Advanced Math Notes

Yuchen Wang

February 12, 2019

Contents

1 Free parameter

A variable in a mathematical model which cannot be predicted precisely or constrained by the model and must be estimated experimentally or theoretically.

2 Rectifier

2.1 Definition

An activation function defined as the positive part of its argument:

$$f(x) = \max(0, x)$$

Also known as: ramp function
A unit employing the rectifier is also called a **rectified linear unit (ReLU)**

2.2 Softplus

A smooth approximation to the rectifier is the analytic function

$$f(x) = \log(1 + e^x)$$

Also known as: SmoothReLU
The derivative of softplus is

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

(the logistic function)

Notes The logistic function is a smooth approximation of the derivative of the rectifier, the **Heaviside step function**

2.3 Multivariable Generalization to Softplus

LogSumExp with the first argument set to zero

$$LSE_0^+(x_1, \ldots, x_n) := LSE(0, x_1, \ldots, x_n) = \log(1 + e^{x_1} + \ldots + e^{x_n})$$

Notes   The LogSumExp function itself is:

$$LSE(x_1, \ldots, x_n) = \log(e^{x_1} + \ldots + e^{x_n})$$

and its gradient is the softmax.

The softmax with the first argument set to zero is the multivariable generalization of the logistic function.

3   Softmax Function

The softmax function takes an un-normalized vector, and normalizes it into a probability distribution. That is, prior to applying softmax, some vector elements could be negative, or greater than one; and might not sum to 1; but after applying softmax, each element x_i is in the interval $[0, 1]$, and $\sum_i x_i = 1$

$$\sigma : \mathbb{R}^K \to \{\sigma \in \mathbb{R}^K | \sigma_i > 0, \sum_{i=1}^{K} \sigma_i = 1\}$$

$$\sigma(\mathbf{z})_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}}$$

for $j = 1, \ldots, K$