) \leq \frac{1}{p}$$

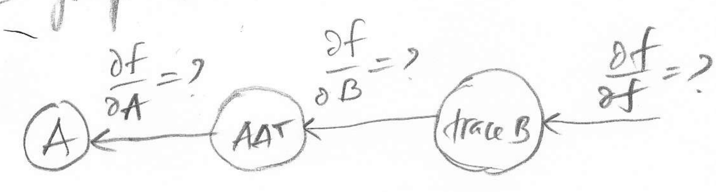
Problem # 4

a.) Given the neural network as below:



given $\sqrt{\text{trace}(AA^T)} = \|A\|_F$

calculate the backward propagation. That is, fill in the ? in the grap below:



Solution:

First: $$\frac{\partial f}{\partial f} = 1.$$

Second: $$\frac{\partial f}{\partial B} = \frac{\partial\, \text{trace}(AA^T)}{\partial\, AA^T} = \frac{\partial \sum_{ij}^{n} A_{ij}^2}{\partial\, AA^T}$$

on the diagonal: $AA^T \to \sum_j A_{ij}^2 \Rightarrow \frac{\partial f}{\partial B_{ii}} = 1$

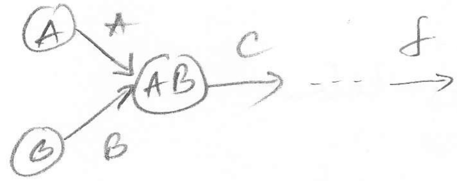
off the diagonal: $AA^T \to \sum_k A_{ik}A_{jk} \Rightarrow \frac{\partial f}{\partial B_{ij}} = 0$

$$\Rightarrow \frac{\partial f}{\partial B} = I \quad \text{(identity matrix)}$$

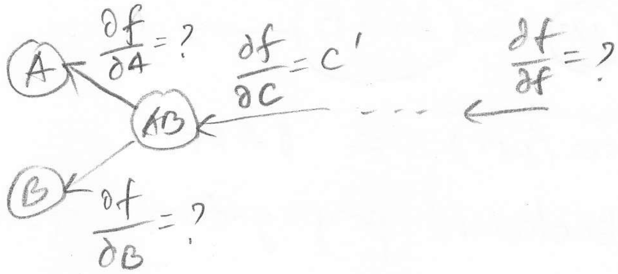
Third: $$\frac{\partial f}{\partial A} = \sum_{ij} \underbrace{\frac{\partial f}{\partial B_{ij}}}_{I} \underbrace{\frac{\partial B_{ij}}{\partial A}}_{2A} = 2A$$

b.)

Given



Find ? :



Solution

$$\frac{\partial f}{\partial f} = 1$$

$$\frac{\partial f}{\partial A} = \left[\frac{\partial f}{\partial A_{ij}} \right] = \left[\sum_{k,l} \frac{\partial f}{\partial C_{kl}} \frac{\partial C_{kl}}{\partial A_{ij}} \right]$$

$$= \left[\sum_{l=1}^{p} \frac{\partial f}{\partial C_{il}} \frac{\partial C_{il}}{\partial A_{ij}} \right] = \sum_{l=1}^{p} C'_{il} B_{jl} = C' B^T$$

Similarly:

$$\frac{\partial f}{\partial B} = A^T C'$$