

§7 贝叶斯分类器. BM

1. Bayes

$$p(c|x) = \frac{p(x,c)}{p(x)} = \frac{p(c)p(x|c)}{p(x)}$$

2. 极大似然估计 MLE

MLE估计过程: 编写似然函数 \rightarrow 对似然函数取对数 \rightarrow 求导 \rightarrow 求解

$$L(x; \theta) \Rightarrow \ell(\theta) = \ln L(x; \theta) \rightarrow \frac{\partial \ell}{\partial \theta} \quad \arg \max_p p(x_i; p) = \arg \max_p \ln(p(x_i; p))$$

参数未知, 但都是客观存在同一固定值, 通过优化似然函数来求解参数

区别: 最大后验概率估计 MAP 求使 $p(x|\theta)p(\theta)$ 值最大

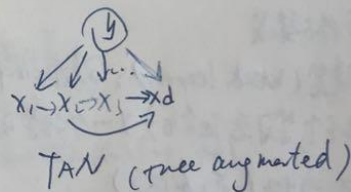
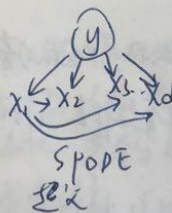
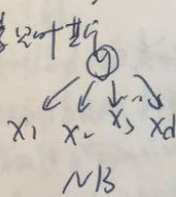
$$p(\theta|x) = \frac{p(\theta)p(x|\theta)}{p(x)} \rightarrow \arg \max_{\theta'} p(\theta'|x) \rightarrow \arg \max_{\theta'} p(\theta')p(x|\theta')$$

3. 朴素贝叶斯

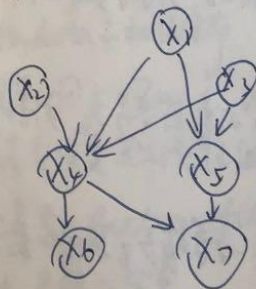
假设各特征独立性的假设 $p(c|x) = \frac{p(c)p(x|c)}{p(x)} = \frac{p(c)}{p(x)} \prod_{i=1}^d p(x_i|c)$

$$h_{nb}(x) = \arg \max_{c \in Y} p(c) \prod_{i=1}^d p(x_i|c)$$

朴素贝叶斯图



4. 贝叶斯网. 有向无环图模型 (directed acyclic graphical model, DAG)



$$p(x_1, x_2, x_3, x_4, x_5, x_6, x_7) = p(x_1)p(x_2|x_1)p(x_3|x_1)p(x_4|x_1, x_2)p(x_5|x_1, x_3)p(x_6|x_4)p(x_7|x_5)$$

$$p(x_5|x_1, x_3)p(x_6|x_4)p(x_7|x_5, x_4)$$

5 EM 算法

latent variable $L(\theta|X) = \ln p(x|\theta) = \ln \sum_z p(x, z|\theta)$

EM 基本思路: 若 θ 已知, 则可依据训练数据推断出最优隐变量 z 之值 (E);

反之若 z 已知, 则可对参数 θ 做极大似然估计 (M)

E (expectation): $Q(\theta|\theta^t) = E_{z|x, \theta^t} L(\theta|x, z)$

M (Maximization): $\theta^{t+1} = \arg \max_{\theta} Q(\theta|\theta^t)$

E 、 M 交替 M 步得到参数值用于 E 步, 直到收敛到局部最优解

GMM 高斯混合模型, 由多个高斯模型线性叠加混合而成

$$p(x) = \sum_{k=1}^K p(k) p(x|k) = \sum_{k=1}^K \pi_k p(x; \mu_k, \Sigma_k)$$

模型评价: AIC, BIC 越小越好