JOHNS HOPKINS
WHITING SCHOOL of ENGINEERING

Introduction to Machine Learning
Decision Tree Learning

Prof. John W. Sheppard

JOHNS HOPKINS
WHITING SCHOOL *of* ENGINEERING

Handling continuous variables

Purity/Impurity Measures

- Attribute selection in decision trees is based on different purity/impurity measures.
- Here we refresh the two most common.
- Classification: Information gain

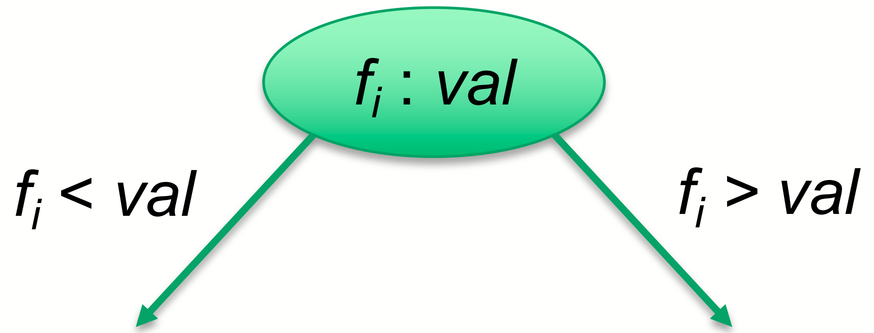
$$f_{sel} = \arg\max gain(f_i)$$
$$gain(f_i) = H(\mathcal{D}_{part}) - \sum_{val(f_i)} P\left(\mathcal{D}_{part}^{val(f_i)}\right) H(\mathcal{D}_{part}^{val(f_i)} | val(f_i))$$

- Regression: Mean squared error

$$f_{sel} = \arg\max mseRed(f_i)$$
$$mseRed = MSE\left(\mathcal{D}_{part}\right) - MSE(\mathcal{D}_{part}^{f_i})$$

Continuous Attributes

- Standard decision trees assume discrete attributes.
- Decision trees are able to handle continuous attributes as well.
- Generally, when dealing with continuous attributes, binary splits are generated for those attributes.



Continuous Attributes – Classification

f_i	0.1	0.3	0.4	0.9	1.3	1.8	2.5	3.7	4.1	4.2	4.4	5.9	6.3	7.1	7.7

f_i : 0.65

f_i : 3.9

f_i : 6.1

Continuous Attributes – Classification

f_i	0.1	0.3	0.4	0.9	1.3	1.8	2.5	3.7	4.1	4.2	4.4	5.9	6.3	7.1	7.7

f_i : 0.0…0.65

f_i : 0.65…3.9

f_i : 0.3.9…6.1

f_i : 6.1…8.0

Continuous Attributes – Regression

Equal width: four bins, every 1.9 from baseline

f_i	0.1	0.3	0.4	0.9	1.3	1.8	2.5	3.7	4.1	4.2	4.4	5.9	6.3	7.1	7.7

f_i : 2.0 f_i : 3.9 f_i : 5.8

Continuous Attributes – Regression

Equal frequency: four bins, every 4 points from baseline

f_i	0.1	0.3	0.4	0.9	1.3	1.8	2.5	3.7	4.1	4.2	4.4	5.9	6.3	7.1	7.7

f_i : 1.1

f_i : 3.9

f_i : 6.1

What Needs to be Tuned?

- Classification
  - Whether or not to consider all changes in class.
  - If not, how many changes?
- Regression – equal width
  - The number of bins
- Regression – equal frequency
  - The number of points in the bin

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING