ASSIGNMENT 2: Introduction to Neural Network

Détagnon Bernic GBAGUIDI

March 9, 2024

Problem 1 : The name of Softmax

Let $z \in \mathbb{R}^d$. Let calculate $\lim_{C \longrightarrow +\infty} softmax(Cz)$ and $\lim_{C \longrightarrow -\infty} softmax(Cz)$.

We have,

$$softmax(z)_i = \frac{\exp(z_i)}{\sum_{j=1}^{d} exp(z_j)}$$

So,

$$softmax(Cz)_i = \frac{\exp(Cz_i)}{\sum_{j=1}^{d} exp(Cz_j)}$$

But we can rewrite like ,

$$softmax(Cz)_i = \frac{1}{1 + \sum_{j \neq i} exp(C(z_j - z_i))}$$

Let compute the limit in the following cases.

- $\max_{j \in [|1,d|]}(z_j) = z_i$

  Let denote $Z = \{z_i, \max_{j \in [|1,d|]}(z_j) = z_i\}$

  We have

$$\begin{aligned}\lim_{C \longrightarrow +\infty} softmax(Cz)_i &= \lim_{C \longrightarrow +\infty} \frac{1}{1 + \sum_{j \neq i} exp(C(z_j - z_i))} \\ &= \frac{1}{card(Z)}\end{aligned}$$

  Because in this case those z_j that are equal to the max will lead to 1 in the denominator while the remain will be equal to 0.

$$\begin{aligned}\lim_{C \longrightarrow -\infty} softmax(Cz)_i &= \lim_{C \longrightarrow -\infty} \frac{1}{1 + \sum_{j \neq i} exp(C(z_j - z_i))} \\ &= 0\end{aligned}$$

  Because in this case those z_j that are equal to the max will lead to 1 in the denominator while the remain will be equal to $+\infty$.

- $\min_{j \in [|1,d|]}(z_j) = z_i$ and $z_i \neq max_{j \in [|1,d|]}(z_j)$

  Let denote $Z1 = \{z_i, \min_{j \in [|1,d|]}(z_j) = z_i\}$

We have

$$\begin{aligned}\lim_{C \longrightarrow +\infty} softmax(Cz)_i =& \lim_{C \longrightarrow +\infty} \frac{1}{1 + \sum_{j \neq i} exp(C(z_j - z_i))} \\ =& 0\end{aligned}$$

Because in this case those z_j that are equal to the min will lead to 1 in the denominator while the remain will be equal to $+\infty$.

$$\begin{aligned}\lim_{C \longrightarrow -\infty} softmax(Cz)_i =& \lim_{C \longrightarrow -\infty} \frac{1}{1 + \sum_{j \neq i} exp(C(z_j - z_i))} \\ =& \frac{1}{card(Z1)}\end{aligned}$$

Because in this case those z_j that are equal to the min will lead to 1 in the denominator while the remain will be equal to 0.

- It exists z_{j_0} and z_{j_1} such that $z_{j_0} > z_i$ and $z_{j_1} < z_i$.

  We have

$$\lim_{C \longrightarrow +\infty} softmax(Cz)_i = 0$$
$$\lim_{C \longrightarrow -\infty} softmax(Cz)_i = 0$$

  Because each time one of the limit will give $+\infty$ in denominator.

  With all of those conditions can conclude that:

- The limit of $softmax(Cz)$ when C goes to $+\infty$ give a vector which components are $\frac{1}{card(Z)}$ where are located the maximums of element of z and 0 for other components.
- The limit of $softmax(Cz)$ when C goes to $-\infty$ give a vector which components are $\frac{1}{card(Z1)}$ where are located the minimums of element of z and 0 for other components.

Problem 2 : ReLU of a gaussian

Let $G \sim \mathcal{N}(0,1)$ be a gaussian variable. Let $Y = ReLU_\alpha(G)$.
Let calculate $E(Y)$ and $Var(Y)$

- $E(Y)$ $ReLu$ is a continous function and $G \sim \mathcal{N}(0,1)$, then ,

$$\begin{aligned}
E(Y) =& E(ReLU(G)) \\
=& \int_{\mathbb{R}} ReLU(t)\phi(t)dt \text{ Where } \phi(t)\text{is the pdf of } G \\
=& \int_O^{+\infty} ReLU(t)\phi(t)dt + \int_{-\infty}^0 ReLU(t)\phi(t)dt \\
=& \int_O^{+\infty} t\frac{1}{\sqrt{2\pi}}\exp(-t^2)dt + \int_{-\infty}^0 \alpha t\frac{1}{\sqrt{2\pi}}\exp(-t^2)dt \\
=& [-\frac{1}{\sqrt{2\pi}}\exp(-t^2)]_0^{+\infty} + [-\alpha\frac{1}{\sqrt{2\pi}}\exp(-t^2)]_0^{+\infty} \\
=& \frac{1}{\sqrt{2\pi}} - \frac{\alpha}{\sqrt{2\pi}} \\
=& \frac{1-\alpha}{\sqrt{2\pi}}
\end{aligned}$$

- Var(Y)

  But for this purpose let compute $E(Y^2)$

- $E(Y^2)$

$$
\begin{aligned}
E(Y^2) =& E(ReLU(G)^2) \\
=& \int_{\mathbb{R}} ReLU(t)^2 \phi(t) dt \text{ Where } \phi(t) \text{is the pdf of } G \\
=& \int_O^{+\infty} ReLU(t)^2 \phi(t) dt + \int_{-\infty}^0 ReLU(t)^2 \phi(t) dt \\
=& \int_O^{+\infty} t^2 \frac{1}{\sqrt{2\pi}} \exp(-t^2) dt + \int_{-\infty}^0 \alpha^2 t^2 \frac{1}{\sqrt{2\pi}} \exp(-t^2) dt \\
& \text{We need to integrate by part} \\
& \text{We will choose } u(t) = t \text{ and } v'(t) = t \frac{1}{\sqrt{2\pi}} \exp(-t^2) \text{, so} \\
E(Y^2) =& [-\frac{1}{\sqrt{2\pi}} t \exp(-t^2)]_0^{+\infty} + \int_0^{+\infty} \frac{1}{\sqrt{2\pi}} \exp(-t^2) dt + \\
& [-\alpha^2 \frac{1}{\sqrt{2\pi}} t \exp(-t^2)]_{-\infty}^0 + \int_{-\infty}^0 \frac{\alpha^2}{\sqrt{2\pi}} \exp(-t^2) dt \\
=& 0 + \frac{1}{2} + 0 + \frac{\alpha^2}{2} \\
=& \frac{1 + \alpha^2}{2}
\end{aligned}
$$

We can then deduce $Var(Y)$.

$$Var(Y) = E(Y^2) - (E(Y))^2$$

$$Var(Y) = \frac{1 + \alpha^2}{2} - (\frac{1 - \alpha}{\sqrt{2\pi}})^2$$

$$Var(Y) = \frac{(1 + \alpha^2)\pi - (1 - \alpha)^2}{2\pi}$$

Problem 3 The power of Linear Neural Network

Let $d \geq 3$. let define $\mathcal{X} : \{-1, 1\}^d \longrightarrow \{-1, 1\}$ as

$$\mathcal{X}(x_1, \ldots, x_d) = \prod_{i=1}^{d} x_i$$

Let show that there is no neural network $\Phi(x) = v^T(Wx + b)$ that compute $\mathcal{X}$, where, $W \in \mathbb{R}^{n \times d}, b, v \in \mathbb{R}^n$.

Firstly let rewrite $\Phi(x) = v^T(Wx + b)$
We have
$$\Phi(x) = \sum_{i=1}^{n}(\sum_{j=1}^{d} W_{ij}x_j v_i + b_i v_i)$$
Let reason by contradiction assuming that Φ compute $\mathcal{X}$. Let assume also that d is even.

- Let $x \in \{-1, 1\}^d$, if all the component of x is 1 then
$$\Phi(x) = 1$$
i.e
$$\sum_{i=1}^{n}(\sum_{j=1}^{d} W_{ij}v_i + b_i v_i) = 1$$
if all the component are -1 then ,
$$\Phi(x) = 1$$
i.e
$$\sum_{i=1}^{n}(\sum_{j=1}^{d} -W_{ij}v_i + b_i v_i) = 1$$

- Let $x_k \in \{-1, 1\}^d$ such that at the position k it is 1 but -1 everywhere else.

  We have $\Phi(x_k) = -1$.

  $\Phi(x_k) = -1$ implies that,
$$\sum_{i=1}^{n}(\sum_{j \neq k} -W_{ij}v_i + W_{ik}v_i + b_i v_i) = -1$$
but,
$$\sum_{i=1}^{n}(\sum_{j \neq k} -W_{ij}v_i + W_{ik}v_i + W_{ik}v_i - W_{ik}v_i + b_i v_i) = -1$$
so
$$\sum_{i=1}^{n}(\sum_{j=1}^{d} -W_{ij}v_i + 2W_{ik}v_i + b_i v_i) = -1$$

what is again,

$$\sum_{i=1}^{n}(\sum_{j=1}^{d} -W_{ij}v_i + b_iv_i) + \sum_{i=1}^{n} 2W_{ik}v_i = -1$$

but we know that $\sum_{i=1}^{n}(\sum_{j=1}^{d} -W_{ij}v_i + b_iv_i) = 1$, then ,

$$1 + \sum_{i=1}^{n} 2W_{ik}v_i = -1$$

so we can conclude that,

$$\sum_{i=1}^{n} W_{ik}v_i = -1 \tag{1}$$

- Let take another x_k such that at the same position k like previously it is -1 but 1 everywhere else.

  We have $\Phi(x_k) = -1$.

  $\Phi(x_k) = -1$ implies that,

$$\sum_{i=1}^{n}(\sum_{j\neq k} W_{ij}v_i - W_{ik}v_i + b_iv_i) = -1$$

  but,

$$\sum_{i=1}^{n}(\sum_{j\neq k} W_{ij}v_i - W_{ik}v_i + W_{ik}v_i - W_{ik}v_i + b_iv_i) = -1$$

  so

$$\sum_{i=1}^{n}(\sum_{j=1}^{d} W_{ij}v_i - 2W_{ik}v_i + b_iv_i) = -1$$

  what is again,

$$\sum_{i=1}^{n}(\sum_{j=1}^{d} W_{ij}v_i + b_iv_i) - \sum_{i=1}^{n} 2W_{ik}v_i = -1$$

  but we know that $\sum_{i=1}^{n}(\sum_{j=1}^{d} W_{ij}v_i + b_iv_i) = 1$, then ,

$$1 - \sum_{i=1}^{n} 2W_{ik}v_i = -1$$

so we can conclude that,

$$\sum_{i=1}^{n} W_{ik} v_i = 1 \tag{2}$$

From (1) and (2) we have a contraction , then we can then conclude that the task is not possible .