Meat N,Y	Crust D,S,T	Veg N,Y	Quality B,G,Gr
Y	Thin	N	Great
N	Deep	N	Bad
N	Stuffed	Y	Good
Y	Stuffed	Y	Great
Y	Deep	N	Good
Y	Deep	Y	Great
N	Thin	Y	Good
Y	Deep	N	Good
N	Thin	N	Bad

Decision Tree Homework

$$Info(S) = -\sum_{i=1}^{|C|} p_i log_2 p_i$$

$$Info_A(S) = \sum_{j=1}^{|A|} \frac{|S_j|}{|S|} Info(S_j) = \sum_{j=1}^{|A|} \frac{|S_j|}{|S|} \cdot -\sum_{i=1}^{|C|} p_i log_2 p_i$$

- Finish the first level, find the best attribute and split
- Then find the best attribute for the left most node at the second level
  - Assume sub-nodes are sorted alphabetically left to right by attribute
  - Label any leaf nodes with their majority class
  - You could do the other nodes if you want for practice

Meat N,Y	Crust D,S,T	Veg N,Y	Quality B,G,Gr
Y	Thin	N	Great
N	Deep	N	Bad
N	Stuffed	Y	Good
Y	Stuffed	Y	Great
Y	Deep	N	Good
Y	Deep	Y	Great
N	Thin	Y	Good
Y	Deep	N	Good
N	Thin	N	Bad

Decision Tree Homework

$$Info(S) = -\sum_{i=1}^{|C|} p_i log_2 p_i$$

$$Info_A(S) = \sum_{j=1}^{|A|} \frac{|S_j|}{|S|} Info(S_j) = \sum_{j=1}^{|A|} \frac{|S_j|}{|S|} \cdot -\sum_{i=1}^{|C|} p_i log_2 p_i$$

- $Info_{Meat}(S)$ = 4/9·(-2/4log$_2$2/4 - 2/4·log$_2$2/4 - 0·log$_2$0/4) +
  5/9·(-0/5·log$_2$0/5 - 2/5·log$_2$2/5 - 3/5·log$_2$3/5) = .98
- $Info_{Crust}(S)$ = 4/9·(-1/4log$_2$1/4 - 2/4·log$_2$2/4 - 1/4·log$_2$1/4) +
  2/9·(-0/2·log$_2$0/2 - 1/2·log$_2$1/2 - 1/2·log$_2$1/2) +
  3/9·(-1/3·log$_2$1/3 - 1/3·log$_2$1/3 - 1/3·log$_2$1/3) = 1.41
- $Info_{Veg}(S)$ = 5/9·(-2/5log$_2$2/5 - 2/5·log$_2$2/5 - 1/5·log$_2$1/5) +
  4/9·(-0/4·log$_2$0/4 - 2/4·log$_2$2/4 - 2/4·log$_2$2/4) = 1.29
- Attribute with least remaining info is Meat

Meat N,Y	Crust D,S,T	Veg N,Y	Quality B,G,Gr
N	Deep	N	Bad
N	Stuffed	Y	Good
N	Thin	Y	Good
N	Thin	N	Bad

Decision Tree Homework

$$Info(S) = -\sum_{i=1}^{|C|} p_i log_2 p_i$$

$$Info_A(S) = \sum_{j=1}^{|A|} \frac{|S_j|}{|S|} Info(S_j) = \sum_{j=1}^{|A|} \frac{|S_j|}{|S|} \cdot -\sum_{i=1}^{|C|} p_i log_2 p_i$$

- Left most node will be Meat = N containing 4 instances
- $Info_{Crust}(S)$ = 1/4·(-1/1log$_2$1/1 - 0/1·log$_2$0/1 - 0/1·log$_2$0/1) +
  1/4·(-0/1·log$_2$0/1 - 1/1·log$_2$1/1 - 0/1·log$_2$0/1) +
  2/4·(-1/2·log$_2$1/2 - 1/2·log$_2$1/2 - 0/2·log$_2$0/2) = .5
- $Info_{Veg}(S)$ = 2/4·(-2/2log$_2$2/2 - 0/2·log$_2$0/2 - 0/2·log$_2$0/2) +
  2/4·(-0/2·log$_2$0/2 - 2/2·log$_2$2/2 - 0/2·log$_2$0/2) = 0
- Attribute with least remaining info is Veg which will split into two pure nodes labeled Bad and Good