Assignment2: Introduction to Neural Network

Done by:

HABIMANA Martin

Problem 1: The name of softmax

Let $z \in \mathbb{R}^d$. Calculate $\lim_{C\to\infty} \text{softmax}(Cz)$ and $\lim_{C\to-\infty} \text{softmax}(Cz)$.

By the definition of softmax,

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

C approaches Positive infinity:

$$\begin{aligned}\lim_{C\to\infty} \text{softmax}(Cz)_i &= \lim_{C\to\infty} \frac{\exp(Cz_i)}{\sum_{j=1}^{d} \exp(Cz_j)} \\ &= \lim_{C\to\infty} \frac{\exp(Cz_i)}{\exp(Cz_i) + \sum_{j\neq i} \exp(Cz_j)} \\ &= \lim_{C\to\infty} \frac{1}{1 + \frac{\sum_{j\neq i}\exp(Cz_j)}{\exp(Cz_i)}}\end{aligned}$$

Since e^{Cz_i} dominates, this limit is 1 with the maximum value in z and the others are 0. Mathematically:

$$\boxed{\lim_{C\to+\infty} \text{softmax}(Cz) = \begin{cases} 1 & \text{if } i = \text{argmax}(z) \\ 0 & \text{otherwise} \end{cases}}$$

C approaches negative infinity:

$$\begin{aligned}\lim_{C\to-\infty} \text{softmax}(Cz)_i &= \lim_{C\to-\infty} \frac{\exp(Cz_i)}{\sum_{j=1}^{d} \exp(Cz_j)} \\ &= \lim_{C\to-\infty} \frac{1}{1 + \frac{\sum_{j\neq i}\exp(Cz_j)}{\exp(Cz_i)}}\end{aligned}$$

$$\boxed{\lim_{C\to-\infty} \text{softmax}(Cz) = \begin{cases} 1 & \text{if } z_i \text{ is the minimum component of } z \\ 0 & \text{otherwise} \end{cases}}$$

Problem 2: ReLU of a gaussian

Let $G \sim \mathcal{N}(0,1)$ be a Gaussian random variable. Recall that for $0 \leq \alpha \leq 1$, the leaky ReLU activation is given as

$$\text{ReLU}_\alpha(x) = \begin{cases} x & \text{if } x \geq 0, \\ \alpha x & \text{if } x < 0. \end{cases}$$

Let $Y = \text{ReLU}_\alpha(G)$. calculate $E[Y]$ and $\text{Var}[Y]$.

Answer

The ReLU is given by:

$$\text{ReLU}_\alpha(x) = \begin{cases} x, & \text{if } x \geq 0 \\ \alpha x, & \text{if } x < 0 \end{cases}$$

The pdf of the standard normal distribution $G \sim N(0, 1)$ is as follow:

$$f_G(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

The expected value $E[Y]$ is given by:

$$\begin{aligned}
E[Y] = E[\text{ReLU}_\alpha(G)] &= \int_{-\infty}^{\infty} \text{ReLU}_\alpha(x) \cdot f_G(x)\, dx && (1)\\
&= \int_0^{\infty} x \cdot f_G(x)\, dx + \int_{-\infty}^{0} \alpha x \cdot f_G(x)\, dx && (2)\\
&= \int_0^{\infty} x \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx + \int_{-\infty}^{0} \alpha x \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx && (3)\\
&= \frac{1}{\sqrt{2\pi}} \int_0^{\infty} x e^{-x^2/2} dx + \frac{\alpha}{\sqrt{2\pi}} \cdot \int_{-\infty}^{0} x e^{-x^2/2} dx && (4)
\end{aligned}$$

We can solve this integration by part

Let:

- $u = \frac{x^2}{2}$
- $du = \frac{2x}{2} dx \quad \Rightarrow du = x dx$

If:

- $x = 0$ then $u = 0$
- $x = \infty$ then $u = \infty$

By replace them into (4), we get:

$$\begin{aligned}
E[Y] &= \frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{-u}\, du + \frac{\alpha}{\sqrt{2\pi}} \int_{\infty}^{0} e^{-u}\, du\\
&= -\frac{1}{\sqrt{2\pi}} e^{-u} \Big|_0^{\infty} - \frac{\alpha}{\sqrt{2\pi}} e^{-u} \Big|_{\infty}^{0}\\
&= -\frac{1}{\sqrt{2\pi}} (0 - 1) - \frac{\alpha}{\sqrt{2\pi}} (1 - 0)\\
&= \frac{1}{\sqrt{2\pi}} - \frac{\alpha}{\sqrt{2\pi}}\\
&= \frac{1}{\sqrt{2\pi}} - \frac{\alpha}{\sqrt{2\pi}}\\
&= \frac{1 - \alpha}{\sqrt{2\pi}}
\end{aligned}$$

Hence,

$$\boxed{E(Y) = \frac{1 - \alpha}{\sqrt{2\pi}}}$$

$Var[Y]$:

The variance $Var[Y]$ is given by:

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

Where Y^2:

$$Y^2 = \begin{cases} x^2, & \text{if } x \geq 0 \\ (\alpha x)^2, & \text{if } x < 0 \end{cases}$$

Therefore:

$$\begin{aligned} E(Y^2) &= \int_{-\infty}^{\infty} \mathrm{ReLU}_\alpha^2(x) \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx \\ &= \int_0^{\infty} x^2 \cdot f_G(x)\, dx + \int_{-\infty}^{0} (\alpha x)^2 \cdot f_G(x)\, dx \\ &= \int_0^{\infty} x^2 \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx + \alpha^2 \int_{-\infty}^{0} x^2 \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx \\ &= \frac{1}{\sqrt{2\pi}} \int_0^{\infty} x^2 \cdot e^{-\frac{x^2}{2}}\, dx + \frac{\alpha^2}{\sqrt{2\pi}} \int_{-\infty}^{0} x^2 \cdot e^{-\frac{x^2}{2}}\, dx \end{aligned}$$

We integrate by parts:

- $\mu = \frac{x^2}{2}, dv = e^{-\mu} d\mu.$
- $\int_0^{\infty} \mu e^{-\mu} d\mu = \frac{\sqrt{2\pi}}{2}$

Then we will finally get:

$$\begin{aligned} E[Y^2] &= \frac{1}{\sqrt{2\pi}} \cdot \frac{\sqrt{2\pi}}{2} + \frac{\alpha^2}{\sqrt{2\pi}} \cdot \frac{\sqrt{2\pi}}{2} && (5) \\ &= (1+\alpha^2) \frac{\sqrt{2\pi}}{2\sqrt{2\pi}} && (6) \\ &= \frac{1}{2}(1+\alpha^2) && (7) \end{aligned}$$

Now, let's substitute $E[Y^2]$ into the formula of variance, we get:

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

Hence,

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

Problem 3: Power of linear neural networks

Let $d \geq 3$. Define the function $\chi : \{-1,1\}^d \to \{-1,1\}$ as

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

Let $n \geq 1$, $W \in \mathbb{R}^{n \times d}$, $b \in \mathbb{R}^n$, and $v \in \mathbb{R}^n$. Define a linear neural network $\Phi(x) = v^T \cdot (Wx + b)$.
We say that Φ computes χ if $\Phi(x) = \chi(x)$ for every $x \in \{-1,1\}^d$. For every $d \geq 3$, either find n, W, b, v such that Φ computes χ or prove that the task is impossible.

Answer:

- $\chi(x) = \prod_{i=1}^{d} x_i$
- $\Phi(x) = v^T.(Wx + b) = \sum_{i=1}^{n} \left(\sum_{j=1}^{d} W_{ij} x_j v_i + b_i \right)$

let's verify that: $\chi(x) = \Phi(x)$ or show that it is impossible.

Let's take k belonging to {1,2,...,d} such that $x_j = 1$ for all $j \neq k$ and $x_k = -1$

We have then:

$\chi(x) = -1 \quad \Longrightarrow \quad \phi(x) = -1$

So,

$$\sum_{i=1}^{x} (-W_{ik} v_i + \sum_{j=1 \neq k}^{d} W_{ij} v_j + b_i) = -1 \tag{8}$$

$$\sum_{i=1}^{n} (-W_{ik} v_i - W_{ik} v_i + W_{ik} v_i + \sum_{j=1 \neq k}^{d} W_{ij} v_j + b_i) = -1 \tag{9}$$

$$\sum_{i=1}^{n} (-2W_{ik} v_i + \sum_{j=1}^{d} W_{ij} v_j + b_i) = -1 \tag{10}$$

Suppose then $x_i = 1$ for all i, then :

$$\chi(x) = 1 \iff \sum_{i=n}^{n} (\sum_{j=1}^{d} W_{ij} v_i + b_i) = 1 \tag{11}$$

If we combine the equation(11) and (10), we get $\sum_{i=1}^{n} W_{ik} v_i = 1$ for all k.

Now, let's take $x_1 = -1$ and $x_2 = -1$ and $x_i = 1$ for all i , then we get

$$\chi(x) = 1$$

Therefore,

$$\sum_{i=1}^{n}(-W_{i2}v_i - W_{i2}v_i + \sum_{j>2}^{d} W_{ij}v_j + b_i) = 1$$

$$\sum_{i=1}^{n}(-2W_{i1}v_i - 2W_{i2}v_i + W_{i1}v_i + W_{i2}v_i + \sum_{j>2}^{d} W_{ij}v_j + b_i) = 1$$

$$\sum_{i=1}^{n}(-2W_{i1}v_i - 2W_{i2}v_i) + \sum_{i=1}^{n}(\sum_{j=1}^{d} W_{ij}v_j + b_i) = 1$$

From the equation (10) we have:

$$\sum_{i=1}^{n}(-2W_{i1}v_i - 2W_{i2}v_i) = 0 \tag{12}$$

$$\sum_{i=1}^{n} W_{i1}v_i = -\sum_{i=1}^{n} W_{i2}v_i \tag{13}$$

From the equation (11) take $k = 1, 2$, then we can write:

$$\sum_{i=1}^{n} W_{i1}v_t = 1 = \sum_{i=1}^{n} W_{i2}v_i$$

This leads to a contradiction because:

$$\begin{cases} \sum_{i=1}^{n} W_{i1}v_i = \sum_{i=1}^{n} W_{i}v_i \\ \sum_{i=1}^{n} W_{i1}v_i = -\sum_{i=1}^{n} W_{i2}v_i \end{cases}$$

We conclude that it is not possible to find $, W, b, \text{and} v \text{such that}$ $\chi(x) = \Phi(x)$.