

## 1 线性回归

### 1.1 单变量线性拟合

#### 1.1.1 机器学习的一般过程

#### 1.1.2 误差函数与线性回归

定义误差函数(**Error Function**)(或称代价函数(**Cost Function**))衡量预测值与实际值之间的差距.机器学习中常用的误差函数为平方和误差函数(**Sum-of-Square Error Function**),即二乘误差函数:

$$E(\mathbf{w}) = \frac{1}{2} \sum_{i=1}^N (f(x_i; \mathbf{w}) - t_i)^2$$

其中 $\mathbf{w}$ 为权重矢量(**Wight Vector**),是拟合多项式的参数构成的矢量.例如,多项式拟合的 $\mathbf{w} = (w_0, \dots, w_M)$ 表明拟合函数 $f$ 为

$$f(x; \mathbf{w}) = w_0 + w_1x + \dots + w_Mx^M = \sum_{j=0}^M w_jx^j$$

各 $x_i$ 为容易看出这与统计学中最小二乘法的关系.

考虑测试数据 $(x_1, y_1), \dots, (x_n, y_n)$ ,使用一次函数模型 $y = f(x) = ax + b$ 进行拟合.

#### 1.1.3 过拟合,维度灾难与正则化