

ADS Assignment 8

JJ Yang

Decision Trees

1. Entropy = un-homogeneity within groupings

$$= - \sum_{k=1}^K \hat{P}_{mk} \log(\hat{P}_{mk})$$

where \hat{p} = sample prob
 m = grouping
 k = outcome/class

Information gain = reduction of entropy = $E_n - E_{n-1}$
 = reduction of mis-classification error

2.

A	B	C	Label
a	a	a	1
b	b	a	2
a	a	b	1
b	b	a	2

E_{initial}

$$= -.5 \ln(.5) = .69$$

Feature A split:

$$E_A = - \sum_{k=1}^2 \hat{P}_{mk} \log(\hat{P}_{mk}) = -1 \ln(1) - 1 \ln(1) = 0$$

Feature B split:

$$E_B =$$

$$-1 \ln(1) - 1 \ln(1) = 0$$

Feature C split:

$$E_C =$$

$$-.5 \ln(.5) - .5 \ln(.5) = .69$$