Math basis for Machine Leanring

Qi Wang

December 3, 2021

Linear Algebra

Jacobian Matrix

Given a function of mapping a n-dimensional input vector to a m-dimensional output vector, $f : \mathbb{R}_n \mapsto \mathbb{R}^m$, the matrix of the first-order partial derivatives is called **Jacobian Matrix**,**J**:

$$J = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

Probability

Expectation

$$\mathbb{E}_{x \sim p_r(x)} f(x) = \int_x p_r(x) f(x) dx$$

where $x \sim p_r(x)$

Reparameterization trick

Assume we have a normal distribution q that is parameterized by θ, specifically $q_\theta(x) = N(\theta, I)$. And we want solve the following:

$$min_\theta \mathbb{E}_q[x_2]$$

By calculating its gradient:

$$\begin{aligned} \nabla_\theta \mathbb{E}_q[x^2] &= \nabla_\theta \int q_\theta(x) x^2 dx \\ &= \mathbb{E}_q[x^2 \nabla_\theta log q_\theta(x)] \end{aligned}$$

For example, we give $q_\theta(x) = N(\theta, I)$, this changes to:

$$\nabla_\theta \mathbb{E}_q[x^2] = \mathbb{E}_q[x^2(x - \theta)]$$

Now comes the trick to Reparameterize, so that the gradient will be independent of θ:

$$x = \theta + \epsilon, \epsilon \sim N(0, I)$$

Then rewrite the equation as:

$$\mathbb{E}_p[x^2] = \mathbb{E}[(\theta + \epsilon)^2]$$

Where p is the distribution of ϵ, i.e. $N(0, I)$. Again, rewriting:

$$\nabla_\theta \mathbb{E}_q x^2 = \mathbb{E}_q[2(\theta + \epsilon)]$$