

· 概序:

频率派 \rightarrow 统计机器学习 \rightarrow 模型优化. 模型
- θ : 常量 - loss function
Algorithm
- $\theta_{MLE} = \arg \log P(X|\theta)$

贝叶斯派 \rightarrow 概率图模型 \rightarrow 求积分.

θ : r.v.

$$P(\theta|X) = \frac{P(\theta) \cdot P(X|\theta)}{P(X)}$$

· 假设偏好问题: 没有免费的午餐
假设与实际问题的 (NFL)

$$\theta_{MAP} = \arg P(\theta|X) = \arg (P(\theta) \cdot P(X|\theta))$$

· 模型评估.

\Rightarrow 先复习概率论!

· 比较检验

· 周志华《机器学习》的最大优点在于本书贯穿了“西瓜”的例子.
这也是学习一个抽象概念的最好方法.

- 8.11

· 高斯分布

$$X \sim (\mu, \sigma^2)$$

· 一维高斯:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$$E(\mu_{MLE}) = \mu$$

$$\mu_{MLE} = \frac{1}{N} \sum_{i=1}^N x_i \rightarrow \text{无偏估计}$$

$$\sigma^2_{MLE} = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2 \rightarrow \text{有偏估计}$$

$$\hookrightarrow E(\sigma^2_{MLE}) = \left[\frac{n-1}{n}\right] \sigma^2$$

PDF:

$$\cdot \text{多维高斯分布: } p(x) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2} (x-\mu)^T \Sigma^{-1} (x-\mu)\right)$$

$$x \in \mathbb{R}^p$$

二次型 \hookrightarrow 马氏距离: x 与 μ

Σ : $p \times p$ 半正定 \rightarrow 正定

$$\cdot (x-\mu)^T \Sigma^{-1} (x-\mu):$$

$$\Sigma \text{ 正定} = U \Lambda U^T \quad \Lambda = \text{diag}(\lambda_i) \quad U = (u_1, u_2, \dots, u_p)_{p \times p}$$

$$\Delta = \sum_{i=1}^p \frac{y_i^2}{\lambda_i}$$

$$y_i = (x-\mu)^T u_i$$

$$\text{当 } \Delta = c \Rightarrow \frac{y_1^2}{c\lambda_1} + \frac{y_2^2}{c\lambda_2} = 1 \quad \text{椭圆 (x与y关于u映射)}$$

$$\therefore p(x) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \cdot \exp\left(-\frac{1}{2} \Delta\right) \quad 0 \leq p(x) \leq 1 \Rightarrow \text{等高线取椭圆}$$

· 局限性

· 参数复杂

$$\text{已知 } x = \begin{pmatrix} x_a \\ x_b \end{pmatrix} \begin{matrix} \rightarrow m \\ \rightarrow n \end{matrix} \quad x \in \mathbb{R}^p \quad \mu = \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix} \quad \Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}$$

· 边缘概率

线性回归:

· 最小二乘法.

$$X = \begin{pmatrix} x_1 & x_2 & \dots & x_i & \dots & x_n \end{pmatrix}^T \quad x_i \in \mathbb{R}^p$$

$n \times p$

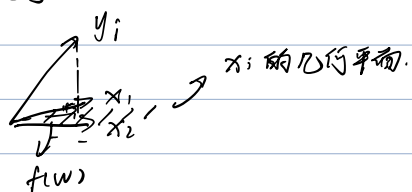
$$\mathcal{L}(w) = \sum \|w_i^T x_i - y_i\|^2 \quad f(w) = w^T x$$

$$w = \arg \min \mathcal{L}(w) \quad \frac{\partial \mathcal{L}(w)}{\partial w} = 0$$

$$\Rightarrow w = \underbrace{(X^T X)^{-1}}_{p \times n \quad n \times p \quad p \times n} X^T Y \quad \begin{matrix} p \times 1 \\ N \times 1 \end{matrix}$$

$p \times n: X^T$

几何角度: y_i 关于 x_i 的几何平面的投影: w



$$\Rightarrow x_i^T \cdot (y_i - f(w)) = 0 \quad (0 \text{ 矩阵})$$

$$X^T (Y - w^T X) = 0$$

• 决策树. \rightarrow 与条件概率有什么关系

• 组成

- 叶结点 (根结点)

-

· 神经网络

· 支持向量机 ~~书~~

✱ 能不能快速定位?

Ipod 屏反光问题

-

$$\sum_i f(x, y_i) = f(x, y)$$

$$f(x) = \sum x$$

$$x_i \quad \frac{\partial f(x)}{\partial x} = \frac{\partial \sum x}{\partial x}$$