

损失函数 是单个样例的损失, 公式为

$$L(Y, f(x)) = (Y - f(x))^2$$

代价函数 是所有数据的损失, 公式为

$$J(w, b) = \frac{1}{m} \sum_{i=1}^m L(y_i, f(x))$$

我们要求出代价函数最小的时候的 w 和 b 。但并不是越小越好, 因为有时候这样会过度拟合给定的训练数据, 造成函数过于复杂。我们不仅要让 cost function 最小化, 还要让结构风险最小化, 这时候就定义了一个函数 $J(f)$, 这个函数专门来度量模型的复杂度。这在机器学习中也叫做正则化, 常用的有 L_1 L_2

所以最终的优化函数或者说目标函数其实是

$$\min \frac{1}{m} \sum_{i=1}^m L(y_i, f(x)) + \lambda J(f)$$