American University of Armenia
Spring 2023
Generative AI (DS 235)
Homework 1 Part2
Tigran Gaplanyan

Task 3: Deriving and Understanding the Sigmoid Function

Sigmoid Function:

The sigmoid function is defined as:

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

To find its derivative with respect to x, we apply the quotient rule of differentiation:

$$\frac{d\sigma(x)}{dx} = \frac{d}{dx}\left(\frac{1}{1 + e^{-x}}\right)$$

Let $f(x) = 1$ and $h(x) = 1 + e^{-x}$. Thus, $f'(x) = 0$ and $h'(x) = -e^{-x}$. Using the quotient rule, we have:

$$\frac{d\sigma(x)}{dx} = \frac{0 \cdot (1 + e^{-x}) - 1 \cdot (-e^{-x})}{(1 + e^{-x})^2}$$

$$\frac{d\sigma(x)}{dx} = \frac{e^{-x}}{(1 + e^{-x})^2}$$

Expressing this derivative in terms of $\sigma(x)$ itself, and since $\sigma(x) = \frac{1}{1+e^{-x}}$, we can rewrite e^{-x} as:

$$e^{-x} = \frac{1 - \sigma(x)}{\sigma(x)}$$

Substituting this back into the derivative, we get:

$$\frac{d\sigma(x)}{dx} = \frac{\frac{1-\sigma(x)}{\sigma(x)}}{(1 + \frac{1-\sigma(x)}{\sigma(x)})^2}$$

Simplifying this expression, we obtain the derivative of the sigmoid function in terms of $\sigma(x)$ itself:

$$\frac{d\sigma(x)}{dx} = \sigma(x) \cdot (1 - \sigma(x))$$

This demonstrates the special property of the sigmoid function where its derivative can be expressed in terms of the function itself.

Task 4: Connecting Sigmoid and Softmax Functions

Softmax Function for Two Classes

In a binary classification problem, the softmax function for an arbitrary vector $\mathbf{z}$ with components z_1 and z_2 for two classes can be expressed as:

$$\text{softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{2} e^{z_j}}$$

This yields the probabilities for each class:

- Probability of class 1: $P(y = 1) = \frac{e^{z_1}}{e^{z_1}+e^{z_2}}$
- Probability of class 2: $P(y = 2) = \frac{e^{z_2}}{e^{z_1}+e^{z_2}}$

Derivation of the Sigmoid Function from Softmax

To derive the sigmoid function from the softmax function for the probability of the first class, consider $P(y = 1)$:

$$P(y = 1) = \frac{e^{z_1}}{e^{z_1} + e^{z_2}}$$

In binary classification, by setting $z_2 = 0$ or considering $z = z_1 - z_2$ and setting $z_2 = 0$, the expression simplifies to:

$$P(y = 1) = \frac{e^z}{e^z + 1}$$

Multiplying the numerator and the denominator by e^{-z} simplifies this to the sigmoid function:

$$P(y = 1) = \frac{1}{1 + e^{-z}}$$

This shows that the softmax function simplifies to the sigmoid function $\sigma(z)$ in the context of binary classification, demonstrating the connection between the two functions.

Task 5: Understanding Logits and Log Odds

Inverse of the Sigmoid Function

Let's derive the inverse of the sigmoid function, $S(x)$:

$$\begin{aligned} y = S(x) &= \frac{1}{1 + e^{-x}} \\ y(1 + e^{-x}) &= 1 \\ e^{-x} &= \frac{1}{y} - 1 \\ x &= -\ln\left(\frac{1}{y} - 1\right) \end{aligned}$$

Simplifying further, we express $\frac{1}{y} - 1$ as $\frac{1-y}{y}$, obtaining the logit function:

$$x = \ln\left(\frac{y}{1 - y}\right)$$

Therefore, the inverse of the sigmoid function, or the logit function, is given by:

$$\sigma^{-1}(y) = \ln\left(\frac{y}{1 - y}\right)$$

This function takes a probability y and maps it to the log odds, which is the natural logarithm of the odds ratio $\frac{y}{1-y}$.

Task 6: Understanding Backpropagation and the Chain Rul

Computing Partial Derivatives

Consider a neural network with a single neuron that takes two inputs x_1 and x_2, with weights w_1 and w_2 respectively, and a bias b. The output of the neuron is passed through a hyperbolic tangent activation function, defined as $F = \tanh(z)$, where $z = w_1 x_1 + w_2 x_2 + b$.

The hyperbolic tangent function is given by:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

To compute the partial derivatives of the function F with respect to w_1 and w_2, we apply the chain rule.

Partial Derivative with respect to w_1

The partial derivative of F with respect to w_1 is given by:

$$\frac{\partial F}{\partial w_1} = \frac{\partial F}{\partial z} \cdot \frac{\partial z}{\partial w_1}$$

First, we find $\frac{\partial F}{\partial z}$, the derivative of $\tanh(z)$ with respect to z:

$$\frac{\partial F}{\partial z} = 1 - \tanh^2(z)$$

Then, we compute $\frac{\partial z}{\partial w_1}$:

$$\frac{\partial z}{\partial w_1} = x_1$$

Combining these, we get:

$$\frac{\partial F}{\partial w_1} = (1 - \tanh^2(z)) \cdot x_1$$

Partial Derivative with respect to w_2

Similarly, the partial derivative of F with respect to w_2 is given by:

$$\frac{\partial F}{\partial w_2} = \frac{\partial F}{\partial z} \cdot \frac{\partial z}{\partial w_2}$$

With $\frac{\partial F}{\partial z}$ already known, and $\frac{\partial z}{\partial w_2} = x_2$, we have:

$$\frac{\partial F}{\partial w_2} = (1 - \tanh^2(z)) \cdot x_2$$