Dicionário
μ = Mu θ = Theta Ξ = Xi Δ = Delta

Premissa: Existe um conjunto universo denominado **μ**

Seja, **a'** um atributo que pertence a este conjunto, então

$$a' \in \mu$$

Considera então **θ** é formado pelo subconjunto de **μ**

$$\theta = a_1 \ldots a_n, n \epsilon \mathbb{N}, \theta \subset \mu$$

Considere **c'** uma das possíveis classes determinada por meio da intersecção dos atributos de **μ**, isto é uma regra de decisão, então

$$c' = c \ldots c_n = \Xi \therefore c' \rightarrow \bigcup_1^n \Theta$$

c' implica na interseção de *n* atributos no subconjunto θ

De outra forma,

$$\int \bigcup_1^n \Theta = \int \bigcup_1^n a_1 \ldots a_n = c', a_x \epsilon \, \Theta$$

Conclui-se, aqui, que em uma arvore de decisão, a função acima busca diminuir o número de atributos que fazem intersecção entre si para determinar uma classe. Essa "diminuição" se dá em usar uma função adequada em selecionar o atributo que simboliza o menor caminho, ou matematicamente, o caminho é determinado pelo atributo com maior relevância/significância em relação aos demais.

Considere **Ξ** como o conjunto de possíveis classes determinadas por meio de um conjunto **μ** e o conjunto **μ** determina mais de uma classe, então,

$$\sum_1^n \int \bigcup_1^m \Theta \epsilon \, \Xi$$

Ou,

$$\sum_1^n \int_1^n \bigcup_1^m a_1 \ldots a_m \epsilon \, \Xi$$

Considere **p** como a probabilidade de para cada classe dado o conjunto **θ**, então,

$$\sum_{i=1} p_i = 1$$

A soma das probabilidades de cada classe é igual a 1

A probabilidade de uma classe é determinada por,

$$\int c' = p'$$

Conclui-se então que existe uma função que calcula a probabilidade de uma regra de decisão que determina uma classe.
A regra de decisão por sua vez é determinada por meio da intersecção de atributos pertencentes a um conjunto universo.

Então,

$$\sum_{i=1}^{n} \int c'_n = \sum_{i=1} p_i = 1$$

Ou,

$$\sum_{i=1}^{n} \int \left(\int \bigcup_{i=1}^{m} \Theta \right)_n = \sum_{i=1} p_i = 1$$

Ou,

$$\sum_{1}^{n} \int \left(\int \bigcup_{1}^{m} a_1 \dots a_m \right)_n = \sum_{1} p_i = 1$$

Conclui-se que esta probabilidade é determinada a partir de uma função aplicada a uma regra de decisão, no entanto, para uma mesma classe existe *n* possibilidades chamadas caminhos para se chegar a mesma classe ou resultado.

Considere que **Δ** é um conjunto das probabilidades para cada classe determinadas em **μ**, então

$$\sum_{i=1} p_i \in \Delta, \sum \Delta = 1$$

Conclui-se aqui que a probabilidade de cada classe ocorrer é a soma das probabilidades obtidas a partir do conjunto de regras de decisão possíveis. Igualando a um, significa que as probabilidades estão distribuídas entre 0 e 100% segundo suas proporções calculadas.

Premissa: Se **Δ** contém as probabilidades de determinar cada possível classe de **Ξ**, estas possibilidades não exprimem relação entre si para compor a arvore de decisão, o entendimento correto é de que estas probabilidades determinam qual classes podem resultar dado um exemplo a ser observado.

Então, a probabilidade **p** para cada classe obtida pelas probabilidades calculadas para cada regra de decisão em **c’** pertencente a **θ**, então

$$\left(\sum_{1}^{n} p_n = \sum_{1}^{m} c'_m\right) \epsilon\ \Delta$$

Ou,

$$\sum_{i}^{n} p_n = \sum_{1}^{m} \left(\int \bigcup_{1}^{n} \Theta\right)_m$$

Ou,

$$\sum_{i}^{n} p_n = \sum_{1}^{m} \left(\int \bigcup_{1}^{n} a_1 \dots a_n\right)_m$$

Considere então que existe uma função que determina a probabilidade **p** para cada classe obtida pelas probabilidades calculadas para cada regra de decisão em **c’** pertencente a **θ**, então

$$\sum_{i}^{n} p_n = \int \sum_{1}^{m} \left(\int \bigcup_{1}^{n} a_1 \dots a_n\right)_m$$

Conclui-se que,
A soma de todas as probabilidades **p** aplicadas as possíveis classes pertencentes ao conjunto **Δ** é igual 1 ou 100%.
A probabilidade **p** é proporcional ao número de exemplos por classe do conjunto **Ξ**.
A probabilidade para cada classe é obtida pela soma das probabilidades de ocorrência de cada regra de decisão formada em **θ**.
Se a soma das probabilidades das classes for diferente de um ou um valor muito inferior a arvore de decisão pode estar mal formada o que a torna imprecisa.

Se a soma das probabilidades de cada regra de decisão diferirem da probabilidade da classe correspondente então a arvore está desbalanceada.

Premissa: Existe uma probabilidade **p** de um exemplo ser classificado em uma classe pertencente a **Ξ**, no entanto, esta probabilidade é determinada pelos atributos pertencentes a **θ**, então existe relação entre p, **Ξ** e **θ** em que a probabilidade dos atributos é possível ser determinada e conhecida.

Considere que para cada atributo **a’** existe um conjunto finito de respostas denominada **r**, então

$$\sum_{1}^{i} r_i \subset a'$$

A probabilidade **q** de cada resposta pode ser determinada pela frequência relativa das respostas em **a’**, então,

$$q = \frac{r_i}{\sum_1^i r_i}$$

No entanto, a relação existe entre atributo, probabilidade e classe, então o conjunto **υ** contém as probabilidades de **r’** determinar ou não a classe.

$$\upsilon = \sum_{1}^{n} q_n + \sum_{1}^{m} q_m \, , n \epsilon \, \Xi', m \notin \Xi'$$

Ou,

$$\upsilon = \sum_{1}^{n} \left(\frac{r_i}{\sum_1^i r_i} \right)_n + \sum_{1}^{m} \left(\frac{r_i}{\sum_1^i r_i} \right)_m \, , n \epsilon \, \Xi', m \notin \Xi'$$

E,

$$\upsilon = \sum_{1}^{n} \left(\frac{r_i}{\sum_1^i r_i} \right)_n + \sum_{1}^{m} \left(\frac{r_i}{\sum_1^i r_i} \right)_m = 1$$

Concluí-se que as probabilidades são calculadas em função da frequência de respostas para cada atributo explorando a possibilidade de o atributo determinar e não determinar a classe. Os cálculos para ambos os sentidos devem ser feitos e caso sua soma se diferente de 1 ou um valor muito inferior então a arvore de decisão possui problemas com os dados de treinamento.

Por fim, existe uma regra generalizada para a

$$\sum_{i=1} f(a'_i) = \left(\sum_1^n \left(\frac{r_j}{\sum_1^j r_j} \right)_n , \sum_1^m \left(\frac{r_k}{\sum_1^k r_k} \right)_m \right), a' \in \theta . n \epsilon \Xi', m \notin \Xi'$$