判断填空简答计算

编码

1. DMS渐进无差错编码定理：当编码速率$R = \frac{N}{L}logD$ 大于H(U)，可以实现**渐进无差错编码**
2. 不等长编码的基本要求：

- **唯一可译性**：需要用后缀分解法判断（除S0外没有任何一个后缀分解集中包含码字）
    - 如果一个码唯一可译->Kraft不等式成立->存在一个具有同样长度的异字头码
    - 由**不等长编码定理**知：任何**D元唯一可译码平均码**均满足 $\bar{n} \geq \frac{H(U)}{logD}$，一定存在一个D元唯一可译码使得 $\bar{n} \leq \frac{H(U)}{logD} + 1$
    - 唯一可译码不一定是异字头码
- 即时可译性：后缀分解S1是空集

3. Kraft不等式：存在D元异字头码的充要条件

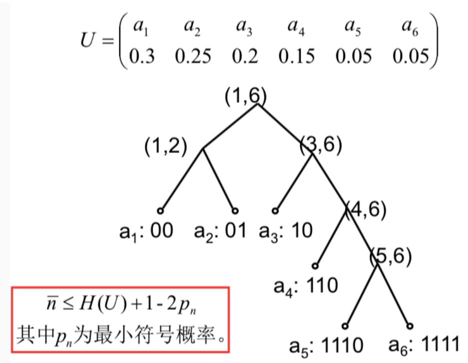
存在长度为 $n_1, n_2, \cdots, n_K$ 的D元异字头码的充要条件为

$$\sum_{k=1}^{K} D^{-n_k} \leq 1$$

4. 不等长编码具体算法：

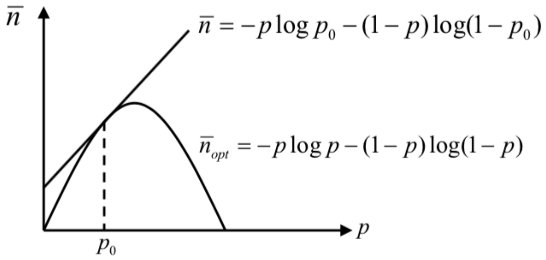
   都先将字符出现概率从大到小排列

   1. Huffman
   2. Shannon：给a_k编码，截取长度$l_k = \lceil log(p_k) \rceil$，截取码字为$P_K = \sum_{i=1}^{k-1} p_i$的小数点后展开串
   3. Fano：消息分成两组，使两组概率和尽量相同，然后重复分成等概两组，给每个部分指定0/1



5. 通用信源编码的必要性：

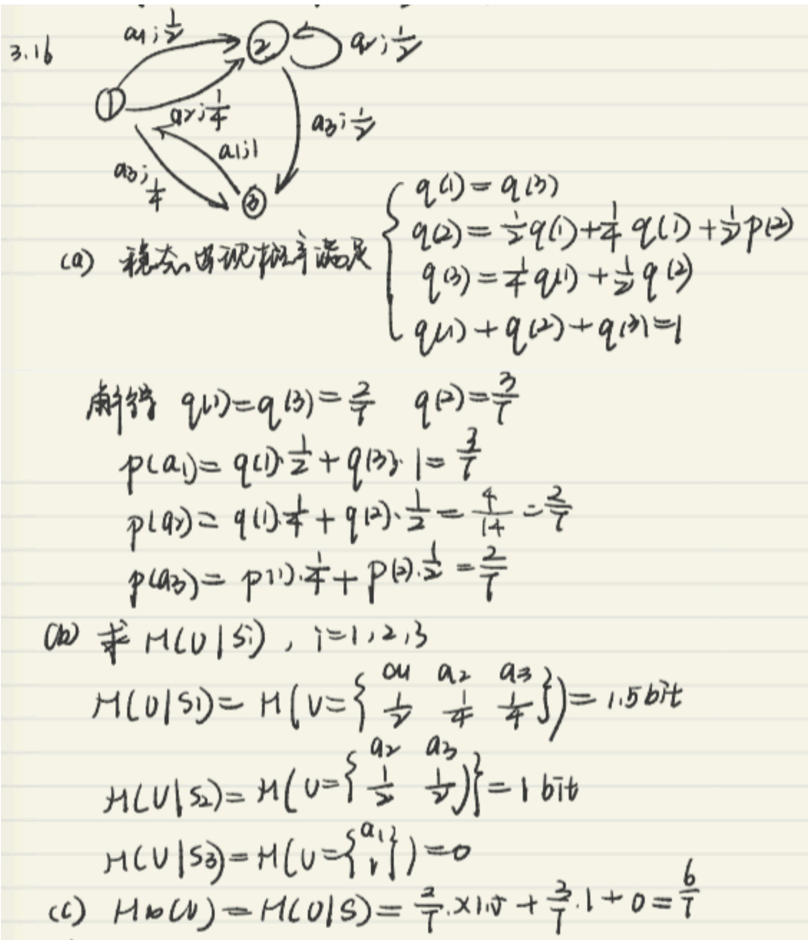
   不知道字符统计特性时，采用通用编码技术。考虑shannon编码的例子，当p和p0失配加大，平均码长迅速偏离shannon编码性能



Shannon编码在不能准确获得信源分布时的性能恶化

6. 马尔科夫源的编码：
    - 先用**转移概率矩阵的转置**和**状态概率和为1**，求出信源各状态S的概率
    - 然后**信源熵速率**等于**各状态下信源熵的加权和**：

$$H_\infty(U) = \sum_{s\in S} q(S=s)H(U\mid S=s)$$
$$= H(U\mid S)$$



3.16

(a) 稳态出现概率满足
$$\begin{cases} q(1)=q(3) \\ q(2)=\frac{1}{2}q(1)+\frac{1}{4}q(1)+\frac{1}{2}p(2) \\ q(3)=\frac{1}{4}q(1)+\frac{1}{2}q(2) \\ q(1)+q(2)+q(3)=1 \end{cases}$$

解得 $q(1)=q(3)=\frac{2}{7}$ $q(2)=\frac{3}{7}$

$p(a_1)=q(1)\cdot\frac{1}{2}+q(3)\cdot 1=\frac{3}{7}$

$p(a_2)=q(1)\cdot\frac{1}{4}+q(2)\cdot\frac{1}{2}=\frac{4}{14}=\frac{2}{7}$

$p(a_3)=p(1)\cdot\frac{1}{4}+p(2)\cdot\frac{1}{2}=\frac{2}{7}$

(b) 求 $H(U\mid S_i)$，$i=1,2,3$

$H(U\mid S_1)=H\left(U=\begin{Bmatrix} a_1 & a_2 & a_3 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4}\end{Bmatrix}\right)=1.5\text{bit}$

$H(U\mid S_2)=H\left(U=\begin{Bmatrix} a_2 & a_3 \\ \frac{1}{2} & \frac{1}{2}\end{Bmatrix}\right)=1\text{bit}$

$H(U\mid S_3)=H\left(U=\begin{Bmatrix} a_1 \\ 1\end{Bmatrix}\right)=0$

(c) $H_\infty(U)=H(U\mid S)=\frac{2}{7}\cdot 1.5+\frac{3}{7}\cdot 1+0=\frac{6}{7}$

率失真(Distortion)信源编码——有损压缩

- 信源的无损表达：$R >= H(U)$
- 信源的有损表达：$R < H(U)$

1. 零速率编码可达到的最小失真

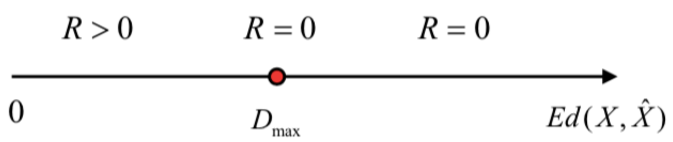
定义： D_{MAX}是无论信源产生什么输出，都恢复到一个固定点，并且这个固定点满足取它时，平均失真最小。那么信源和这个点的平均（可以用汉明失真距离）距离就是D_{MAX}

含义： 无需对信源做任何描述，可以达到的最小失真；对信源进行适当描述时允许的最大失真

当对源$\mathcal{X}$进行零速率编码时，即无论它输出什么符号，都用某个固定的恢复点来表达它，此时可以达到的最小平均失真为

$$D_{\max} = Ed(X,\hat{x}^*) = \min_{\hat{x}\in\hat{\mathcal{X}}} \sum_{x\in\mathcal{X}} p(x)d(x,\hat{x})$$

这个失真既是不必对该源进行编码的可容许失真的最小起点，也是必须考虑对该源进行适当编码的可容许失真的最大值。



2. **率失真函数及其支撑集****

对符号独立同分布的信源$(X, p(x), \mathcal{X})$进行限失真编码，恢复点集为$\hat{\mathcal{X}}$，相应的单符号有界失真度量为$d(x,\hat{x})$。则其率失真函数$R(D)$等于相应的信息率失真函数

$$\begin{aligned} R^I(D) &= \min_{\hat{X}:Ed(X,\hat{X})\le D} I(X;\hat{X}) \\ &= \min_{\{q(\hat{x}|x)\}\in\mathcal{Q}_D} I(\{q(\hat{x}\mid x)\}) \end{aligned}$$

其中，$\mathcal{Q}_D = \left\{\{q(\hat{x}\mid x)\}: \sum_x\sum_{\hat{x}} p(x)q(\hat{x}\mid x)d(x,\hat{x}) \le D\right\}.$

$$D_{\min} = \sum_{x\in\mathcal{X}} p(x)C_x = \sum_{x\in\mathcal{X}} p(x)\min_{\hat{x}\in\hat{\mathcal{X}}} d(x,\hat{x})$$

$$D_{\max} = Ed(X,\hat{x}^*) = \min_{\hat{x}\in\hat{\mathcal{X}}} \sum_{x\in\mathcal{X}} p(x)d(x,\hat{x})$$

思路：

- D_{MAX}是恢复到一个点，让平均距离最小；
- D_{MIN}是让每个x恢复到最小距离的点，算出来的最小平均距离；
- R(D)信源失真函数是在**某一种信道上**能达到函数的下界，所以要求一个转移概率矩阵让其：在平均距离小于D的前提下，找到最小互信息。

  该信道满足：

  存在与失真度量矩阵具有同样对称性的转移概率分布达到率失真 R(D)

信源分布为$P=\{p_1,p_2,\cdots,p_r\}$，失真度量矩阵为$(d(i,j))_{r\times s}$.
π为$\{1,2,\cdots,r\}$上的一个置换，ρ为$\{1,2,\cdots,s\}$上的一个置换。
如果

1) $p_i=p_{\pi(i)}$,

2) $d(i,j)=d(\pi(i),\rho(j))$,

对所有$i=1,2,\cdots,r;j=1,2,\cdots,s$均成立，则存在一个达到率失真函数$R(D)$的转移概率矩阵$(q(j\mid i))_{r\times s}$具有与失真度量矩阵$(d(i,j))_{r\times s}$相同的置换对称性，即

3) $q(j\mid i)=q(\rho(j)\mid\pi(i))$,

对所有$i=1,2,\cdots,r;j=1,2,\cdots,s$均成立.

$$\hat{\mathcal{X}}=\{\hat{x}_0,\hat{x}_1,\hat{x}_2\}\qquad (\boldsymbol{d})_{2\times 3}=\begin{bmatrix}0 & \infty & 1\\ \infty & 0 & 1\end{bmatrix}$$

$$(\boldsymbol{q}(j\mid i))_{2\times 3}=\begin{bmatrix}\alpha & \beta & \gamma\\ \beta & \alpha & \gamma\end{bmatrix},\quad \alpha+\beta+\gamma=1$$

一般都是n元对称信道，然后需要解一下什么样的条件转移概率矩阵，可以使平均距离等于D

- 然后得到信道后就可以算互信息了，即为R(D)

（一）

4.1 一个四元对称信源 $\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{Bmatrix} 0 & 1 & 2 & 3 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{Bmatrix}$，接收符号 $Y=\{0, 1, 2, 3\}$ ，其失真矩阵为 $\begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}$，求 D_{max} 和 D_{min} 及信源的 R(D)函数，并画出其曲线（取 4 至 5 个点）。

解：

$$D_{max} = \min_j D_j = \min_j \sum_i p(x_i) d(x_i, y_j) = \frac{1}{4}\times 1 + \frac{1}{4}\times 1 + \frac{1}{4}\times 1 + \frac{1}{4}\times 0 = \frac{3}{4}$$

$$D_{min} = \sum_i p(x_i) \min_j d(x_i, y_j) = \frac{1}{4}\times 0 + \frac{1}{4}\times 0 + \frac{1}{4}\times 0 + \frac{1}{4}\times 0 = 0$$

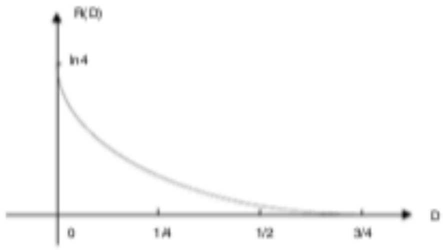
因为 n 元等概信源率失真函数：

$$R(D) = \ln n + \frac{D}{a}\ln\frac{\frac{D}{a}}{n-1} + \left(1-\frac{D}{a}\right)\ln\left(1-\frac{D}{a}\right)$$

其中 a＝1，n＝4，所以率失真函数为：

$$R(D) = \ln 4 + D\ln\frac{D}{3} + (1-D)\ln(1-D)$$

函数曲线：



其中：

$D = 0, R(0) = \ln 4 \ \ nat/symbol$

$D = \frac{1}{4}, R(D) = \ln 4 - \frac{1}{2}\ln\frac{16}{3} \ \ nat/symbol$

$D = \frac{1}{2}, R(D) = \ln 4 - \frac{1}{2}\ln 12 \ \ nat/symbol$

$D = \frac{3}{4}, R(D) = 0 \ \ nat/symbol$

(二)

3:

X= [0 1 失真矩阵为[0 a

0.5 0.5]; a 0]

求Dmin Dmax， R(D)

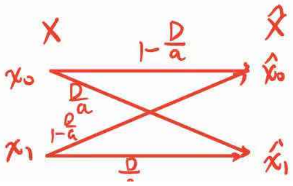
$D_{min}=\sum_x p(x)\min_{\hat{x}\in\hat{X}} d(x,\hat{x})=0$ $\quad (d)_{2\times2}=\begin{bmatrix}0&a\\a&0\end{bmatrix}$ $\quad \hat{X}=\{\hat{x}_0,\hat{x}_1,\hat{x}_2\}$

8 $D_{max}=\min_{\hat{x}}\sum_x p(x)d(x,\hat{x})=0.5a$ $\quad \begin{pmatrix}X\\P(x)\end{pmatrix}=\begin{pmatrix}x_0&x_1\\0.5&0.5\end{pmatrix}$ $\quad \hat{X}=\{\hat{x}_0,\hat{x}_1\}$

- 存在与失真度量矩阵具有同样对称性的转移概率分布达到率失真 $R(D)$

$Q=\begin{bmatrix}\alpha&1-\alpha\\1-\alpha&\alpha\end{bmatrix}$, $D=\sum_{x,\hat{x}} p(x)q(\hat{x}|x)d(x,\hat{x})=(1-\alpha)a=D.$

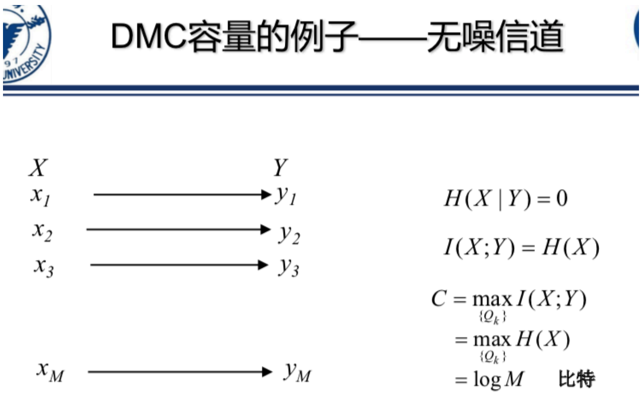
$\therefore \alpha=1-\frac{D}{a}$

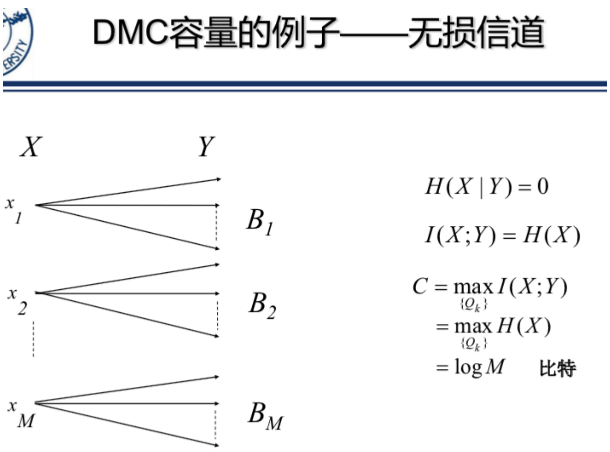


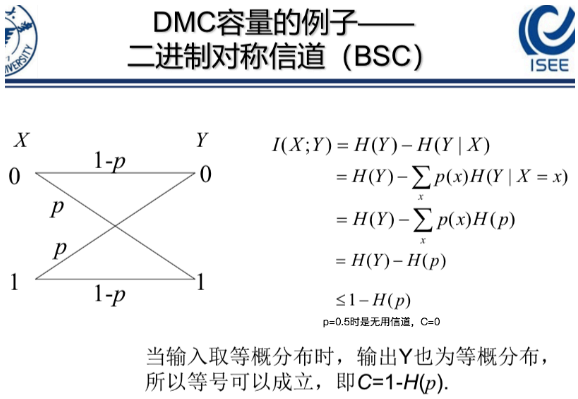
二进制对称信道：$R(D)=I(X;Y)=1-H(\frac{D}{a})$

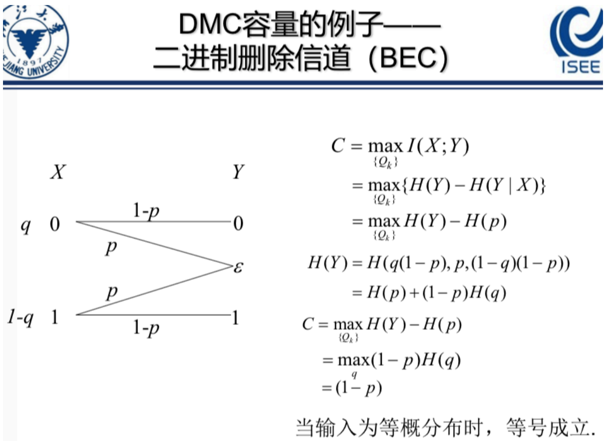
信道

1. 信道容量：输入输出符号互信息的最大值 $\boxed{C=\lim_{n\to\infty}\frac{1}{n}\max_{\{Q(\mathbf{x}^n)\}} I(X_1X_2\cdots X_n;Y_1Y_2\cdots Y_n)}$
2. 特殊信道 DMC的容量：$\boxed{C=\max_{\{Q_k\}} I(X;Y)}$
3. DMC例子









4. 离散无记忆信道容量定理

概率分布$\{Q_0,Q_1,\cdots Q_{K-1}\}$达到转移概率为$\{p(j|k)\}$的离散无记忆信道容量C的充要条件为：

$$I(X=k;Y)=C \qquad \forall k, Q_k>0$$

$$I(X=k;Y)\le C \qquad \forall k, Q_k=0$$

其中$I(X=k;Y)$表示通过信道传送字符$X=k$时，信道的输入与输出之间可获得的互信息的期望值，即

$$I(X=k;Y)=\sum_{j=0}^{J} p(j|k)\log\frac{p(j|k)}{\sum_{i=0}^{K-1} Q_i p(j|i)}$$

所有送到j的概率和

insight：达到离散无记忆信道容量时，发送符号集中的每个符号虽然被利用的概率Q_i不一定相同，但是一旦被利用$Q_i>0$，其通过信道传输的信息量必然相等，且等于信道容量

5. 离散无记忆**信道编码定理**：传输速率R<C,可达，即存在(M,n)码，使最大误码率趋近于0。

$$R=\frac{\log M}{n}\approx I(X;Y)\to C$$

6. 带限、加性白高斯噪声信道

T秒信道容量：$C_T=WTlog(1+\frac{P}{NW})$

每秒信道容量：$C=Wlog(1+\frac{P}{NW})$ bit/s

频带效率：$\eta = \frac{\text{每秒传输速率}R}{\text{传输带宽}W} \leq \frac{C}{W}$ bit/s/Hz

控制

1. 状态空间模型

$\varphi = [IS - A]^{-1}$

$X = \varphi BU(S)$

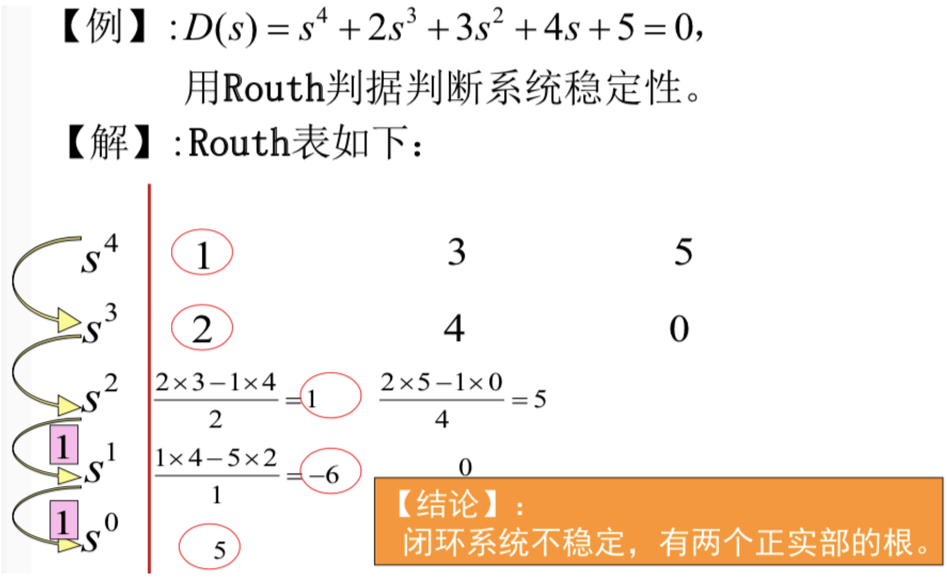
$Y = C\varphi BU(S) + DU(S)$ 传递函数$W = C\varphi B + D$

可控性：$Uc = [B, AB, A^2B, \ldots]$满秩

可观性：$Uo = [C, CA, CA^2, \ldots]^T$满秩

2.稳定性判断

- Routh：**特征方程**有传递函数分母得到，s次数从高往底排



- 李雅普诺夫第一法、间接法

$$\mathbf{W}(s) = \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}$$

外部稳定（有界输入有界输出）：传递函数**W(s)**极点全在左平面

内部稳定：求det(Iλ-A)=0（或者传递函数的分母），A**所有特征值全小于0**（渐进稳定）；**小于0有一个等于0**（李雅普诺夫稳定）；else不稳定

- 李雅普诺夫第二法、直接法

能量函数V(x)，如果存在连续一阶偏导数的标量函数V'(x)，对t求导，x上带一点代表对t的导数

$$V(x) = \frac{1}{2}kx_1^2 + \frac{1}{2}mx_2^2 \qquad \dot{V}(\mathbf{x}) = kx_1\dot{x}_1 + mx_2\dot{x}_2 = 0$$

李雅普诺夫意义下的稳定

V(x) 正定， V'(x)负定，大范围渐进稳定

V(x) 正定， V'(x)负半定且V'(x)不恒等于0，大范围渐进稳定

V(x) 正定， V'(x)负半定且V'(x)恒等于0，李雅普诺夫意义下的稳定

计算

1.贝叶斯决策

计算使后验/似然概率最大的假设，哪个**假设**h算出来h_{MAP}大，就是最大后验得到的结论：假设h成立

最大后验(Maximum a posteriori, MAP)决策规则

给定数据D,在候选假设集合H中寻找可能性最大的假设h

$$h_{\mathrm{MAP}} = \arg\max_{h\in H} P(h|D)$$
$$= \arg\max_{h\in H} P(D|h)P(h)$$

最大似然(Maximum likelihood, ML)决策规则

假设集合H中的每个假设有相同的先验概率$P(h_i) = P(h_j)$

$$h_{\mathrm{ML}} = \arg\max_{h\in H} P(D|h)$$

- ML：最大-似然度 $P(+|\text{cancer})$和P $(+|\neg\text{cancer})$
  - $P(+|\text{cancer}) = 0.98, P(+|\neg\text{cancer}) = 0.03$
  - 结论：$h_{\mathrm{ML}} = \text{cancer}$
- MAP：最大-后验概率$P(\text{cancer}|+)$和$P(\neg\text{cancer}|+)$
  - $P(+|\text{cancer})P(\text{cancer}) = 0.0078$
  - $P(+|\neg\text{cancer})P(\neg\text{cancer}) = 0.0298$
  - 后验概率：$P(\text{canner}|+) = 0.21$；$P(\neg\text{cancer}|+) = 0.79$
  - 结论：$h_{\mathrm{MAP}} = \neg\text{cancer}$
- 结论：
  - 贝叶斯推理的结果很大程度上依赖于先验概率
  - 不是完全接受或拒绝假设，只是在观察到较多的数据后增大或减小了假设的可能性

2.朴素贝叶斯分类

在已知若干属性的情况下，计算不同分类的后验概率，最大的分类即是分类的结果

- 适用场景：
  - 每个实例x可由若干属性描述， $x = \langle a_1, a_2, \cdots, a_n\rangle$
  - 目标函数$f(x)$在某有限集合V中取值
- 决策规则：
  - 给定实例的属性$\langle a_1, a_2, \cdots, a_n\rangle$，取最可能的目标值$v_{\mathrm{MAP}}$

$$v_{\mathrm{MAP}} = \arg\max_{v_j\in V} P(v_j|a_1, a_2, \cdots, a_n)$$
$$v_{\mathrm{MAP}} = \arg\max_{v_j\in V} P(a_1, a_2, \cdots, a_n|v_j)P(v_j)$$

□从训练数据中估计：

■ $P(a_1, a_2, \cdots, a_n | v_j)$和$P(v_j)$

□假设：

■ 在给定目标值时，属性值之间相互独立

$$P(a_1, a_2, \cdots, a_n | v_j) = \prod_i P(a_i | v_j)$$

□决策规则

$$v_{NB} = \arg\max_{v_j \in V} P(v_j) \prod_i P(a_i | v_j)$$

例子：分类v是 Yes or No，属性是Sunny/Cool/High/Strong

计算结果：

$P(\text{PlayTennis} = \text{Yes}) = 9/14 = 0.64;$

$P(\text{PlayTennis} = \text{No}) = 5/14 = 0.36$

$P(\text{Wind} = \text{Strong} | \text{PlayTennis} = \text{Yes}) = 3/9 = 0.33$

$P(\text{Wind} = \text{Strong} | \text{PlayTennis} = \text{No}) = 3/5 = 0.60$

$P(\text{Yes})P(\text{Sunny}|\text{Yes})P(\text{Cool}|\text{Yes})P(\text{High}|\text{Yes})P(\text{Strong}|\text{Yes}) = 0.0053$

$P(\text{No})P(\text{Sunny}|\text{No})P(\text{Cool}|\text{No})P(\text{High}|\text{No})P(\text{Strong}|\text{No}) = 0.02$

3.构造决策树（Entropy熵）

□核心问题：那个属性是最佳的分类属性?

□衡量属性价值的标准：信息增益

■ 用熵来定义样例集合S的纯度，对于布尔型分类

$$\text{Entropy}(S) = -p_+ \log_2 p_+ - p_- \log_2 p_-$$

其中p_+是S中正例的比例，p_-是S中反例的比例

■ 用**信息增益**度量期望熵降低：使用属性A相对于样例集合S的信息增益Gain(S, A)被定义为：

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

其中Values(A)是属性A所有可能值的集合，S_v是S中属性A的值为v的子集。

S是：布尔型分类的统计集合

□按属性Wind分类14个样例得到的信息增益计算：

■ $\text{Values}(\text{Wind}) = \text{Weak}, \text{Strong}$

■ $S = [9+, 5-]$

■ $S_{\text{Weak}} = [6+, 2-]$

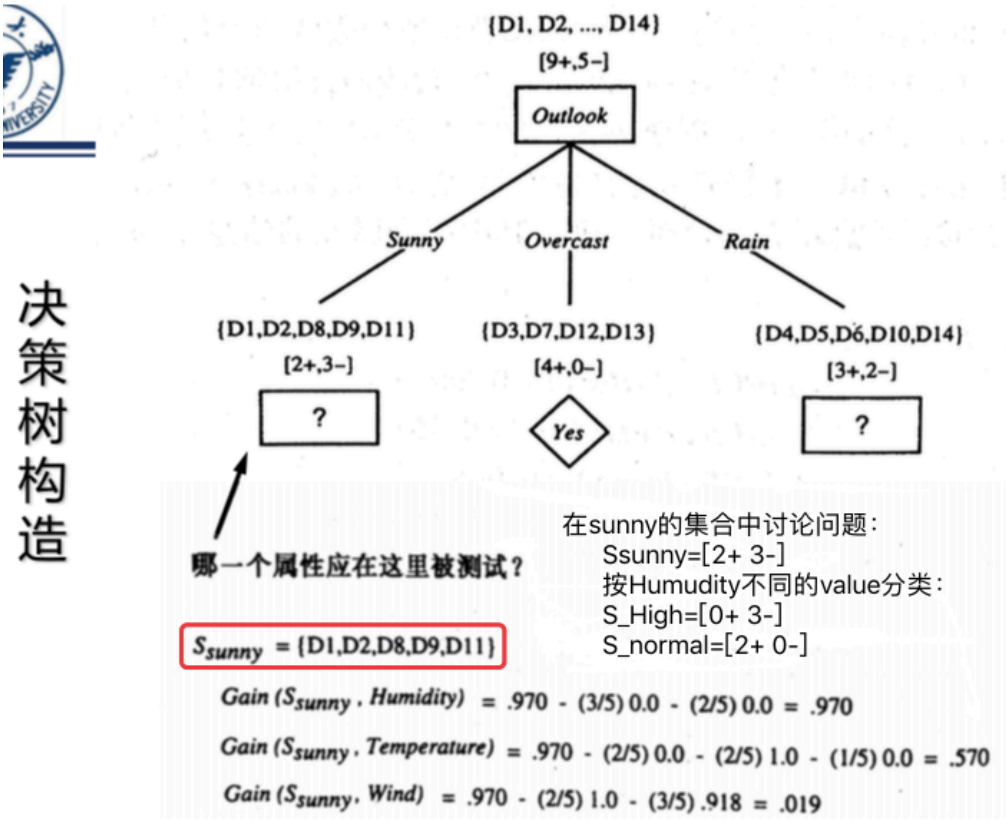
■ $S_{\text{Strong}} = [3+, 3-]$

$$\text{Gain}(S, \text{Wind}) = \text{Entropy}(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

$$= \text{Entropy}(S) - \left(\frac{8}{14}\right) \text{Entropy}(S_{\text{Weak}}) - \left(\frac{\cancel{8}}{14}\right) \text{Entropy}(S_{\text{Strong}})$$

$$= 0.94 - \frac{8}{14} * 0.811 - \frac{6}{14} * 1 = 0.048$$

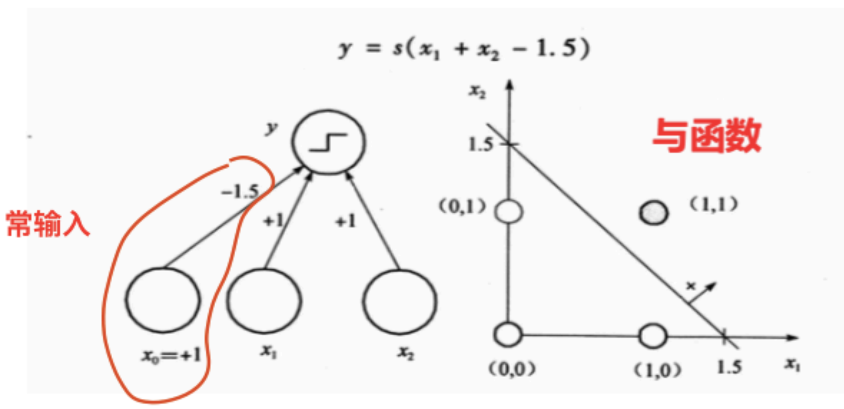
Gain(S,属性)=**S的熵**-属性取value1的概率×value1的熵-属性取value2的概率×value3的熵



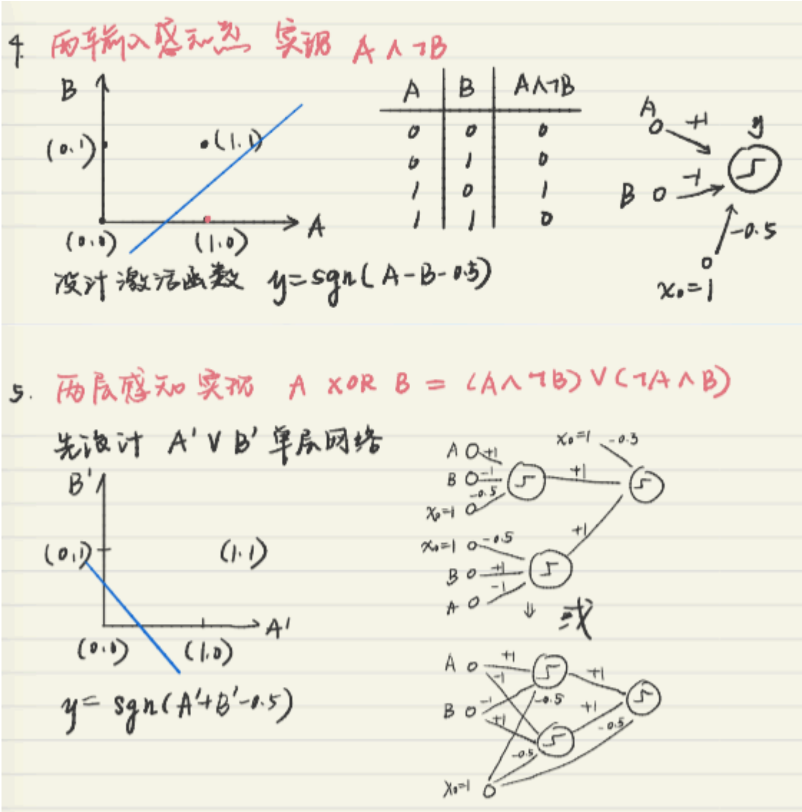
Gain(Ssunny,Humidity)=Ssunny的熵 - sunny里High的频率×High的熵 - sunny里normal的频率×normal的熵

4.人工智能网络

- **单层感知器：**
  - n维空间的超平面（hyperplane）决策面
  - 能够表示所有原子布尔函数：与；与非；或；或非
  - 不能表示异或（XOR）



2. 设计一个两输入的感知器来实现布尔函数$A \wedge \neg B$
3. 设计一个两层的感知器网络来实现异或布尔函数$A \text{ XOR } B$。提示： $A \text{ XOR } B = (A \wedge \neg B) \vee (\neg A \wedge B)$



6.K复杂度

Kolmogorov复杂度上界:

$$K(x) \le K(x|l(x)) + 2\log l(x) + c$$

$$K(x|l(x)) \le l(x) + c$$

1.一般上界:

例3：整数n

$$K(n) \le c + \log^* n, \forall n$$

套娃式告诉计算机读取长度

2.更松的上界:

2. 整数和的复杂度

a) 证明：$K(n) \leq \log n + 2 \log \log n + c$

b) 证明：$K(n_1 + n_2) \leq K(n_1) + K(n_2) + c$

双写法告诉计算机读取长度

eg：”**11000010**1101…“ 加黑部分告诉计算机读取100=4位，于是读到1101。