Formula Sheet for First Midterm Exam

-

$$Entropy(S) = \sum_{i=1}^{n} -Prob(S_i) \times log_2 Prob(S_i)$$

where S is divided into n subsets, $S_1, \ldots, S_n$

-

$$GainRatio(A) = \frac{InfoGain(S, A)}{SplitInfo(S, A)}$$

where

$$SplitInfo(S, A) = -\sum_{i=1}^{k} \frac{|S_i|}{|S|} \times log(\frac{|S_i|}{|S|})$$

where S is divided into k subsets $S_1, \ldots, S_k$

-

$$Gini(S) = 1 - \sum_{i=1}^{m} p_i^2$$

where S is divided into m subsets

-

$$\chi^2 = \sum_{allcells} \frac{(C_{i,j} - E_{i,j})^2}{E_{i,j}}$$

where

$$E_{i,j} = \frac{RowTotal_i \times ColumnTotal_j}{N}$$

-

$$p = \frac{f + \frac{z^2}{2N} \pm z\sqrt{\frac{f}{N} - \frac{f^2}{N} + \frac{z^2}{4N^2}}}{1 + \frac{z^2}{N}}$$

-

$$f(x) = \frac{e^{-\frac{(x-mean)^2}{2*(StandDev)^2}}}{\sqrt{2\pi} * StandDev}$$

-

$$pessimistic \;\; error: \quad q = \frac{e + \frac{z^2}{2N} + z\sqrt{\frac{e}{N} - \frac{e^2}{N} + \frac{z^2}{4N^2}}}{1 + \frac{z^2}{N}}$$