Announcement: You get extra credits if you write your own program to generate the decision tree correctly (subject to one-on-one interview)

Type Overall Type Arrangement Margin Margin sub-category Count

Dogwood Simple Opposite Entire X         2

Maple Simple Opposite Lobed X 2

Ash Compound Opposite X X 40

Hickey Compound X Toothed X 2

Locust Compound Alternate Entire X 2

Cherry Simple Alternate Toothed X 2

White Oak Simple Alternate Lobed Rounded 100

Red Oak Simple Alternate Lobed Pointed 100

2 3 4 3

Decision Tree:

Bad news: Grows exponentially

Good news: Practically, many cases are already eliminated as it goes.

E[I({Margin Sub-category; Margin}-> Type)] = 1\* 4/250 + 1\*4/250 = 8/250 = 0.008

{0..7} {0,1} {0,1,2} {0,1,2,3} {0,1,2}

Type Overall Type Arrangement Margin Margin sub-category Count

0 0 0 0 0 2

1 0 0 1 0 15

2 1 0 2 0 5

=================================

I(Q-> Type) = -sum\_Type Pr(Type|Q)Log\_2 Pr(Type|Q)

E[I(Q-> Type)] = \sum\_Q I(Q-> Type) Pr(Q)

I(Overall\_Type: Simple->Type)

=-sum\_Type Pr(Type|Q)Log\_2 Pr(Type|Q)

= -Pr(Type:Dogwood|Overall\_Type:Simple)Log\_2 Pr(Type:Dogwood|Overall\_Type:Simple) +

  -Pr(Type:Maple|Overall\_Type:Simple)Log\_2 Pr(Type:Maple|Overall\_Type:Simple) +

  -Pr(Type:Ash|Overall\_Type:Simple)Log\_2 r(Type:Ash|Overall\_Type:Simple) +

   ...

  -Pr(Type:Red Oak|Overall\_Type:Simple)Log\_2 Pr(Type:Red Oak|Overall\_Type:Simple)

= (2/206)\*Log\_2 (206/2) + (2/206)\*Log\_2 (206/2) + 0 + 0 + 0 + (2/206)\*Log\_2 (206/2) + (100/206)\*Log\_2 (206/100) + (100/206)\*Log\_2 (206/100)

= V1 = 1.207

I(Overall\_Type: Compound->Type)

= (40/44)Log\_2(44/40) + 2\*(2/44)Log\_2(44/2)

= V2 = 0.53

E[I(Q-> Type)] = \sum\_Q I(Q-> Type) Pr(Q)

= V1\*Pr(Overall\_Type: Simple) + V2\*Pr(Overall\_Type: Compound)

= V1\*206/250 + V2\*44/250

= Information left that you need to figure out in order to identify the tree leaf type

= 1.08785

=================================

I(Arrangement:Opposite-> Type)

= 2\* 2/44 Log\_2(44/2) + 40/44 Log\_2(44/40)

= V4 = 0.53

I(Arrangement:X-> Type) = 0

I(Arranagement: Alternate->Type)

= 2\* 2/204 Log\_2(204/2) + 2\* 100/204 Log\_2(204/100)

= V5 = 1.139

E[I(Q-> Type)]

= V4\*44/250 + V5\*204/250

= 1.0227

=================================

I(Margin: Entire->Type)

= 2\*2/4 Log\_2(4/2)

= 1

I(Margin: Lobed->Type)

= 2/202\* Log\_2(202/2) + 2\* 100/202 \*Log\_2(202/100)

= V6 = 1.07

I(Margin: X -> Type) = 0

I(Margin: Toothed -> Type)

= = 2\*2/4 Log\_2(4/2)

= 1

E[I(Q-> Type)]

= 4/250 + V6\* 202/250 + 4/250

= 8/250 + V6\* 202/250

= 0.89656

=================================

I(Margin\_subCategory: X -> Type)

= (40/50)\*Log\_2(50/40) + 5\*(2/50)\*Log\_2(50/2)

= V3 = 1.1863

I(Margin\_subCategory: Rounded -> Type)

= -sum\_Type Pr(Type|Margin\_subCategory: Rounded)Log\_2 Pr(Type|Margin\_subCategory: Rounded)

= 0

I(Margin\_subCategory: Pointed -> Type)

= 0

E[I(Q-> Type)]

= \sum\_Q I(Q-> Type) Pr(Q)

= V3\*Pr(Margin\_subCategory: X) + 0 + 0

= 0.2\*V3

= 0.2373

Winner: Margin\_Subcategory

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Type Overall Type Arrangement Margin Margin sub-category Count

Dogwood Simple Opposite Entire X         2

Maple Simple Opposite Lobed X 2

Ash Compound Opposite X X 40

Hickey Compound X Toothed X 2

Locust Compound Alternate Entire X 2

Cherry Simple Alternate Toothed X 2

White Oak Simple Alternate Lobed Rounded 100

Red Oak Simple Alternate Lobed Pointed 100

QUESTION 2

I({Q}-> Type) = -sum\_Type Pr(Type|{Q})Log\_2 Pr(Type|{Q})

E[I({Q}-> Type)] = \sum\_Q I({Q}-> Type) Pr({Q})

I({Margin sub-category: X, Margin: Entire} -> Type) = 2\* (2/4)\*Log\_2(2) = 1

I({Margin sub-category: X, Margin: Lobed} -> Type) = 0

I({Margin sub-category: X, Margin: X} -> Type) = 0

I({Margin sub-category: X, Margin: Toothed} -> Type) = 1

I({Margin sub-category: Rounded, Margin: Entire} -> Type) = 0

I({Margin sub-category: Rounded, Margin: Lobed} -> Type) = 0

I({Margin sub-category: Rounded, Margin: X} -> Type) = 0

I({Margin sub-category: Rounded, Margin: Toothed} -> Type) = 0

I({Margin sub-category: Pointed, Margin: Entire} -> Type) = 0

I({Margin sub-category: Pointed, Margin: Lobed} -> Type) = 0

I({Margin sub-category: Pointed, Margin: X} -> Type) = 0

I({Margin sub-category: Pointed, Margin: Toothed} -> Type)  = 0‘

E(I(Q-> Type)) = 1\*4/250+ 1\*4/250 = 0.032  = S1

I({Margin sub-category: X, ,Arrangement: Opposite} -> Type)

= 2\*2/44 Log\_2(44/2) + 40/44 Log-2(44/40)

=0.5304

I({Margin sub-category: X, Arrangement: X} -> Type)

= 0

I({Margin sub-category: X, Arrangement: Alternate} -> Type)

= 2\*2/4 Log\_2(4/2)

= 1

I({Margin sub-category: Pointed, Arrangement: Opposite} -> Type) = 0

I({Margin sub-category: Pointed, Arrangement: X} -> Type) = 0

I({Margin sub-category: Pointed, Arrangement: Alternate: X} -> Type) = 0

I({Margin sub-category: Rounded, Arrangement: Opposite} -> Type) = 0

I({Margin sub-category: Rounded, Arrangement: X} -> Type) = 0

I({Margin sub-category: Rounded, Arrangement: Alternate: X} -> Type) = 0

E(I(Q-> Type)) = S2 = 0.5304\*44/250 + 1\*4/250 = 0.10935

I({Margin sub-category: X, Overall Type: Simple } -> Type) = 3\*2/6\*Log\_2(3)

I({Margin sub-category: X, Overall Type: Compound} -> Type)

= 2\*2/44 \*Log\_2(44/2) + 40/44\*Log\_2(44/40)

=0.5034

I({Margin sub-category: Pointed, Overall Type: Simple } -> Type) =0

I({Margin sub-category: Pointed, Overall Type: Compound} -> Type) =0

I({Margin sub-category: Rounded, Overall Type: Simple } -> Type) =0

I({Margin sub-category: Rounded, Overall Type: Compound} -> Type) =0

E(I(Q-> Type)) = S3 = 0.5034\*44/250 = 0.08859

ArgMin (S1, S2, S3) = S1

Winner= Margin-Subcategory -> Margin

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Verify and confirm whether Margin is the best question for the second round.

Assume Margin is the best question for the second round.

I({Margin Sub-category, Margin, Arrangement} -> Type) … 36 of them

E(I({Margin Sub-category, Margin, Arrangement} -> Type))

I({Margin Sub-category, Margin, Overall type} -> Type) … 24 of them

E(I({Margin Sub-category, Margin, Overall type} -> Type))

Complete solution - final

