Author：Liu Jian

Time：2020-06-03

概要

应用：

1. 数据挖掘；
2. 计算机视觉；
3. 自然语言处理；
4. 机器决策（无人驾驶等）。

解决的问题：

1. 回归问题；
2. 分类问题。

概念解释：

1. 归纳偏好，可以理解为几个模型都可以时，根据我们的偏好对此进行选择。比如，奥卡姆剃刀会选择模型最“简单”的那一个。此外，还有多释原则，(principle of multiple explanations)，即主张保留与经验观察一致的所有模型，这与集成学习 (ensemble learning) 的理念相吻合。
2. No free lunch theorem (NFL) 告诉我们：脱离具体问题，空泛地谈论模型的好坏毫无意义，因为若考虑所有潜在的问题 (针对所有问题求模型性能的期望)，则所有模型都一样好。我们可以这样理解，对于模型 A ，若它在某些问题 (记问题的真实解为 S_1) 上比模型 B 好 (也就是 A 与 S_1 更贴近)，则必然存在另一些问题 (记问题的真实解为 S_2)，使得模型 B 比模型 A 好 (也就是 B 与 S_2 更贴近)，不存在一个模型能解决所有的问题 (即不存在一条具体的曲线对所有的解曲线都拟合得很好)，应该具体问题具体分析 (即对于具体问题求解相应的合适的拟合曲线)。

> 所谓“没有免费的午餐”，如果不假定观察到的数据和未来的数据之间有一定的联系的话，这个任务是无法完成的。而在统计学习中，建立两者之间联系是通过一个共享的概率模型来实现的。-- pluskid

常用知识点

- **直线法向量与方向向量：**

直线：$Ax+By+C=0\;(y=wx+b)$

点 (x_1,y_1) , (x_2,y_2) 在直线上，相减得方向向量。

方向向量 $\langle x_1-x_2, y_1-y_2\rangle \implies \langle B,-A\rangle$ 或 $\langle 1,w\rangle$ 。

法向量与方向向量垂直，可得法向量 $\langle A,B\rangle$ 。

- **矩阵求微分常用公式：**

$$\frac{\partial \boldsymbol{x}^T \boldsymbol{a}}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{a}^T \boldsymbol{x}}{\partial \boldsymbol{x}} = \boldsymbol{a}$$

$$\frac{\partial \boldsymbol{x}^T B \boldsymbol{x}}{\partial \boldsymbol{x}} = (B + B^T)\boldsymbol{x}$$

- **概率分布常用公式：**

若有分布：

$$P(y) = \begin{cases} p, & y = 1 \\ 1 - p, & y = 0 \end{cases}$$

则概率分布可以写成如下等价的形式：

$$P(y) = p^y (1-p)^{(1-y)} \quad \text{或} \quad P(y) = yp + (1-y)(1-p)$$

在实际应用时，根据问题选用相应的形式。

- **二元函数凹凸性及最值定理：**
    - 凹凸性：设$f(x,y)$在区域$\mathcal{D}$上具有二阶连续偏导数，记 $A = f''_{xx}(x,y), B = f''_{xy}(x,y), C = f''_{yy}(x,y)$则：

      (1) 在$\mathcal{D}$上恒有$A > 0$，且$AC - B^2 \geqslant 0$时，$f(x,y)$在区域$\mathcal{D}$上是凸函数；

      (2) 在$\mathcal{D}$上恒有$A < 0$，且$AC - B^2 \geqslant 0$时，$f(x,y)$在区域$\mathcal{D}$上是凹函数；
    - 最值：设$f(x,y)$是在开区域$\mathcal{D}$内具有连续偏导数的凸(或者凹)函数，$(x_0, y_0) \in \mathcal{D}$且 $f'_x(x_0,y_0) = 0, f'_y(x_0,y_0) = 0$，则$f(x_0,y_0)$为$f(x,y)$在$\mathcal{D}$内的最小值(或最大值)。

- **多元实值函数凹凸性及最值：**
    - 多元实值函数凹凸性判定定理：设$\mathcal{D} \in R^n$是非空开凸集，$f: \mathcal{D} \subset R^n \to R$，且$f(x)$在$\mathcal{D}$上二阶连续可微，如果$f(x)$的Hessian矩阵$\nabla^2 f(x)$在$\mathcal{D}$上是正定的，则$f(x)$是$\mathcal{D}$上的严格凸函数。
    - 凸充分性定理：若$f: R^n \to R$是凸函数，且$f(x)$一阶连续可微，则x^*是最小值的充要条件为$\nabla f(x^*) = 0$.

机器学习1-线性回归

内容： 一元线性回归；多元线性回归；广义线性模型；对数几率回归（逻辑斯谛回归）

1 一元线性回归

模型： $f(x) = wx + b$, s.t. $f(x_i) \simeq y_i$

参数估计方法： 最小化平方损失函数/最小二乘法

参数计算公式：

$$w = \frac{\boldsymbol{x}_d^T \boldsymbol{y}_d}{\boldsymbol{x}_d^T \boldsymbol{x}_d} b = \overline{y} - w\overline{x}$$

其中，$\boldsymbol{x}_d = \langle x_1 - \overline{x}, x_2 - \overline{x}, \ldots, x_m - \overline{x} \rangle^T$，$\boldsymbol{y}_d = \langle y_1 - \overline{y}, x_2 - \overline{x}, \ldots, x_m - \overline{x} \rangle^T$。平均值 $\overline{x} = \frac{1}{m}\sum_{i=1}^m x_i$，$\overline{y} = \frac{1}{m}\sum_{i=1}^m y_i$。

2 多元线性回归

模型： $f(\boldsymbol{x}) = \boldsymbol{w}^T \boldsymbol{x} + \boldsymbol{b}$, s.t. $f(\boldsymbol{x}_i) \simeq y_i$

参数估计方法： 最小化平方损失函数/最小二乘法

参数计算公式： 见西瓜书P55