公式1.2

第8页-第9页。

$$
\begin{aligned}
\sum_f E_{ote}(\mathfrak{L}_a|X,f) &= \sum_f\sum_h\sum_{x\in\mathcal{X}-X}P(x)\mathbb{I}(h(x)\neq f(x))P(h|X,\mathfrak{L}_a) \\
&= \sum_{x\in\mathcal{X}-X}P(x)\sum_h P(h|X,\mathfrak{L}_a)\sum_f\mathbb{I}(h(x)\neq f(x)) \\
&= \sum_{x\in\mathcal{X}-X}P(x)\sum_h P(h|X,\mathfrak{L}_a)\frac{1}{2}2^{|\mathcal{X}|} \\
&= \frac{1}{2}2^{|\mathcal{X}|}\sum_{x\in\mathcal{X}-X}P(x)\sum_h P(h|X,\mathfrak{L}_a) \\
&= 2^{|\mathcal{X}|-1}\sum_{x\in\mathcal{X}-X}P(x)\cdot 1
\end{aligned}
$$

[解析]：

第一步到第二步是因为$\sum_i^m\sum_j^n\sum_k^o a_ib_jc_k=\sum_i^m a_i\cdot\sum_j^n b_j\cdot\sum_k^o c_k$。

第二步到第三步：首先要知道此时f的定义为**任何能将样本映射到{0,1}的函数+均匀分布**，也即不止一个f且每个f出现的概率相等，例如样本空间只有两个样本时：$\mathcal{X}=\{x_1,x_2\},|\mathcal{X}|=2$，那么所有的真实目标函数$f$为：

$f_1: f_1(x_1)=0, f_1(x_2)=0;$
$f_2: f_2(x_1)=0, f_2(x_2)=1;$
$f_3: f_3(x_1)=1, f_3(x_2)=0;$
$f_4: f_4(x_1)=1, f_4(x_2)=1;$

一共$2^{|\mathcal{X}|}=2^2=4$个真实目标函数。所以此时通过算法$\mathfrak{L}_a$学习出来的模型$h(x)$对每个样本无论预测值为0还是1必然有一半的f与之预测值相等，例如，现在学出来的模型$h(x)$对x_1的预测值为1，也即$h(x_1)=1$，那么有且只有f_3和f_4与$h(x)$的预测值相等，也就是有且只有一半的f与它预测值相等，所以$\sum_f\mathbb{I}(h(x)\neq f(x))=\frac{1}{2}2^{|\mathcal{X}|}$。

第三步一直到最后有点概率论的基础应该都能看懂了。$P(h|X,\mathfrak{L}_a)$的含义是算法$\mathfrak{L}_a$基于训练数据X产生假设h的概率，那么对所有的h的概率进行求和，自然就是1了。