HW1_316228022_316539535

יום שישי 26 ינואר 2024   15:08

Hw 1

Akiyad Gerish 316228022
Over Sela 316539535

Q1:

Q1

Consider a multivariate logistic regression problem. Recall that given a vector $x \in \mathbb{R}^d$, and a parameter matrix $W \in \mathbb{R}^{d\times c}$, where c is the number of classes, the model can be written in the following manner using the Softmax function:

$$softmax(z)_{[i]} = \frac{exp(z_i)}{\sum_j exp(z_j)},$$

where, $z = W^T x$.

1. Show that $softmax(z) = softmax(z+m)$ for every constat m (Where $z+m$ means adding m to every element in z).
2. For $c = 2$, show that the Sigmoid function is equivalent to the Softmax function.
3. Present an alternative to the Sigmoid function which also maps from the real line to the $[0,1]$ interval (any valid solution will be acceptable here).

Solution 1:

@ Let m be a constant. We will show $softmax(z+m) = softmax(z)$ where

$$softmax(z)_{[i]} = \frac{e^{z_i}}{\sum_j e^{z_j}} \qquad z = W^T x \in R^c$$

Proof:

$$softmax(z+m)_{[i]} = \frac{e^{z_i+m_i}}{\sum_j e^{z_j+m_j}} =$$

$$= \frac{e^{z_i}\cdot e^{m_i}}{\sum_j e^{z_j}\cdot e^{m_j}} \underset{m \text{ is const}}{=} \frac{e^{m_i}\cdot e^{z_i}}{e^{m_i}\sum_j e^{z_j}} =$$

$\forall i,j \; m_j = m_i$.

e^{z_i}

$$= \frac{e^{z_i}}{\sum_j e^{z_j}} = \text{Softmax}(z)_{[i]}$$

$\boxed{}$

<u>Solution 2 i</u> $c = 2$: $z \in R^2 \Rightarrow$

$$z = (z_1, z_2) \Rightarrow$$

$$j \neq i$$

$$\text{softmax}_{(i)} = \frac{e^{z_i}}{e^{z_i} + e^{z_j}} \cdot \frac{e^{-z_j}}{e^{-z_j}} =$$

define $z_i - z_j = x$

$$= \frac{e^{z_i - z_j}}{e^{z_i - z_j} + 1} = \frac{e^x}{1 + e^x} \cdot \frac{e^{-x}}{e^{-x}} = \frac{1}{1 + e^{-x}} =$$

$$= \frac{1}{1 + e^{-}} = \text{Sigmoid}(x)$$

$\boxed{}$

<u>Solution 3</u>

An Alternative to Sigmoid (x) which also maps from $\mathbb{R} \longrightarrow [0,1]$ is:

$$g(x) = \frac{1}{2}\left(tgh(x) + 1\right)$$

where $tgh(x) = \dfrac{e^x - e^{-x}}{e^x + e^{-x}}$

$g(x): \mathbb{R} \to [-1, 1]$

hence $g(x): \mathbb{R} \to [0, 1]$

Q2:

$\boxed{V}$

Q2

Let $x \in \{0,1\}^2$ be an input vector. Consider the following model (scalar function):

$$f(x) = w^T h + b_2$$
$$h = max(U^T x + b_1, 0)$$

Where $U \in \mathbb{R}^{2\times 2}, b_1 \in \mathbb{R}^2, w \in \mathbb{R}^2, b_2 \in \mathbb{R}$, and the max is taken element-wise.

Suppose we would like to represent with $f(x)$ the XOR function, defined as:

$$XOR(0,0) = 0,$$
$$XOR(0,1) = 1,$$
$$XOR(1,0) = 1,$$
$$XOR(1,1) = 0,$$

using the rule $sign(f(x))$, that is, the answer is 1 if $f(x) \geq 0$ and 0 if $f(x) < 0$.

1. Find a suitable set of parameters for this task. A guess is fine, but show that indeed it solves the above task.
2. Will it be possible to represent the XOR function if we replace the max function with the identity function (i.e., $h = U^T x$)? If so, show how. If not, explain why not.

Solution 1 :

A suitable set parameters:

$U = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$ $b_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ $b_2 = -1$

$w = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Then,

$h = max\left(\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} x + \begin{pmatrix} 0 \\ 0 \end{pmatrix}, 0\right)$

By observation:

$f((0,0)) = Sign(w^T h + b_2) =$

$Sign((1,1)(max((0,0) + (0,0), 0)) + (-1))$

$= \text{Sign}(0-1) = \text{Sign}(-1) = 0$ ✓

$f(0,1) = \text{Sign}\left((1,1)\left(\max\left(\begin{pmatrix}1 & -1\\ -1 & 1\end{pmatrix}\begin{pmatrix}0\\1\end{pmatrix} + \begin{pmatrix}0\\0\end{pmatrix}, 0\right)\right.\right.$

$\left.-1\right) = \left((1,1)\left(\max\left(\begin{pmatrix}-1\\1\end{pmatrix}, \begin{pmatrix}0\\0\end{pmatrix}\right)\right)\right) =$

$= \text{Sign}\left((1\ 1)\begin{pmatrix}0\\1\end{pmatrix} - 1\right) = \text{Sign}(0) = 1$ ✓

$f(1,0) = \text{Sign}\left((1,1)\max\left(\begin{pmatrix}1 & -1\\ -1 & 1\end{pmatrix}\begin{pmatrix}1\\0\end{pmatrix} + \begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}0\\0\end{pmatrix}\right) - 1\right)$

$= \text{Sign}\left((1,1)\max\left(\begin{pmatrix}1\\-1\end{pmatrix}, \begin{pmatrix}0\\0\end{pmatrix}\right) - 1\right)$

$= \text{Sign}\left((1,1)\begin{pmatrix}1\\0\end{pmatrix} - 1\right) = \text{Sign}(0)$

$= \text{Sign}(0) = 1$ ✓

$f(1,1) = \text{Sign}\left((1,1)\max\left(\begin{pmatrix}1 & -1\\ -1 & 1\end{pmatrix}\begin{pmatrix}1\\1\end{pmatrix} + \begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}0\\0\end{pmatrix}\right) - 1\right)$

$= \text{Sign}\left((1,1)\max\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}0\\0\end{pmatrix}\right) - 1\right)$

$= \text{Sign}(-1) = 0$ ✓

Solution 2 No ∇

Replacing the max function with the

Replacing the max function with the identity i.e $h = U^T x$ would result in a linear transformation without introducing non-linearity. The XOR function is non-linear.

$$f(x) = w^T(U^T x) + b_2 = (Uw)^T x + b_2$$

$f(x)$ is linear while XOR doesn't.

So we can't replace h with the identity function.