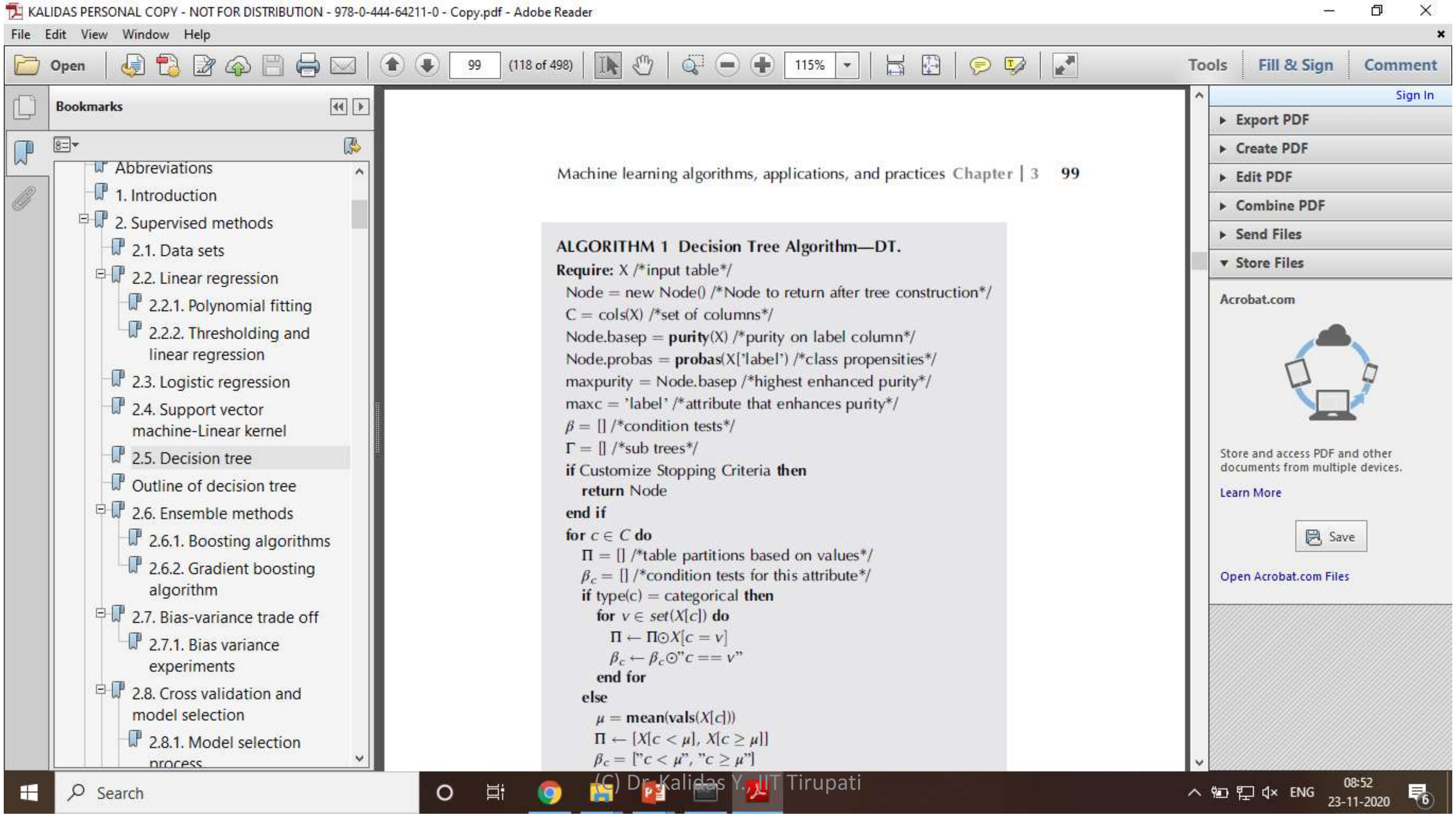
Decision Tree

Dr. Kalidas Y., IIT Tirupati

By the end of this lecture, you will understand Decision Tree based Classification and Regression formulation

100) key phrase… “Decision Tree”

a) Table, b) Numeric comparison, c) String equality and d) Loss function



101) key phrase… “impurity”

- Bag 1 – 5 Red balls, 0 Blue balls
- Bag 2 – 5 Red balls, 1 Blue ball
- Bag 3 – 5 Red balls, 2 Blue balls
- Bag 4 – 5 Red balls, 5 Blue balls
- Bag 5 – 0 Red balls, 5 Blue balls

Which is more ***pure***?

102) key phrase… “gini impurity index”

- $gi = p_{red} * (1 - p_{red}) + p_{blue} * (1 - p_{blue})$
- $gi = p_{red} * (1 - p_{red}) + p_{blue} * (1 - p_{blue}) + p_{green} * (1 - p_{green})$

And so on and so forth…

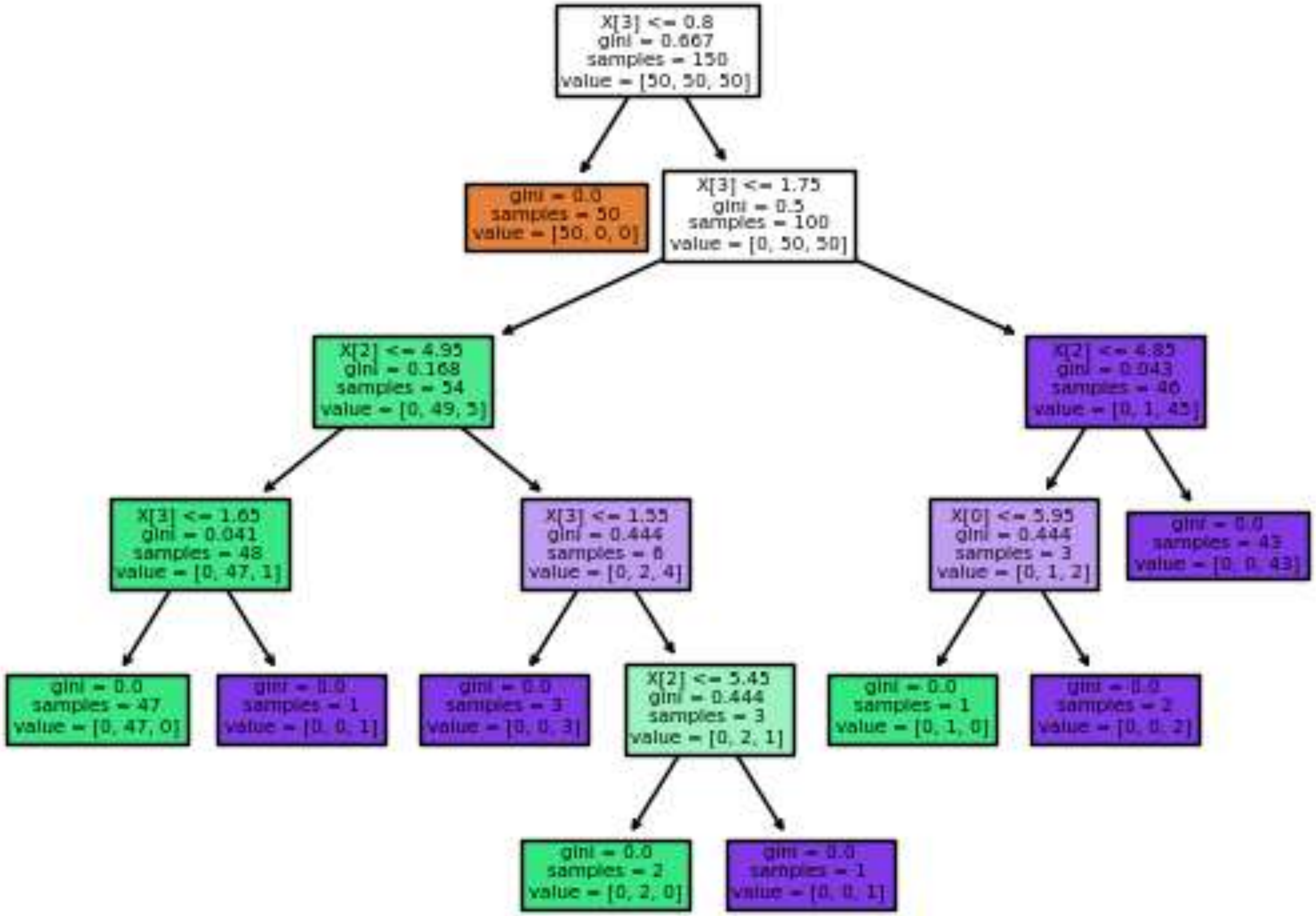
- k classes: $gi = \sum_{j=1}^{j=k} p_j * (1 - p_j)$

103) key phrase… “entropy”

- k classes: $E = -\sum_{j=1}^{j=k} p_j * \log(p_j)$

104) key phrase… “Misclassification”

- k classes: $M = 1 - \max(\{p_1, \ldots, p_k\})$


X[3] <= 0.8
gini = 0.667
samples = 150
value = [50, 50, 50]
gini = 0.0
samples = 50
value = [50, 0, 0]
X[3] <= 1.75
gini = 0.5
samples = 100
value = [0, 50, 50]
X[2] <= 4.95
gini = 0.168
samples = 54
value = [0, 49, 5]
X[2] <= 4.85
gini = 0.043
samples = 46
value = [0, 1, 45]
X[3] <= 1.65
gini = 0.041
samples = 48
value = [0, 47, 1]
X[3] <= 1.55
gini = 0.444
samples = 6
value = [0, 2, 4]
X[0] <= 5.95
gini = 0.444
samples = 3
value = [0, 1, 2]
gini = 0.0
samples = 43
value = [0, 0, 43]
gini = 0.0
samples = 47
value = [0, 47, 0]
gini = 0.0
samples = 1
value = [0, 0, 1]
gini = 0.0
samples = 3
value = [0, 0, 3]
X[2] <= 5.45
gini = 0.444
samples = 3
value = [0, 2, 1]
gini = 0.0
samples = 1
value = [0, 1, 0]
gini = 0.0
samples = 2
value = [0, 0, 2]
gini = 0.0
samples = 2
value = [0, 2, 0]
gini = 0.0
samples = 1
value = [0, 0, 1]

106) key phrase… “base impurity”

compute gini impurity for the label column of the “***base table***”

base impurity (bi) $= \frac{2}{7} * \left(1 - \frac{2}{7}\right) + \frac{2}{7} * \left(1 - \frac{2}{7}\right) + \frac{3}{7} * \left(1 - \frac{3}{7}\right) = \frac{32}{49} = 0.65$

A1	A2	A3	A4	A5	Label
?	cat	?	?	?	Red
?	cat	?	?	?	Red
?	cat	?	?	?	Blue
?	rat	?	?	?	Blue
?	dog	?	?	?	Green
?	dog	?	?	?	Green
?	dog	?	?	?	Green

107) key phrase… "attribute selection and sub-table"

$bi = 0.65$

Sub table: $A2 = cat$

gi(A2=cat): $\frac{2}{3} * \left(1 - \frac{2}{3}\right) + \frac{1}{3} * \left(1 - \frac{1}{3}\right) = \frac{4}{9} = 0.44$

A1	**A2**	**A3**	**A4**	**A5**	**Label**
?	cat	?	?	?	Red
?	cat	?	?	?	Red
?	cat	?	?	?	Blue

Pros: $0.44 < 0.65$

Cons: It's only $\frac{3}{7}$ of the *"base table"*

… "attribute selection and sub-table"

$bi = 0.65$

gi(A2)* = gi(A2=cat)**sup(A2=cat)* + gi(A2=rat)**sup(A2=rat)*+gi(A2=dog)**sup(A2=dog)

$bi = 0.65$
$gi(A2) = 0.19$

$gi(A2) < bi$

gi(A2=cat): $\frac{2}{3} * \left(1 - \frac{2}{3}\right) + \frac{1}{3} * \left(1 - \frac{1}{3}\right) = \frac{4}{9} = 0.44 * \frac{3}{7}$

gi(A2=rat): $\frac{1}{1} * \left(1 - \frac{1}{1}\right) = \frac{0}{1} = 0 * \frac{1}{7}$

gi(A2=dog): $\frac{3}{3} * \left(1 - \frac{3}{3}\right) = \frac{0}{9} = 0 * \frac{3}{7}$

gi(A2) = $0.44 * \frac{3}{7} + 0 * \frac{1}{7} + 0 * \frac{1}{7} = 0.19$

→ We can split the table based on A2

but wait…

108) key phrase… "weighted impurity"

$bi = 0.65$
$gi(A2) = 0.19$
$gi(A3) = 0.19$
$gi(A4) = 0.19$

How to handle numeric attributes???

A1	A2	A3	A4	A5	Label
0	cat	0	cat	0.1	Red
-10	cat	0	cat	0.1	Red
1.2	cat	0	cat	0.2	Blue
1.3	rat	1	rat	0.3	Blue
12	dog	2	dog	0.5	Green
200	dog	2	dog	0.6	Green
14	dog	2	eagle	0.6	Green

Numeric attributes – About the Mean Value

A1 – Mean value = $\frac{0+(-10)+1.2+1.3+12+200+14}{7} = 31.21$

gi(A1<31.21) = $\frac{2}{6} * \left(1 - \frac{2}{6}\right) + \frac{2}{6} * \left(1 - \frac{2}{6}\right) + \frac{2}{6} * \left(1 - \frac{2}{6}\right) = 0.22$

gi(A1) $= 0.22 * \frac{6}{7} + 0 * \frac{1}{7} = 0.19$

A1	A2	A3	A4	A5	Label
0	cat	0	cat	0.1	Red
-10	cat	0	cat	0.1	Red
1.2	cat	0	cat	0.2	Blue
1.3	rat	1	rat	0.3	Blue
12	dog	2	dog	0.5	Green
14	dog	2	eagle	0.6	Green

gi(A1>=31.21) = $\frac{1}{1} * \left(1 - \frac{1}{1}\right) = 0$

A1	A2	A3	A4	A5	Label
200	dog	2	dog	0.6	Green

… "weighted impurity"

$bi = 0.65$
$gi(A2) = 0.19$
$gi(A3) = 0.19$
$gi(A4) = 0.19$

$gi(A1) = 0.19$
$gi(A5) = 0.14$

A1	A2	A3	A4	A5	Label
0	cat	0	cat	0.1	Red
-10	cat	0	cat	0.1	Red
1.2	cat	0	cat	0.2	Blue
1.3	rat	1	rat	0.3	Blue
12	dog	2	dog	0.5	Green
200	dog	2	dog	0.6	Green
14	dog	2	eagle	0.6	Green

109) key phrase… “Information Gain”

- information gain = base impurity – attribute impurity
  - ig(A1) = 0.65 – 0.19
  - ig(A2) = 0.65 – 0.19
  - ig(A3) = 0.65 – 0.19
  - ig(A4) = 0.65 – 0.19
  - ig(A5) = 0.65 – 0.15
- Select the attribute with maximal information gain – “A5”

see the magic… next!

Recursive Decision Tree

- Decision Tree (base table T)

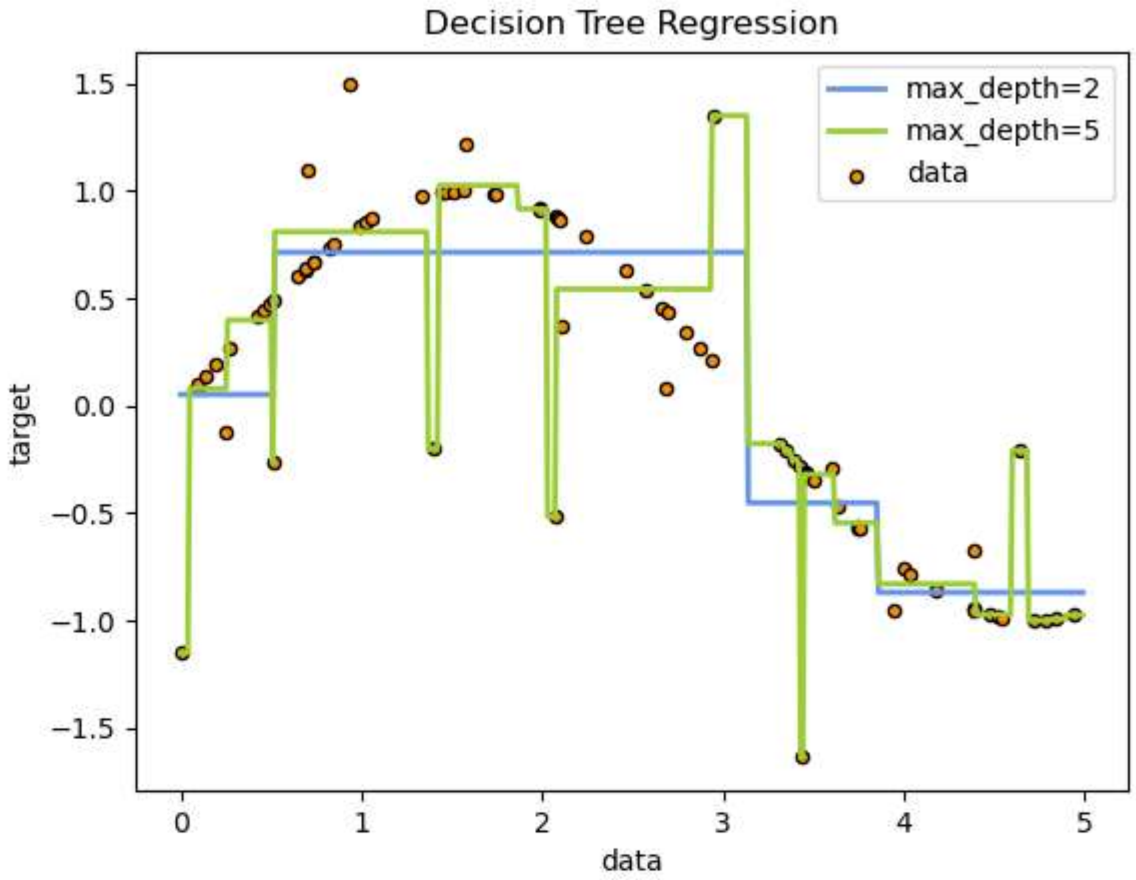
1. Compute – bi

2. For Each Attribute Ax, compute Information Gain – ig(Ax)

3. Select Attribute which give Maximal Information Gain.. Ay = argmax_Az ig(Az)

4. Split the table
   1. Sub Tables = []
   2. If Ay is numeric, Sub Tables = [T[Ay<Mean], T[Ay>=Mean]]
   3. If Ay is categorical {val1,…,valx}, Sub Tables = [T[Ay=val1], T[Ay=val2], … , T[Ay=valx]]

5. For each subtab ∈ Sub Tables, recursively do Decision Tree (subtab)


Decision Tree Regression
max_depth=2
max_depth=5
data
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
target
0
1
2
3
4
5
data

110) key phrase… “Decision Tree Regressor”

- ~~Categorical Impurity Index or Error Function~~
- Numeric Impurity Function – ***Variance!***

A1	A2	A3	A4	A5	Label
0	cat	0	cat	0.1	0.1
-10	cat	0	cat	0.1	0.2
1.2	cat	0	cat	0.2	1.0
1.3	rat	1	rat	0.3	1.3
12	dog	2	dog	0.5	30.5
14	dog	2	eagle	0.6	40.1

110) key phrase… “Decision Tree Regressor”

- ***Base Variance: 429.16***

A1	A2	A3	A4	A5	Label
0	cat	0	cat	0.1	10.1
-10	cat	0	cat	0.1	0.2
1.2	cat	0	cat	0.2	-1.0
1.3	rat	1	rat	0.3	1.3
12	dog	2	dog	0.5	-30.5
14	dog	2	eagle	0.6	40.1

Attribute Specific Variance – Example categorical

Base Variance: 429.16

H(A2=cat) = Var({10.1, 0.2, -1.0}) = 24.74

H(A2=rat) = Var({1.3}) = 0

H(A2=dog) = Var({-30.5, 40.1}) = 1246.1

Base Variance: 429.16

*H(A2=cat) = Var({10.1, 0.2, -1.0}) = 24.74 * sup(A2=cat)*

*H(A2=rat) = Var({1.3}) = 0 * sup(A2=rat)*

*H(A2=dog) = Var({-30.5, 40.1}) = 1246.1 * sup(A2=dog)*

Weighted Variance of A2: V(A2) $=24.74 * \frac{3}{6} + 0 * \frac{1}{6} + 1246.1 * \frac{2}{6} = 427.74$

A1	A2	A3	A4	A5	Label
0	cat	0	cat	0.1	10.1
-10	cat	0	cat	0.1	0.2
1.2	cat	0	cat	0.2	-1.0
1.3	rat	1	rat	0.3	1.3
12	dog	2	dog	0.5	-30.5
14	dog	2	eagle	0.6	40.1

Attribute Specific Variance – Example Numeric

Base Variance: 429.16

H(A1<3.1) = Var({10.1, 0.2, -1.0, 1.3}) = 19.16

H(A1>=3.1) = Var({-30.5, 40.1}) =1246.1

Base Variance: 429.16

V(A1) = $19.16 * \frac{4}{6} + 1246.1 * \frac{2}{6} = 428.14$

A1	**A2**	**A3**	**A4**	**A5**	**Label**
0	cat	0	cat	0.1	10.1
-10	cat	0	cat	0.1	0.2
1.2	cat	0	cat	0.2	-1.0
1.3	rat	1	rat	0.3	1.3
12	dog	2	dog	0.5	-30.5
14	dog	2	eagle	0.6	40.1

Decision Tree

Looks great isn't it? ;-)

Decision Tree Regression – Multi Variate

types?

A1	A2	A3	Y1	Y2
cat	0	0	0.1	-0.1
cat	0.3	0	0.3	0.4
cat	2.4	0	-0.2	0.9
rat	1.2	1	1.5	0.7
dog	-9.1	2	3	2.7
dog	8.3	2	1.9	3.3

Decision Tree Regression – Multi Variate

Row#	A1	A2	A3	Y1	Y2
0	cat	0	0	0.1	-0.1
1	cat	0.3	0	0.3	0.4
2	cat	2.4	0	-0.2	0.9
3	rat	1.2	1	1.5	0.7
4	dog	-9.1	2	3	2.7
5	dog	8.3	2	1.9	3.3

var*(A1=cat): var(Y1[0,1,2])+var(Y2[0,1,2])=0.21

var*(A1=rat): var(Y1[3])+var(Y2[3])=0

var*(A1=dog):var(Y1[4,5])+var(Y2[4,5])=9.1

Y_mean = $(1.1, 1.32) = \left(\sum_{i=0}^{i=5} Y_i[0], \sum_{i=0}^{i=5} Y_i[1]\right)$

base variance*: var(Y1)+var(Y2)

var* =

1) compute variance for each dimension np.var(Y,axis=0)
2) sum those variances, np.sum(np.var(Y,axis=0))

This is equivalent to,
average distance from centroid of all points.

Decision Tree Regression STEPs

- STEP 1: For the given base table, compute "base variance"
- STEP 2: For each attribute compute weighted variance
- STEP 3: Select the attribute – maximal variance reduction
- STEP 4: Split the table based on that attribute
- STEP 5: Recursion: For each sub-table, do Decision Tree Regression

Decision Tree – Prediction???

- Identify the leaf node
- ***Classification***: Propensity of all classes in the leaf
- ***Regression***: *Mean value* in the leaf

- Multi-variate regression???
- ***Mean value itself is multivariate***

Decision Tree Hyper Parameters

- Tree Depth
- Minimum number of entities in a leaf
- Impurity function