Common Mathematical Tools

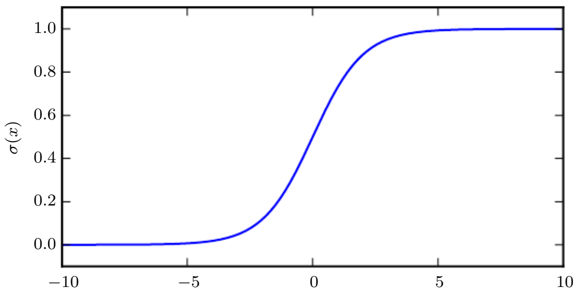
CHEN Si

Contents

Part I
Common Functions

1 Logistic Sigmoid

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



Formulas

1. $\sigma(x) = \dfrac{\exp(x)}{\exp(x) + \exp(0)}$
2. $\dfrac{\mathrm{d}}{\mathrm{d}x}\sigma(x) = \sigma(x)(1 - \sigma(x))$
3. $1 - \sigma(x) = \sigma(-x)$
4. $\log \sigma(x) = -\zeta(-x)$
5. $\sigma(x) = \dfrac{\mathrm{d}}{\mathrm{d}x}\zeta(x)$
6. $\forall x \in (0,1),\ \sigma^{-1}(x) = \log\left(\dfrac{x}{1-x}\right)$
   - The function $\sigma^{-1}(x)$ is called the **logit** in statistics.

Properties

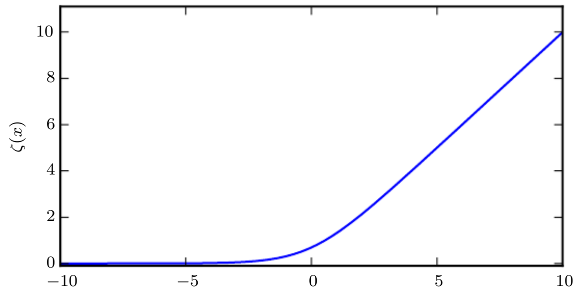
1. **Saturates** when its argument is very positive or very negative.
   - very flat
   - insensitive to small changes in its input

Applications

1. Produce the parameter of a Bernoulli distribution.

2 Softplus Function

$$\zeta(x) = \log(1 + \exp(x))$$



Formulas

1. $-\zeta(-x) = \log \sigma(x)$
2. $\frac{\mathrm{d}}{\mathrm{d}x}\zeta(x) = \sigma(x)$
3. $\forall x > 0,\ \zeta^{-1}(x) = \log(\exp(x) - 1)$
4. $\zeta(x) = \int_{-\infty}^{x} \sigma(y)\mathrm{d}y$
5. $\zeta(x) - \zeta(-x) = x$

Properties

1. $\zeta(x) \in (0, \infty)$
2. $\zeta(x)$ is a smoothed (or "softened") version of

$$x^+ = \max(0, x)$$

$\zeta(-x)$ is a smoothed (or "softened") version of

$$x^- = \max(0, -x)$$

Similar to $x^+ - x^- = x$, we also have

$$\zeta(x) - \zeta(-x) = x$$

Applications

1. Produces the β or σ parameter of a normal distribution.

Part II
Common Probability Distributions

3 Gaussian Distribution

Introduction

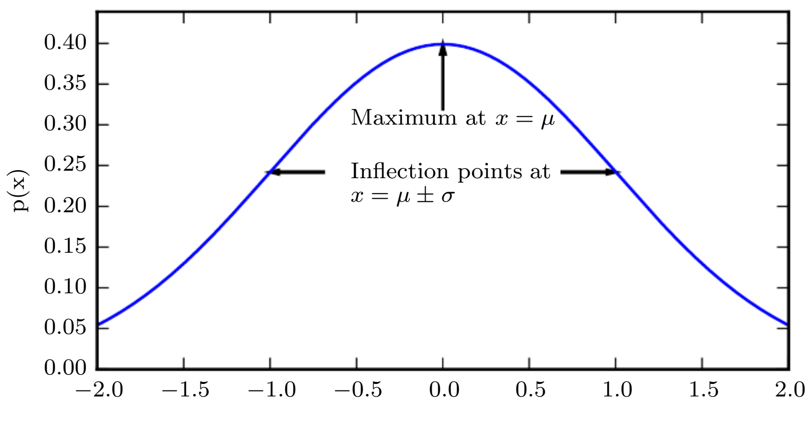
- Also known as the **Normal Distribution**.
- The most commonly used distribution over real numbers.

Definition

For real numbers

$$\mathcal{N}(x;\mu,\sigma^2) = \sqrt{\frac{1}{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$$

- $\mu \in \mathbb{R}$
- $\sigma \in (0,\infty)$



For $\mathbb{R}^n$

$$\mathcal{N}(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma}) = \sqrt{\frac{1}{(2\pi)^n\det(\boldsymbol{\Sigma})}}\exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right)$$

- Also known as **Multivariate Normal Distribution**.
- The covariance matrix $\boldsymbol{\Sigma}$ is positive definite symmetric.

Variant

For real numbers If we need to frequently evaluate the PDF with diffrent parameter values, a more efficient way of parametrizing the distribution is to use a parameter $\beta \in (0, \infty)$ to control the **precision** (or inverse variance) of the distribution

$$\mathcal{N}(x; \mu, \beta^{-1}) = \sqrt{\frac{\beta}{2\pi}} \exp\left(-\frac{1}{2}\beta(x-\mu)^2\right)$$

For $\mathbb{R}^n$ We can use a **precision matrix $\boldsymbol{\beta}$**

$$\mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\beta}^{-1}) = \sqrt{\frac{\det(\boldsymbol{\beta})}{(2\pi)^n}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\beta}(\boldsymbol{x}-\boldsymbol{\mu})\right)$$

We often fix the covariance matrix to be a diagonal matrix. An even simpler version is the **isotropic** Gaussian distribution, whose covariance matrix is a scalar times the identiy matrix.

Properties

- $\mathbb{E}[x] = \mu$

Advantages

1. The **central limit theorem** shows that the sum of many independent random variables is approximately normally distributed.

2. Out of all possible probability distributions with the same variance, the normal distribution encodes the maximum amount of uncertainty over the real numbers.

    - We can thus think of the normal distribution as being the one that inserts the least amount of prior knowledge into a model.