Machine Learning

Learning with Decision Tree – Regression

Satishkumar L. Varma

Department of Information Technology
SVKM's Dwarkadas J. Sanghvi College of Engineering, Vile Parle, Mumbai.
ORCID | Scopus | Google Scholar | Google Site | Website

Outline

- Learning with Regression and Trees
  - Learning with Regression
    - Simple Linear Regression
    - Multiple Linear Regression
    - Logistic Regression
  - Learning with Trees
    - Decision Trees
    - Constructing Decision Trees using Gini Index
    - Classification and Regression Trees (CART)

Learning with Decision Tree – Regression

- Regression/Prediction Step
- Here's how a regression tree makes predictions for new data:
- 1. Start at the top (root) of the tree.
- 2. At each decision point (node):
- – Look at the feature and split value.
- – If the data point's feature value is smaller or equal, go left.
- – If it's larger, go right.
- 3. Keep moving down the tree until you reach the end (a leaf).
- 4. The prediction is the average value stored in that leaf.

Learning with Decision Tree – Regression

- Decision Tree – Regression

Predictors				Target
Outlook	Temp	Humidity	Windy	Hours Played
Rainy	Hot	High	False	25
Rainy	Hot	High	True	30
Overcast	Hot	High	False	48
Sunny	Mild	High	False	46
Sunny	Cool	Normal	False	52
Sunny	Cool	Normal	True	23
Overcast	Cool	Normal	True	43
Rainy	Mild	High	False	35
Rainy	Cool	Normal	False	38
Sunny	Mild	Normal	False	46
Rainy	Mild	Normal	True	48
Overcast	Mild	High	True	52
Overcast	Hot	Normal	False	44
Sunny	Mild	High	True	30

Learning with Decision Tree – Regression

- Predictive modeling is the process by which a model is created to predict an outcome.
- If the outcome is categorical it is called classification.
- &
- if the outcome is numerical it is called regression.
- Descriptive modeling or clustering is the assignment of observations
  - into clusters so that observations in the same cluster are similar.

Predictors				Target
Outlook	Temp	Humidity	Windy	Hours Played
Rainy	Hot	High	False	25
Rainy	Hot	High	True	30
Overcast	Hot	High	False	48
Sunny	Mild	High	False	46
Sunny	Cool	Normal	False	62
Sunny	Cool	Normal	True	23
Overcast	Cool	Normal	True	43
Rainy	Mild	High	False	36
Rainy	Cool	Normal	False	38
Sunny	Mild	Normal	False	48
Rainy	Mild	Normal	True	48
Overcast	Mild	High	True	52
Overcast	Hot	Normal	False	44
Sunny	Mild	High	True	30

Learning with Decision Tree – Regression

- Decision tree builds regression or classification models in the form of a tree structure.
- It breaks down a dataset into smaller and smaller subsets
  - while at the same time an associated decision tree is incrementally developed.
- The final result is a tree with decision nodes and leaf nodes.

Predictors				Target
Outlook	Temp	Humidity	Windy	Hours Played
Rainy	Hot	High	False	25
Rainy	Hot	High	True	30
Overcast	Hot	High	False	48
Sunny	Mild	High	False	46
Sunny	Cool	Normal	False	62
Sunny	Cool	Normal	True	23
Overcast	Cool	Normal	True	43
Rainy	Mild	High	False	36
Rainy	Cool	Normal	False	38
Sunny	Mild	Normal	False	48
Rainy	Mild	Normal	True	48
Overcast	Mild	High	True	52
Overcast	Hot	Normal	False	44
Sunny	Mild	High	True	30

Learning with Decision Tree – Regression

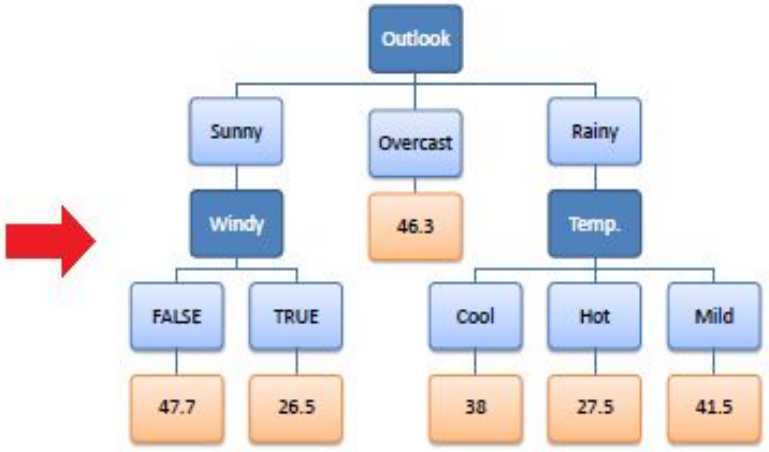
- A decision node (e.g., Outlook) has two or more branches (e.g., Sunny, Overcast and Rainy),
  - each representing values for the attribute tested.
- Leaf node (e.g., Hours Played) represents a decision on the numerical target.
- The topmost decision node in a tree which corresponds to the best predictor called root node.
- Decision trees can handle both categorical and numerical data.

Predictors				Target
Outlook	Temp.	Humidity	Windy	Hours Played
Rainy	Hot	High	False	25
Rainy	Hot	High	True	30
Overcast	Hot	High	False	48
Sunny	Mild	High	False	46
Sunny	Cool	Normal	False	62
Sunny	Cool	Normal	True	23
Overcast	Cool	Normal	True	43
Rainy	Mild	High	False	36
Rainy	Cool	Normal	False	38
Sunny	Mild	Normal	False	48
Rainy	Mild	Normal	True	48
Overcast	Mild	High	True	52
Overcast	Hot	Normal	False	44
Sunny	Mild	High	True	30

Learning with Decision Tree – Regression

- Decision Tree Algorithm
- The core algorithm for building decision trees called ID3 (Iterative Dichotomiser 3) by J. R. Quinlan
  - It is employs a top-down, greedy search through the space of possible branches with no backtracking.
- The ID3 algorithm can be used to construct a decision tree for regression
  - by replacing Information Gain with Standard Deviation Reduction.

Predictors				Target
Outlook	Temp.	Humidity	Windy	Hours Played
Rainy	Hot	High	False	25
Rainy	Hot	High	True	30
Overcast	Hot	High	False	48
Sunny	Mild	High	False	46
Sunny	Cool	Normal	False	52
Sunny	Cool	Normal	True	23
Overcast	Cool	Normal	True	43
Rainy	Mild	High	False	35
Rainy	Cool	Normal	False	38
Sunny	Mild	Normal	False	48
Rainy	Mild	Normal	True	48
Overcast	Mild	High	True	52
Overcast	Hot	Normal	False	44
Sunny	Mild	High	True	30



Learning with Decision Tree – Regression

- Standard Deviation
- A decision tree is built top-down from a root node and involves partitioning the data into subsets
  - that contain instances with similar values (homogenous).
- We use standard deviation to calculate the homogeneity of a numerical sample.
- If the numerical sample is completely homogeneous its standard deviation is zero.
- **a) Standard deviation for one attribute:**

Hours Played
25
30
46
45
52
23
43
35
38
46
48
52
44
30

$$S = \sqrt{\frac{\sum (x-\mu)^2}{n}}$$

Standard Deviation

S = 9.32

Learning with Decision Tree – Regression

- b) Standard deviation for two attributes:

$$S(T,X) = \sum_{c \in X} P(c)S(c)$$

		Hours Played (StDev)	Count
Outlook	Overcast	3.49	4
	Rainy	7.78	5
	Sunny	10.87	5
			14

S(Hours, Outlook) = **P**(Sunny)*S(Sunny) + **P**(Overcast)*S(Overcast) + **P**(Rainy)*S(Rainy)

= (4/14)*3.49 + (5/14)*7.78 + (5/14)*10.87 = 7.66

Learning with Decision Tree – Regression

- Standard Deviation Reduction
- The standard deviation reduction is based on the decrease in standard deviation after a dataset is split on an attribute.
- Constructing a decision tree is all about finding attribute that returns the highest standard deviation reduction (i.e., the most homogeneous branches).
- Step 1: The standard deviation of the target is calculated.
- Standard deviation (Hours Played) = 9.32

Learning with Decision Tree – Regression

- Step 2: The dataset is then split on the different attributes. The standard deviation for each branch is calculated.
- The resulting standard deviation is subtracted from the standard deviation before the split.
- The result is the standard deviation reduction.

		Hours Played (StDev)
Outlook	Overcast	3.49
	Rainy	7.78
	Sunny	10.87
SDR=1.66		

		Hours Played (StDev)
Temp.	Cool	10.51
	Hot	8.95
	Mild	7.65
SDR=0.17		

		Hours Played (StDev)
Humidity	High	9.36
	Normal	8.37
SDR=0.28		

		Hours Played (StDev)
Windy	False	7.87
	True	10.59
SDR=0.29		

$$SDR(T,X) = S(T) - S(T,X)$$

SDR(Hours , Outlook) = **S**(Hours) – **S**(Hours, Outlook)

= 9.32 – 7.66 = 1.66

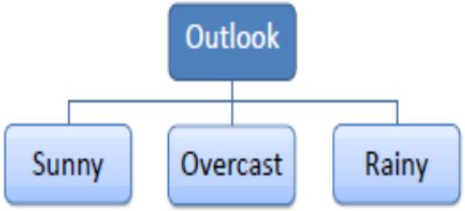
Learning with Decision Tree – Regression

- Step 3: The attribute with the largest standard deviation reduction is chosen for the decision node.

★		Hours Played (StDev)
Outlook	Overcast	3.49
	Rainy	7.78
	Sunny	10.87
SDR=1.66		

Learning with Decision Tree – Regression

- Step 4a: Dataset is divided based on the values of the selected attribute.



Outlook	Temp	Humidity	Windy	Hours Played
Sunny	Mild	High	FALSE	45
Sunny	Cool	Normal	FALSE	52
Sunny	Cool	Normal	TRUE	23
Sunny	Mild	Normal	FALSE	46
Sunny	Mild	High	TRUE	30
Rainy	Hot	High	FALSE	25
Rainy	Hot	High	TRUE	30
Rainy	Mild	High	FALSE	35
Rainy	Cool	Normal	FALSE	38
Rainy	Mild	Normal	TRUE	48
Overcast	Hot	High	FALSE	46
Overcast	Cool	Normal	TRUE	43
Overcast	Mild	High	TRUE	52
Overcast	Hot	Normal	FALSE	44

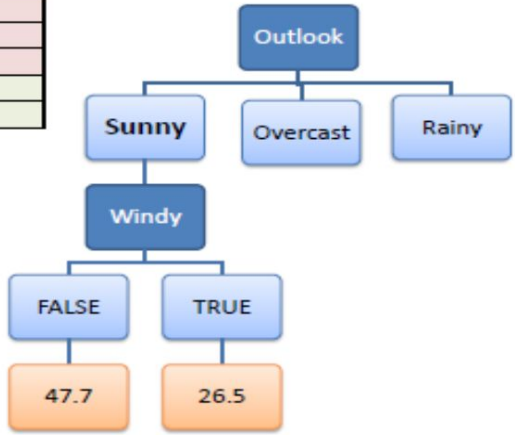
Learning with Decision Tree – Regression

- Step 4b: A branch set with standard deviation more than 0 needs further splitting.
- In practice, we need some termination criteria.
- For example,
  - when standard deviation for the branch becomes smaller than a certain fraction (e.g., 5%) of standard deviation for the full dataset
  - OR when too few instances remain in the branch (e.g., 3).

Temp	Humidity	Windy	Hours Played
Mild	High	FALSE	45
Cool	Normal	FALSE	52
Mild	Normal	FALSE	46
Cool	Normal	TRUE	23
Mild	High	TRUE	30

★		Hours Played (StDev)
Windy	False	3.09
	True	3.50
SDR= 7.62		

SDR = 10.87-((3/5)*3.09 + (2/5)*3.5)



Learning with Decision Tree – Regression

- Step 5: The process is run recursively on the non-leaf branches, until all data is processed.
- When the number of instances is more than one at a leaf node
  - we calculate the average as the final value for the target.

References

Text books:

1. Ethem Alpaydin, “Introduction to Machine Learning , 4th Edition, The MIT Press, 2020.
2. Peter Harrington, “Machine Learning in Action”, 1st Edition, Dreamtech Press, 2012."
3. Tom Mitchell, “Machine Learning”, 1st Edition, McGraw Hill, 2017.
4. Andreas C, Müller and Sarah Guido, “Introduction to Machine Learning with Python: A Guide for Data Scientists”, 1ed, O'reilly, 2016.
5. Kevin P. Murphy, “Machine Learning: A Probabilistic Perspective”, 1st Edition, MIT Press, 2012."

Reference Books:

6. Aurélien Géron, “Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow”, 2nd Edition, Shroff/O'Reilly, 2019.
7. Witten Ian H., Eibe Frank, Mark A. Hall, and Christopher J. Pal., “Data Mining: Practical machine learning tools and techniques”, 1st Edition, Morgan Kaufmann, 2016.
8. Han, Kamber, “Data Mining Concepts and Techniques”, 3rd Edition, Morgan Kaufmann, 2012.
9. Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar, “Foundations of Machine Learning”, 1ed, MIT Press, 2012.
10. H. Dunham, “Data Mining: Introductory and Advanced Topics”, 1st Edition, Pearson Education, 2006.

Thank You.