Shannon’s Model of Entropy:

is probability that random element ***t*** is type ***i***, ***l*** is number of different types of ***t***, and ***s*** is arbitrary but usually 2 for calculating bits.

Information Gain, entropy wrt target feature:



Information Gain, entropy after partitioning:

Information Gain:

Information Gain Ratio:

Binomial probability

n = num trials, x = num successes desired, p = prob success on one trial, q = 1-p

Similarity metric criteria:  
Non-negativity,   
Identity, Symmetry,   
Triangular inequality:

K nearest neighbours:

Weighted k nearest neighbour model:

Range normalisation:



Prior probability

joint prob of descriptive features

conditional prob

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | **Prediction** | | |
|  |  | TRUE | FALSE | Precision |
| **Target** | TRUE | *TP* | *FP* |  |
| FALSE | *FN* | *TN* |  |
| Recall |  |  |  |

Generalised Bayes Theorem:

Chain Rule:

Naïve Bayes’ Classifier:

Laplacian Smoothing (conditional probabilities):