HomeWork

ZAMOLODTCHIKOV Petr

28 février 2019

1 Exercise 1

Let

$$\mathcal{F}_{m,\sigma} = \left\{ f : \mathbb{R}^d \to \mathbb{R} | f = \sum_{j=1}^{m} c_j \sigma(w_j^\top . + v_j), \forall j \in \{1, \ldots, m\}, w_j \in \mathbb{R}^d, v_j \in \mathbb{R} \right\}$$

It is easy to see that any $f \in \mathcal{F}_{m,\sigma}$, $f = \sum_{j=1}^m c_j \sigma(w_j^\top . + v_j)$ can be described as at most m couples of one hyperplane and one coefficient $((\{w_1^\top x = -v_1\}, c_1), \ldots, (\{w_m^\top x = -v_m\}, c_m))$. The union of those hyperplanes associated with a non-zero coefficient form the set of points where f is not differentiable. We are going to use this property to prove that $\mathcal{F}_{m,\sigma}$ is not convex.

Let $(e_1, \ldots, e_d)$ be the cannonic basis of $\mathbb{R}^d$ and

$$f_1 : x \mapsto \sum_{j=1}^{m} \sigma(je_1 x + 1)$$
$$f_2 : x \mapsto \sum_{j=1}^{m} \sigma(je_2 x + 1)$$

and $f : x \mapsto \frac{1}{2}(f_1 + f_2)$. The set of points where f is not differentiable is exactly the reunion of $2m$ distinct hyperplanes, thus f can not be in $\mathcal{F}_{m,\sigma}$ and $\mathcal{F}_{m,\sigma}$ is not convex.

2 Exercise 2

2.a

Let $f \in \mathcal{H}(K)$, $X_f = \{x \in [0,1], f \text{ is not differentiable in } x\}$ and let D be the set of points on which $\sigma(f)$ is not differentiable. Then $\sigma(f) \in \mathcal{H}(\mathrm{Card}(D) - 1)$ (D includes $\{0, 1\}$). We are going to upper-bound $\mathrm{Card}(D)$.
We denote $\overline{f^{-1}(\{0\})} = \{x \in [0,1],\ f(x) = 0 \text{ and } \forall \varepsilon > 0,\ f(]x - \varepsilon, x + \varepsilon[) \neq \{0\}\}$

$$D = \underbrace{(X_f \cap f^{-1}(\mathbb{R}_+^*))}_{\text{non-differentiability because of } f} \cup \underbrace{\overline{f^{-1}(\{0\})}}_{\text{non-differentiability because of } \sigma}$$

$$\mathrm{Card}(D) \leq \underbrace{\mathrm{Card}(X_f \cap f^{-1}(\mathbb{R}_+^*))}_{=k} + \underbrace{\mathrm{Card}(\overline{f^{-1}(\{0\})})}_{\leq 2\min(k-1, \mathrm{Card}(X_f) - k + 1)}$$

$$\mathrm{Card}(D) \leq k + 2(k-1)\mathbb{1}(k \leq K/2 + 3/2) + 2(K + 1 - k + 1)\mathbb{1}(k > K/2 + 3/2)$$

$$\mathrm{Card}(D) \leq (3K/2+2)\mathbb{1}(k \leq K/2+3/2) + (3/2K+3/2)\mathbb{1}(k > K/2+3/2)$$

Because Card(D) $\in \mathbb{N}$, we have $\mathrm{Card}(D) \leq \lfloor 3K/2+2 \rfloor$.
Ultimately we have

$$\sigma(f) \in \mathcal{H}(\lfloor 3K/2+1 \rfloor)$$

2.b

Let $f_i \in \mathcal{H}(K_i)$, $\forall i \in \{1,\dots,r\}$ and let $f = \sum_{1\leq i \leq r} \lambda_i f_i$ be a linear combination of the f_i. Let X_f be defined as in Exercise 1, for any f_i, $\mathrm{Card}(X_{f_i}) \leq K_i + 1$, lets assume that $\forall i, j, i \neq j \Rightarrow X_{f_i} \cap X_{f_j} = \{0,1\}$. We have :

$$f \in \mathcal{H}(\mathrm{Card}((\cup_{i=1}^r (X_{f_i} - \{0,1\})) \cup \{0,1\}) - 1)$$

And

$$\begin{aligned}
\mathrm{Card}((\cup_{i=1}^r (X_{f_i} - \{0,1\})) \cup \{0,1\}) &\leq 2 + \sum_{i=1}^r (\mathrm{Card}(X_{f_i}) - 2) \\
&\leq 2 + \sum_{i=1}^r (K_i - 1) \\
&= 2 - r + \sum_{i=1}^r K_i
\end{aligned}$$

finally :

$$f \in \mathcal{H}(1 - r + \sum_{i=1}^r K_i)$$

2.c

For a 1 layer network, by using the results of previous questions, f being the sum of p_1 functions of $\mathcal{H}(\lfloor 3/2+1 \rfloor)$, we have $f \in \mathcal{H}(\lfloor 3/2(1-p_1+3/2p_1+p_1) \rfloor) \subset \mathcal{H}(\lfloor 3/2(p_1+1) \rfloor)$
The property is then true for $L=1$, for $L \in \mathbb{N}$ we have to apply the ReLU function once more and then sum the results.

f is the sum of p_L functions of $\mathcal{H}(\lfloor (3/2)^L \prod_{i=1}^{L-1}(p_i+1)+1 \rfloor)$, then

$$f \in \mathcal{H}(\underbrace{\lfloor 1 - p_L + \sum_{j=1}^{p_L}((3/2)^L \prod_{i=1}^{L-1}(p_i+1)+1) \rfloor}_{A})$$

$$A = \lfloor 1 + p_L(3/2)^L \prod_{i=1}^{L-1}(1+p_i) \rfloor \leq (3/2)^L \prod_{i=1}^{L}(p_i+1)$$

The upper bound for the number of pieces is then true for any $(L, (p_1, \dots, p_L))$

Extension : Sharper bound for 2.c

Here we are going to prove the following fact :

A deep ReLU network with architecture $(L,(1,p_1,\dots,p_L,1))$ and ρ being the identity is piercewise linear in the input with at most $1+\sum_{i=1}^{L}(\frac{3}{2})^i\prod_{j=1}^{i}p_{L-j+1}$ pieces.

To do so, we are first going to show by inductions on the layers that any node of the k-th (hidden) layer outputs a piercewise linear function of the input with at most $1+3/2+\sum_{i=2}^{k}(\frac{3}{2})^i\prod_{j=1}^{i-1}p_{k-j}$ pieces.
The output of any node of the first hidden layer is piercewise linear in the input with at most $3/2+1$ pieces.
The input of any node in the second hiddel layer is the sum of p_1 functions that are piercewise linear with at most $3/2+1$ pieces. This sum is piercewise linear with at most $1-p_1+p_1(3/2+1)=1+3/2p_1$ pieces, and the output of any node of the second hidden layer is then piercewise linear with at most $\frac{3}{2}(1+\frac{3}{2}p_1)+1=1+\frac{3}{2}+(\frac{3}{2})^2p_1$.

The property holds for $k=1,\,2$ (with $p_0=1$).

Assume that the output of the k-th layer is piercewise linear with at most $1+3/2+\sum_{i=2}^{k}(\frac{3}{2})^i\prod_{j=1}^{i-1}p_{k-j}$. Then the output of the layer $k+1$ is piercewise linear with at most :

$$\begin{aligned}
K &= \frac{3}{2}(1-p_k+p_k(1+3/2+\sum_{i=2}^{k}(\frac{3}{2})^i\prod_{j=1}^{i-1}p_{k-j}))+1\\
&= 1+\frac{3}{2}+(\frac{3}{2})^2p_k+\sum_{i=2}^{k}(\frac{3}{2})^{i+1}p_k\prod_{j=1}^{i-1}p_{k-j})\\
&= 1+\frac{3}{2}+\sum_{i=2}^{k+1}(\frac{3}{2})^i\prod_{j=1}^{i-1}p_{k+1-j}
\end{aligned}$$

Thus the property is true for any $k\in\mathbb{N}$. After the layer L, the final output is the sum of all the outputs of the last layer :

Any node of the layer L outputs a piercewise linear function of the input, that has at most $1+\frac{3}{2}+\sum_{i=2}^{L}(\frac{3}{2})^i\prod_{j=1}^{i-1}p_{L-j}$ pieces.
Thus the final output is piercewise linear with at most :

$$\begin{aligned}
K(L,p_1,\dots,p_L) &= 1-p_L+p_L(1+3/2+\sum_{i=2}^{L}(\frac{3}{2})^i\prod_{j=1}^{i-1}p_{L-j})\\
&= 1+\frac{3}{2}p_L+\sum_{i=2}^{L}(\frac{3}{2})^ip_L\prod_{j=1}^{i-1}p_{L-j}\\
&= 1+\sum_{i=1}^{L}(\frac{3}{2})^i\prod_{j=1}^{i}p_{L-j+1}
\end{aligned}$$

And we have our result.

Now we are going to prove that this bound is indeed sharper than the one we had previously.

$$\Delta = 1 + \sum_{i=1}^{L} (\frac{3}{2})^i \prod_{j=1}^{i} p_{L-j+1} - (\frac{3}{2})^L \prod_{i=1}^{L} (p_i + 1)$$

$$\begin{aligned}\Delta &= 1 + \sum_{i=1}^{L} (\frac{3}{2})^i \prod_{j=1}^{i} p_{L-j+1} - (\frac{3}{2})^L (1 + \sum_{i=1}^{L} \sum_{1 \le k_1 < \dots < k_i \le L} p_{k_1} \dots p_{k_L}) \\ &\le 1 - 3/2 + \sum_{i=1}^{L} (\frac{3}{2})^i \Big[\prod_{j=1}^{i} p_{L-j+1} - \sum_{1 \le k_1 < \dots < k_i \le L} p_{k_1} \dots p_{k_L} \Big] \\ &\le -1/2 - \sum_{i=1}^{L} (\frac{3}{2})^i \sum_{\substack{1 \le k_1 < \dots < k_i \le L \\ k_i \ne L, \dots, k_1 \ne p_{L-i+1}}} p_{k_1} \dots p_{k_L} < 0\end{aligned}$$

Thus this bound is sharper than the previous one.

2.d

Lets majorate $||f-h||_{[0,1],\infty}$ for $h \in \mathcal{H}(K)$.

First of all lets denote $\mathcal{L}(K)$ the set of piercewise linear functions on $[0,1]$ that are not necessarily $\mathcal{C}^0$ with at most K pieces.
We have $\mathcal{H}(K) \subset \mathcal{L}(K)$, so $\underset{h \in \mathcal{L}(K)}{\text{Inf}} ||f-h||_{[0,1],\infty} \le \underset{h \in \mathcal{H}(K)}{\text{Inf}} ||f-h||_{[0,1],\infty}$

Now take a pair $0 \le a < b \le 1$. Let $M > 0$, and suppose $\exists h : x \mapsto \alpha x + \beta$ such that $||f-h||_{[a,b],\infty} \le M$ (For M great enough this assertion is obvious). Now let's characterize the set of linear functions $A(M) = \Big\{ h \text{ linear }, ||f-h||_{[a,b],\infty} \le M \Big\}$:

$$\forall h \in A(M), \forall \lambda \in [0,1], a^2 + \lambda(b^2 - a^2) - M \le h(a + \lambda(b-a)) \le (a + \lambda(b-a))^2 + M$$

$$\begin{aligned}\forall \lambda \in [0,1], a^2 + \lambda(b^2 - a^2) - M &\le (a + \lambda(b-a))^2 + M \\ \forall \lambda \in [0,1], (b-a)^2 \lambda(1-\lambda) &\le 2M \\ \forall \lambda \in [0,1], \underbrace{\lambda(1-\lambda)}_{\le 1/4} &\le \frac{2M}{(b-a)^2}\end{aligned}$$

The optimal M is then $M_{opt} = \frac{(b-a)^2}{8}$. $A(M_{opt}) \ne \emptyset$ because $x \mapsto a^2 + \frac{x-a}{b-a}(b^2 - a^2) - M_{opt} \in A(M_{opt})$ and the overall distance is $||f-h||_{[0,1],\infty} = \frac{1}{8} \max_{1 \le k \le K} ((x_{k-1} - x_k)^2) \ge \frac{1}{8K^2}$

2.e

By removing the inactive nodes, we have that

$$f \in \mathcal{F}(L, p, s) \Rightarrow f \in \mathcal{F}(L, (p_0, p_1 \wedge s, \dots, p_L \wedge s, p_{L+1}), s)$$

Then with regards to the exercise, we have $p_0 = p_{L+1} = 1$ and we can directly apply the result of 2.c to obtain that f has at most $(3/2)^L \prod_{i=1}^{L} (p_i \wedge s + 1)$ pieces. This is upper bounded by $K = (3/2)^L(s+1)^L$.

By using the result of 2.d we have that $\inf\limits_{h\in\mathcal{F}(L,p,s)} ||h-f||_\infty \geq \frac{1}{8K^2}$

From these two facts we conclude that if $s \leq c(L)\epsilon^{-1/2L}$ then $\inf\limits_{h\in\mathcal{F}(L,p,s)} ||h-f||_\infty \geq \epsilon$
With $c(L) = \frac{2}{3}8^{-1/2L}$

3 Exercise 3

3.a

One can easily see that $\hat{f} \in \{f \in \mathcal{F}_{n,\sigma}, f(U_i) = Y_i, \forall i = 1,\dots,n\} \neq \emptyset$. This set is not empty because $\hat{f} \in \mathcal{H}(n+1)$, and one can always find a piercewise linear function which break points are in $(\alpha, U_1, \dots, U_n)$ with $\alpha \leq 0$.

3.b

3.c

For $f = 0$ and $n \geq 2$, we have :

$$\mathbb{E}[\|\hat{f} - f\|_{L^2,[0,1]}] \geq 1/16\mathbb{E}[\max_i U_i - \min_i U_i]$$

Now lets majorate $\mathbb{E}[\max_i U_i - \min_i U_i]$

$$\mathbb{E}[\max_i U_i - \min_i U_i] = \underbrace{\mathbb{E}[\max_i U_i]}_{A} - \underbrace{\mathbb{E}[\min_i U_i]}_{B}$$
$$A = \int_{[0,1]} (1 - \Phi_U(t)^n)\mathrm{d}t = 1 - \frac{1}{n+1}$$
$$B = \int_{[0,1]} (1 - \Phi_U(t))^n\mathrm{d}t = \frac{1}{n+1}$$

Where Φ_U is the CDF of $U \sim \mathcal{U}([0,1])$. Finally,

$$\forall n \geq 2,\ \mathbb{E}[\|\hat{f} - f\|_{L^2,[0,1]}] \geq 1/16$$