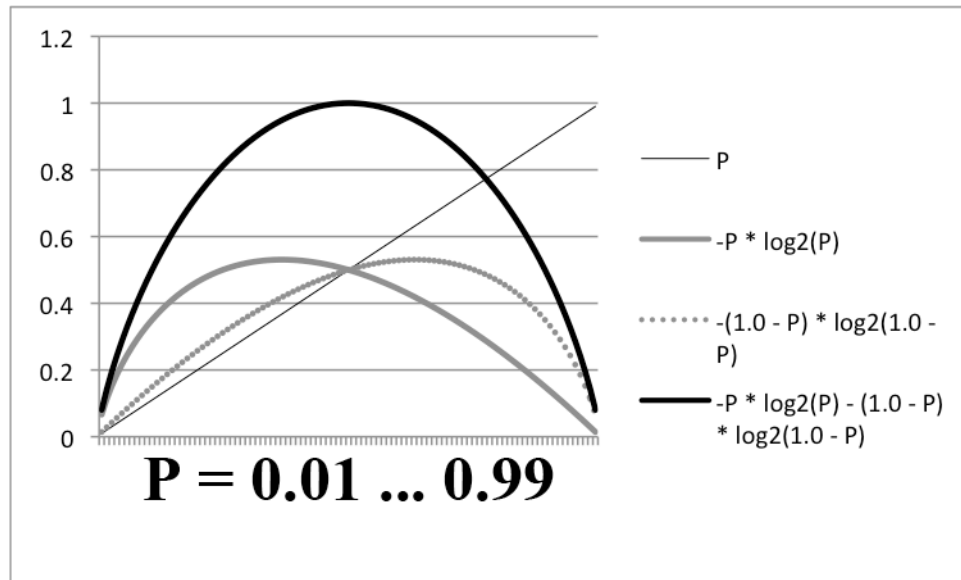


CSC 556 – A little graph on informational entropy

Dr. D. Parson, March 6, 2013

In binary classification / decision making, as the probability of two alternative events becomes equal at 50%, the informational entropy reaches its maximum at 1.0. Think of it as a measure of uncertainty of what to do next. Decision trees and rules typically focus on reducing that entropy (uncertainty) as much as possible at each step.



```
1 # Information entropy discussion, March 6, 2013, D. Parson
2 # CSC 556, DBMS II as it relates to data analysis and mining
3 from math import log10
4
5 # Find log base 2 of n.
6 def log2(n):
7     return log10(n) / log10(2.0)
8
9 # Plot f across the range .01 .. .99 in steps of .01
10 # Avoid potential undefined values at 0.0 and 1.0.
11 def plot(f):
12     p = .01
13     while p < 1.0:
14         print p, f(p)
15         p = p + .01
16
17 def entropy(p):
18     return (-p * log2(p) - (1.0-p) * log2(1.0-p))
19
20 def partEntropy(p):
21     return (-p * log2(p))
```