Exclusiveness preference property of iHD

iHD has exclusiveness preference property which is expected for a good splitting criterion [Taylor and Silverman, 1993]. This property is defined by the following two conditions:

1. Firstly, for a certain value of $\rho_L \rho_R$, the criterion has the maximum value when, $\sum_{j=1}^{k} p_{Lj} p_{Rj} = 0$ which indicates that the two child nodes are mutually exclusive.
2. Secondly, regardless of $\rho_L \rho_R$, it obtains its minimum value when the class probability distributions of child nodes are identical which can be defined as $p_{Lj} = p_{Rj} = p_j, \forall j$.

Theorem 1. *iHD has the exclusivity preference property.*

Proof. iHD can be written as:

$$\begin{aligned} iHD &= \rho_L \Big(1 - \sum_{j=1}^{k} \sqrt{p_{Lj} p_j}\Big) + \rho_R \Big(1 - \sum_{j=1}^{k} \sqrt{p_{Rj} p_j}\Big) \\ &= 1 - \sum_{j=1}^{k} \sqrt{p_j} (\rho_L \sqrt{p_{Lj}} + \rho_R \sqrt{p_{Rj}}) \end{aligned} \tag{1}$$

For a certain value of $\rho_L \rho_R$, (1) is maximum when the value from the summation over all classes is minimum. In other words, (1) is maximum when we get the minimum value separately for each class in the summation. Let, for j^{th} class, $a_j = p_{Lj}$ and $b_j = p_{Rj}$ for a fixed $\rho_L = 1 - \rho_R \neq 0$. Thus $p_j = \rho_L a_j + \rho_R b_j$ which is a nonzero constant. Now, the j^{th} term in the summation can be expressed by a function of a_j as follows:

$$f(a_j) = \sqrt{p_j} (\rho_L \sqrt{a_j} + \sqrt{\rho_R} \sqrt{p_j - a_j \rho_L})$$

The second derivative of $f(a_j)$ is:

$$f^{''}(a_j) = -\frac{\rho_L \sqrt{p_j}}{4 a_j^{\frac{3}{2}}} - \frac{a_j^2 \sqrt{\rho_R} \sqrt{p_j}}{4 (p_j - a_j \rho_L)^{\frac{3}{2}}}$$

Here, $f^{''}(a_j) < 0$ in the interval of $0 \leq a_j \leq \frac{p_j}{\rho_L}$, thus is a concave function. So, $f(a_j)$ has the minimum value at one of the extreme points of the interval which is either $a_j = 0$ or $a_j = \frac{p_j}{\rho_L}$ (equivalent to $b_j = 0$).

From the properties of Hellinger distance, we can say the distance between the class distribution of the parent and child is minimum (=0) when they have identical class distribution. In other words, for $p_{Lj} = p_{Rj} = p_j, \forall j$, regardless of $\rho_L \rho_R$, (1) becomes:

$$\begin{aligned} iHD &= 1 - \sum_{j=1}^{k} \sqrt{p_j} (\rho_L \sqrt{p_j} + \rho_R \sqrt{p_j}) \\ &= 1 - \sum_{j=1}^{k} p_j (\rho_L + \rho_R) = 1 - 1 = 0 \end{aligned}$$

So, iHD is minimum when $p_{Lj} = p_{Rj} = p_j, \forall j$. $\square$

References

[Taylor and Silverman, 1993] Paul C. Taylor and Bernard W. Silverman. Block diagrams and splitting criteria for classification trees. *Statistics and Computing*, 3(4):147–161, Dec 1993.