1 Hw 3.1

设 x 为一个 d 维的二值向量（即其分量取值为 0 或 1），服从多维伯努利分布

$$p(x_i|\Theta) = \prod_{i=1}^{d} \Theta_i^{x_i}(1-\Theta_i)^{1-x_i} \tag{1}$$

其中 $\Theta = (\Theta_1, \cdots, \Theta_d)^T$ 是未知的参数向量，而 Θ_i 为 $x_i = 1$ 的概率。证明，对于 Θ 的最大似然估计为

$$\widehat{\Theta} = \frac{1}{n}\sum_{k=1}^{n} X_k \tag{2}$$

解：由已知条件可得最大似然函数为：

$$p(X|\Theta) = \prod_{k=1}^{n}\prod_{i=1}^{d} \Theta_i^{x_{ki}}(1-\Theta_i)^{1-x_{ki}} \tag{3}$$

即

$$p(X|\Theta) = \prod_{i=1}^{d} \Theta_i^{\sum\limits_{k=1}^{n} x_{ki}}(1-\Theta_i)^{\sum\limits_{k=1}^{n}(1-x_{ki})} \tag{4}$$

等式两边取对数可得

$$\ln p(X|\Theta) = \sum_{i=1}^{d}\Big(\sum_{k=1}^{n} x_{ki}\ln\Theta_i + \sum_{k=1}^{n}(1-x_{ki})\ln(1-\Theta_i)\Big) \tag{5}$$

$\ln p(X|\Theta)$ 对 Θ_i 求导, 并令其等于 0，可得

$$\Big(\sum_{k=1}^{n} x_{ki}\frac{1}{\Theta_i} + \sum_{k=1}^{n}(1-\Theta_i)\frac{1}{\Theta_i - 1}\Big) = 0 \tag{6}$$

两边同时乘以 $\Theta_i(1-\Theta_i)$ 得到

$$\Big(\sum_{k=1}^{n} x_{ki}(\Theta_i - 1) + \sum_{k=1}^{n}(1-x_{ki})\Theta_i\Big) = 0 \tag{7}$$

即

$$(\Theta_i - \frac{1}{n}\sum_{k=1}^{n} x_{ki}) = 0 \tag{8}$$

也就是

$$(\Theta_i = \frac{1}{n}\sum_{k=1}^{n} x_{ki}) \tag{9}$$

$$\widehat{\Theta} = \frac{1}{n}\sum_{k=1}^{n} X_k \tag{10}$$

2 Hw 3.2

令 $D=\{x_1,\cdots,x_n\}$ 为 n 个独立的已标记的集合。令 $D_k(x)=\{x_1',\cdots,x_k'\}$ 为样本 x 的 k 个最邻近。根据 k - 近邻规则，x 将归入 $D_k(x)$ 中出现次数最多的那个类别。考虑一个 2 类别问题，先验概率为 $P(w_1)=P(w_2)=1/2$。进一步假设类条件概率密度 $P(X|w_i)$, 在 10 单位超球体内为均匀分布。

（a）证明如果 k 为奇数，那么平均误差率为

$$P_n(e)=\frac{1}{2^n}\sum_{j=0}^{(k-1)/2}\binom{n}{j} \tag{11}$$

（b）证明在这种情况下，如果 k>1，那么最近邻规则比 k-近邻规则有更低的误差率。

（c）如果 k 随着 n 的增加而增加，同时又受 $k<a\sqrt{n}$ 的限制，那么证明：当 $n\to\infty$ 时，$p_n(e)\to 0$

解：（1）、假设 $P_n(e)$ 表示平均误差率，那么 $P_n(e)$ 等于将 $w1$ 分类为 $w2$ 的概率和将 $w2$ 分类为 $w1$ 的概率之和. 也等于两倍将 $w1$ 错误分类为 $w2$ 的概率之和。那么 $P_n(e)=2P(w1)Pr[$标记为 $w1$ 的点少于 (k-1)/2，其他为标记 $w2$ 的点$]=2\frac{1}{2}\sum_{j=0}^{(k-1)/2}Pr[$n 个样本点中由 j 个被标记为 $w1$，其余标记为 $w2]=\sum_{j=0}^{(k-1)/2}\binom{n}{j}\frac{1}{2_j}\frac{1}{2^{(n-j)}}=P_n(e)=\frac{1}{2^n}\sum_{j=0}^{(k-1)/2}\binom{n}{j}$。

（2）、对于平均误差率，当 k = 1 时，对应最近邻规则，此时误差率为 $P_n(e)=\frac{1}{2^n}$，而对于 k-近邻规则，当 k>1 时，$\sum_{j=0}^{(k-1)/2}\binom{n}{j}$ 是一个大于 1 的数，于是有 $\frac{1}{2^n}<\frac{1}{2^n}\sum_{j=0}^{(k-1)/2}\binom{n}{j}$, 由此证明，如果 k>1，那么最近邻规则比 k-近邻规则有更低的误差率。

（3）、用 $B(\cdot,\cdot)$ 表示一个二项分布，假设 $Y_1,\cdots Y_n$ 相互独立并且有 $Pr[Y_i=1]=Pr[Y_i=0]=1/2$。那么 $P_n(e)$ 可以表示为 $Pr[B(n,1/2)\leq(k-1)/2]$, 即 $Pr[Y_1+\cdots+Y_n\leq\frac{k-1}{2}]$。假如 k 随着 n 增加而增加，同时又受到 $k<a\sqrt{n}$ 的限制，那么有：

$$P_n(e)\leq Pr(Y_1+\cdots+Y_n\leq\frac{a/\sqrt{n}-1}{2}) \tag{12}$$

而当 $n\to\infty$ 时，有：

$$Pr(Y_1+\cdots+Y_n\leq\frac{a/\sqrt{n}-1}{2})=Pr(Y_1+\cdots+Y_n\leq 0)=0 \tag{13}$$

由此可得，当 $n\to\infty$ 时，有 $P_n(e)\to 0$