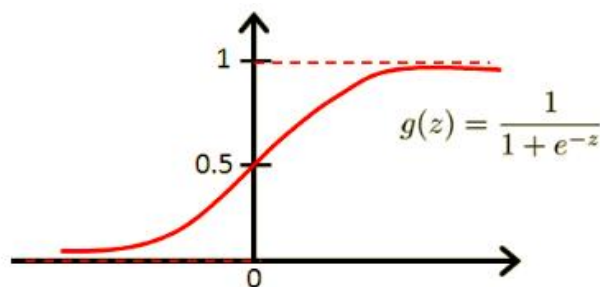


逻辑回归

Logistic Regression

逻辑回归(logistic regression)是一种常用的有监督机器学习算法，因其实现简单，所以应用广泛。



逻辑回归是在线性函数 $\theta^T \mathbf{x}$ 输出预测值的基础上，寻找一个激活函数 $h_{\theta}(\mathbf{x}) = g(\theta^T \mathbf{x})$ ，将实际值映射到 0~1 之间。选择 Sigmoid 函数为激活函数，则

$$h_{\theta}(\mathbf{x}) = \frac{1}{1 + e^{-\theta^T \mathbf{x}}}$$

联邦逻辑回归的目标函数是

$$L = -\frac{1}{N} \sum_{i=1}^n (y_i \log h_{\theta}(x_i) + (1 - y_i) \log (1 - h_{\theta}(x_i)))$$

这里的 $y_i = 0$ 或 1 。在实际计算过程中，参与方 P_1 和 P_2 的模型参数迭代更新策略是

$$\theta^{(1)} \leftarrow \theta^{(1)} - \eta \nabla_{\theta^{(1)}} L$$

$$\theta^{(2)} \leftarrow \theta^{(2)} - \eta \nabla_{\theta^{(2)}} L$$

由于 $\theta^T \mathbf{x} = \theta^{(1)T} x_i^{(1)} + \theta^{(2)T} x_i^{(2)} = u_i^{(1)} + u_i^{(2)}$ ， $\frac{\partial u_i^{(1)}}{\partial \theta^{(1)}} = x_i^{(1)}$ ， $\frac{\partial u_i^{(2)}}{\partial \theta^{(2)}} = x_i^{(2)}$ ，所以梯度

$\nabla_{\theta^{(1)}} L$ 和 $\nabla_{\theta^{(2)}} L$ 可以用下式计算

$$\nabla_{\theta^{(1)}} = \frac{\partial L}{\partial \theta^{(1)}} = -\frac{1}{N} \sum_{i=1}^N (y_i - h_{\theta}(x_i)) x_i^{(1)}$$

$$\nabla_{\theta^{(2)}} = \frac{\partial L}{\partial \theta^{(2)}} = -\frac{1}{N} \sum_{i=1}^N (y_i - h_{\theta}(x_i)) x_i^{(2)}$$

逻辑回归由于算法的简单和高效，在实际中应用非常广泛，常用于数据挖掘，疾病自动诊断，经济预测等领域。