作业2

图灵2201 张祎迪

3.2. 利用例题 3.2 的构造的 kd 树求点 $x=(3,4.5)^T$ 的最近邻点

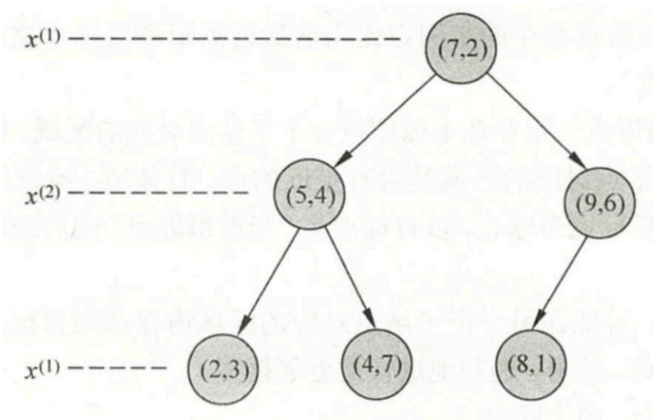

图 3.4 *kd* 树示例

> 注：使用欧氏距离 $L_2(x_i,x_j)=(\sum_{l=1}^{n}|x_i^{(l)-}x_j^{(l)}|^2)^{\frac{1}{2}}$

- 在kd 树中找出包含目标点x 的叶结点：$(7,2)\rightarrow(5,4)\rightarrow(4,7)$

  得到搜索路径$<(7,2),(5,4),(4,7)>$
- 以$(4,7)$为“当前最近点”(其到查询点的距离为2.69)，递归地向上回退
  - 回溯到$(5,4)$ 其到查询点距离为 2.06，将其更新为最近点.
  - 以 $(3,4.5)$ 为圆心，以 2.06 为半径画圆，圆和超平面 $y=4$ 交割，进入 $(5,4)$ 结点的左子空间搜索.
  - 左子空间找到$(2,3)$ 到查询点距离为 1.8，将其更新为最近点.
  - 回溯到$(7,2)$,以 $(3,4.5)$ 为圆心 以 1.8为半径画圆 ，不与$x=7$的超平面相交，所以不用进入其右子空间进行查找
- 结束搜索，得到最近邻点$(2,3)$,最近距离1.8

4.1. 用极大似然估计法推出朴素贝叶斯法中的概率估计公式 (4.8) 及公式 (4.9)。

- $P(Y=c_k)=\frac{\sum_{i=1}^{N}I(y_i=c_k)}{N}$ (4.8)

  设$P(Y=c_k)=p$ 似然函数为 $L(p)=C_N^n p^n(1-p)^{N-n}$ 其中 n 为c_k 在 Y 中出现次数

  $\frac{\partial lnL(p)}{\partial p}=0\Rightarrow p=\frac{n}{N}$

  $\therefore$ 先验概率 $P(Y=c_k)$ 的极大似然估计是 $P(Y=c_k)=p=\frac{\sum_{i=1}^{N}I(y_i=c_k)}{N}$
- $P(X^j=a_{jl}|Y=c_k)=\frac{\sum_{i=1}^{N}I(x_i^j=a_{jl},y_i=c_k)}{\sum_{i=1}^{N}I(y_i=c_k)}$ (4.9)

  设 $P(X^j=a_{jl}|Y=c_k)=p$

  $\because$ 在条件 $Y=c_k$ 下，随机变量满足条件独立性

  $\therefore L(p)=C_n^m p^m(1-p)^{n-m}$ 其中 $n=\sum_{i=1}^{N}I(y_i=c_k)$ $m=\sum_{i=1}^{N}I(x_i^j=a_{jl},y_i=c_k)$

  $\frac{\partial lnL(p)}{\partial p}=0\Rightarrow p=\frac{m}{n}$

$\therefore P(X^j=a_{jl}|Y=c_k)$ 的极大似然估计是 $p=\frac{\sum_{i=1}^N I(x_i^j=a_{jl},y_i=c_k)}{\sum_{i=1}^N I(y_i=c_k)}$

4.2

贝叶斯估计的一般步骤

1确定参数θ的先验概率$p(\theta)$

2.根据样本集$D=x_1,x_2,\ldots,x_n$，计算似然函数$P(D|\theta)$：$P(D|\theta)=\prod\limits_{i=1}^n P(x_n|D)$

3.利用贝叶斯公式，求θ的后验概率：$P(\theta|D)=\dfrac{P(D|\theta)P(\theta)}{\int\limits_\Theta P(D|\theta)P(\theta)d\theta}$

4.计算后验概率分布参数θ的期望，并求出贝叶斯估计值：$\hat{\theta}=\int\limits_\Theta \theta\cdot P(\theta|D)d\theta$

证明（4.11）

$P(Y=c_k)=\frac{\sum_{i=1}^N I(y_i=c_k)+\lambda}{N+K\lambda}\quad(4.11)$

- 条件假设：$P_\lambda(Y=c_k)=u_k$，且服从参数为λ的$Dirichlet$分布；随机变量Y出现$y=c_k$的次数为m_k 即 $m_k=\sum\limits_{i=1}^N I(y_i=c_k)$，可知$\sum\limits_{k=1}^K m_k=N$；

> 参考：https://github.com/datawhalechina/statistical-learning-method-solutions-manual/blob/master/docs/chapter04/ch04.md
>
> 1. 狄利克雷($Dirichlet$)分布
>   参考PRML（Pattern Recognition and Machine Learning）一书的第2.2.1章节：用似然函数(2.34)乘以先验(2.38)，我们得到了参数u_k的后验分布，形式为 $p(u|D,\alpha)\propto p(D|u)p(u|\alpha)\propto\prod_{k=1}^K u_k^{\alpha_k+m_k-1}$
>   该书中第B.4章节：狄利克雷分布是K个随机变量$0\leqslant u_k\leqslant 1$的多变量分布，其中$k=1,2,\ldots,K$，并满足以下约束 $0\leqslant u_k\leqslant 1,\quad \sum_{k=1}^K u_k=1$ 记$u=(u_1,\ldots,u_K)^T,\alpha=(\alpha_1,\ldots,\alpha_K)^T$，有
> $\mathrm{Dir}(u|\alpha)=C(\alpha)\prod_{k-1}^K u_k^{\alpha_k-1}\quad E(u_k)=\frac{\alpha_k}{\sum\limits_{k=1}^K\alpha_k}$
> 2. 为什么假设$Y=c_k$的概率服从$Dirichlet$分布？
>   - 首先，根据PRML第B.4章节，$Dirichlet$分布是$Beta$分布的推广。
>   - 由于，$Beta$分布是二项式分布的共轭分布，$Dirichlet$分布是多项式分布的共轭分布。$Dirichlet$分布可以看作是“分布的分布”；
>   - 又因为，$Beta$分布与$Dirichlet$分布都是先验共轭的，意味着先验概率和后验概率属于同一个分布。当假设为$Beta$分布或者$Dirichlet$分布时，通过获得大量的观测数据，进行数据分布的调整，使得计算出来的概率越来越接近真实值。
>   - 因此，对于一个概率未知的事件，$Beta$分布或$Dirichlet$分布能作为表示该事件发生的概率的概率分布。

这里做先验概率退化为均匀分布时的简单证明

- 先验概率： $pK=1$
- 似然函数： $P(Y=c_k)=p=\frac{\sum_{i=1}^N I(y_i=c_k)}{N}$ (4.8)
- 利用拉格朗日数乘法，有$\lambda(pK-1)+pN-\sum_{i=1}^N I(y_i=c_k)=0$
- 得到$P(Y=c_k)=\frac{\sum_{i=1}^N I(y_i=c_k)+\lambda}{N+K\lambda}$

证明（4.10）同理

$P_\lambda(X^j=a_{jl}|Y=c_k)=\frac{\sum_{i=1}^N I(x_i^j=a_{jl},y_i=c_k)+\lambda}{\sum_{i=1}^N I(y_i=c_k)+S_j\lambda}(4.10)$

- 先验概率： $pK=1$
- 似然函数:$P(X^j=a_{jl}|Y=c_k)=\frac{\sum_{i=1}^N I(x_i^j=a_{jl},y_i=c_k)}{\sum_{i=1}^N I(y_i=c_k)}$ (4.9)
- 利用拉格朗日数乘法同**(4.11)**的证明可以得到

  $P_\lambda(X^j=a_{jl}|Y=c_k)=\frac{\sum_{i=1}^N I(x_i^j=a_{jl},y_i=c_k)+\lambda}{\sum_{i=1}^N I(y_i=c_k)+S_j\lambda}(4.10)$

> 更完整的证明参见 https://github.com/datawhalechina/statistical-learning-method-solutions-manual/blob/master/docs/chapter04/ch04.md