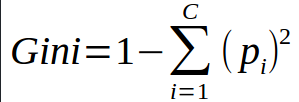
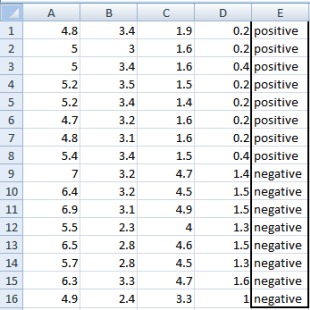
**Gini Index**



Gini Index is a metric to measure how often a randomly chosen element would be incorrectly identified. It means an attribute with lower gini index should be preferred.

**Example: Construct a Decision Tree by using “gini index” as a criterion**

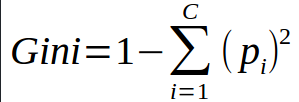


We are going to use same data sample that we used for information gain example. Let’s try to use gini index as a criterion. Here, we have 5 columns out of which 4 columns have continuous data and 5th column consists of class labels.

A, B, C, D attributes can be considered as predictors and E column class labels can be considered as a target variable. For constructing a decision tree from this data, we have to convert continuous data into categorical data.

We have chosen some **random values** to categorize each attribute:

|  |  |  |  |
| --- | --- | --- | --- |
| **A** | **B** | **C** | **D** |
| >= 5 | >= 3.0 | >=4.2 | >= 1.4 |
| < 5 | < 3.0 | < 4.2 | < 1.4 |



**Gini Index for Var A**

Var A has value >=5 for 12 records out of 16 and 4 records with value <5 value.

* For Var A >= 5 & class == positive: 5/12
* For Var A >= 5 & class == negative: 7/12
  + gini(5,7) = 1- ( (5/12)2 + (7/12)2 ) = 0.4860
* For Var A <5 & class == positive: 3/4
* For Var A <5 & class == negative: 1/4
  + gini(3,1) = 1- ( (3/4)2 + (1/4)2 ) = 0.375

By adding weight and sum each of the gini indices:

\textrm{gini(Target, A) = (12/16) * (0.486) + (4/16) * (0.375) = 0.45825}

**Gini Index for Var B**

Var B has value >=3 for 12 records out of 16 and 4 records with value <5 value.

* For Var B >= 3 & class == positive: 8/12
* For Var B >= 3 & class == negative: 4/12
  + gini(8,4) = 1- ( (8/12)2 + (4/12)2 ) = 0.446
* For Var B <3 & class == positive: 0/4
* For Var B <3 & class == negative: 4/4
  + gin(0,4) = 1- ( (0/4)2 + (4/4)2 ) = 0

\textrm{gini(Target, B) = (12/16) * 0.446 + (4/16) * 0 = 0.3345}

**Gini Index for Var C**

Var C has value >=4.2 for 6 records out of 16 and 10 records with value <4.2 value.

* For Var C >= 4.2 & class == positive: 0/6
* For Var C >= 4.2 & class == negative: 6/6
  + gini(0,6) = 1- ( (0/8)2 + (6/6)2 ) = 0
* For Var C < 4.2& class == positive: 8/10
* For Var C < 4.2 & class == negative: 2/10
  + gin(8,2) = 1- ( (8/10)2 + (2/10)2 ) = 0.32

\textrm{gini(Target, C) = (6/16) * 0+ (10/16) * 0.32 = 0.2} 

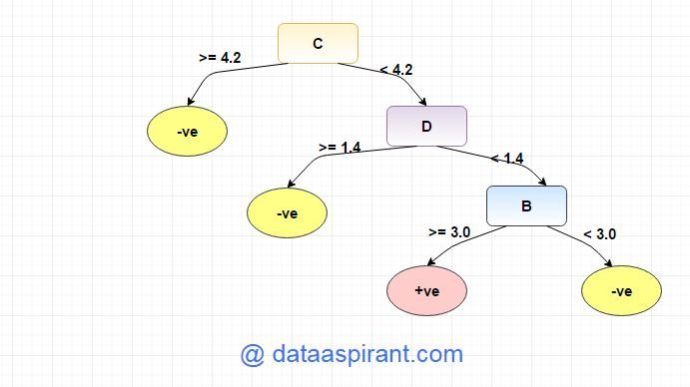
**Gini Index for Var D**

Var D has value >=1.4 for 5 records out of 16 and 11 records with value <1.4 value.

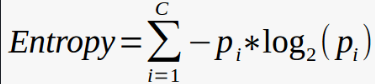
* For Var D >= 1.4 & class == positive: 0/5
* For Var D >= 1.4 & class == negative: 5/5
  + gini(0,5) = 1- ( (0/5)2 + (5/5)2 ) = 0
* For Var D < 1.4 & class == positive: 8/11
* For Var D < 1.4 & class == negative: 3/11
  + gin(8,3) = 1- ( (8/11)2 + (3/11)2 ) = 0.397

\textrm{ gini(Target, D) = (5/16) * 0+ (11/16) * 0.397 = 0.273} 

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  | | --- | --- | --- | --- | |  | | wTarget | | | Positive | Negative | | A | >= 5.0 | 5 | 7 | | <5 | 3 | 1 | | Ginin Index of A = 0.45825 | | | | | |  |  |  |  | | --- | --- | --- | --- | |  | | Target | | | Positive | Negative | | B | >= 3.0 | 8 | 4 | | < 3.0 | 0 | 4 | | Gini Index of B= 0.3345 | | | | |
| |  |  |  |  | | --- | --- | --- | --- | |  | | Target | | | Positive | Negative | | C | >= 4.2 | 0 | 6 | | < 4.2 | 8 | 2 | | Gini Index of C= 0.2 | | | | | |  |  |  |  | | --- | --- | --- | --- | |  | | Target | | | Positive | Negative | | D | >= 1.4 | 0 | 5 | | < 1.4 | 8 | 3 | | Gini Index of D= 0.273 | | | | |

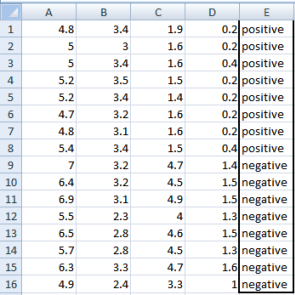


Entropy





**Example: Construct a Decision Tree by using “information gain” as a criterion**

We are going to use this data sample. Let’s try to use information gain as a criterion. Here, we have 5 columns out of which 4 columns have continuous data and 5th column consists of class labels.

A, B, C, D attributes can be considered as predictors and E column class labels can be considered as a target variable. For constructing a decision tree from this data, we have to convert continuous data into categorical data.

We have chosen some random values to categorize each attribute:

|  |  |  |  |
| --- | --- | --- | --- |
| **A** | **B** | **C** | **D** |
| >= 5 | >= 3.0 | >= 4.2 | >= 1.4 |
| < 5 | < 3.0 | < 4.2 | < 1.4 |

There are **2 steps for calculating information gain** for each attribute:

1. Calculate entropy of Target.
2. Entropy for every attribute A, B, C, D needs to be calculated. Using information gain formula we will subtract this entropy from the entropy of target. The result is Information Gain.

**The entropy of Target:** We have 8 records with negative class and 8 records with positive class. So, we can directly estimate the entropy of target as 1.

|  |  |
| --- | --- |
| Variable E | |
| Positive | Negative |
| 8 | 8 |

**Calculating entropy using formula:**

E(8,8) = -1\*( (p(+ve)\*log( p(+ve)) + (p(-ve)\*log( p(-ve)) )  
= -1\*( (8/16)\*log2(8/16)) + (8/16) \* log2(8/16) )  
= 1

**Information gain for Var A**



Var A has value >=5 for 12 records out of 16 and 4 records with value <5 value.

* For Var A >= 5 & class == positive: 5/12
* For Var A >= 5 & class == negative: 7/12
  + Entropy(5,7) = -1 \* ( (5/12)\*log2(5/12) + (7/12)\*log2(7/12)) = 0.9799
* For Var A <5 & class == positive: 3/4
* For Var A <5 & class == negative: 1/4
  + Entropy(3,1) =  -1 \* ( (3/4)\*log2(3/4) + (1/4)\*log2(1/4)) = 0.81128

Entropy(Target, A) = P(>=5) \* E(5,7) + P(<5) \* E(3,1)  
= (12/16) \* 0.9799 + (4/16) \* 0.81128 = 0.937745

\textrm{Information Gain(IG) = E(Target) - E(Target,A) = 1- 0.9337745 = 0.062255}  

**Information gain for Var B**

Var B has value >=3 for 12 records out of 16 and 4 records with value <5 value.

* For Var B >= 3 & class == positive: 8/12
* For Var B >= 3 & class == negative: 4/12
  + Entropy(8,4) = -1 \* ( (8/12)\*log2(8/12) + (4/12)\*log2(4/12)) = 0.39054
* For VarB <3 & class == positive: 0/4
* For Var B <3 & class == negative: 4/4
  + Entropy(0,4) =  -1 \* ( (0/4)\*log2(0/4) + (4/4)\*log2(4/4)) = 0

Entropy(Target, B) = P(>=3) \* E(8,4) + P(<3) \* E(0,4)  
= (12/16) \* 0.39054 + (4/16) \* 0 = 0.292905

\textrm{Information Gain(IG) = E(Target) - E(Target,B) = 1- 0.292905= 0.707095}  

**Information gain for Var C**

Var C has value >=4.2 for 6 records out of 16 and 10 records with value <4.2 value.

* For Var C >= 4.2 & class == positive: 0/6
* For Var C >= 4.2 & class == negative:  6/6
  + Entropy(0,6) = 0
* For VarC < 4.2 & class == positive: 8/10
* For Var C < 4.2 & class == negative: 2/10
  + Entropy(8,2) = 0.72193

Entropy(Target, C) = P(>=4.2) \* E(0,6) + P(< 4.2) \* E(8,2)  
= (6/16) \* 0 + (10/16) \* 0.72193 = 0.4512

\textrm{Information Gain(IG) = E(Target) - E(Target,C) = 1- 0.4512= 0.5488}  

**Information gain for Var D**

Var D has value >=1.4 for 5 records out of 16 and 11 records with value <5 value.

* For Var D >= 1.4 & class == positive: 0/5
* For Var D >= 1.4 & class == negative: 5/5
  + Entropy(0,5) = 0
* For Var D < 1.4 & class == positive: 8/11
* For Var D < 14 & class == negative: 3/11
  + Entropy(8,3) =  -1 \* ( (8/11)\*log2(8/11) + (3/11)\*log2(3/11)) = 0.84532

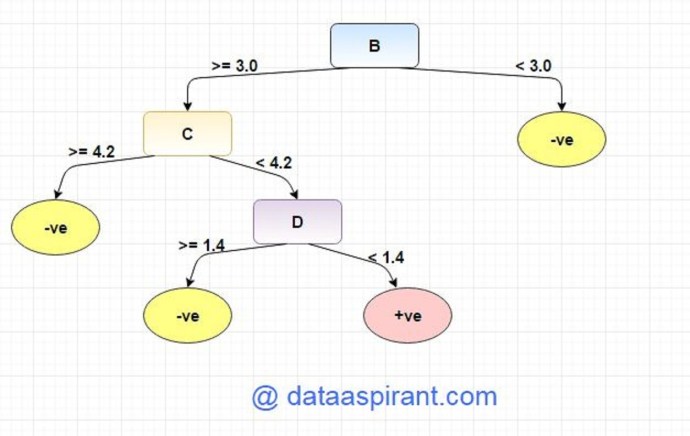
Entropy(Target, D) = P(>=1.4) \* E(0,5) + P(< 1.4) \* E(8,3)  
= 5/16 \* 0 + (11/16) \* 0.84532 = 0.5811575

\textrm{Information Gain(IG) = E(Target) - E(Target,D) = 1- 0.5811575 = 0.41189}  

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  | | --- | --- | --- | --- | |  | | **Target** | | | Positive | Negative | | A | >= 5.0 | 5 | 7 | | <5 | 3 | 1 | | Information Gain of A = 0.062255 | | | | | |  |  |  |  | | --- | --- | --- | --- | |  | | **Target** | | | Positive | Negative | | B | >= 3.0 | 8 | 4 | | < 3.0 | 0 | 4 | | Information Gain of B= 0.7070795 | | | | |
| |  |  |  |  | | --- | --- | --- | --- | |  | | **Target** | | | Positive | Negative | | C | >= 4.2 | 0 | 6 | | < 4.2 | 8 | 2 | | Information Gain of C= 0.5488 | | | | | |  |  |  |  | | --- | --- | --- | --- | |  | | **Target** | | | Positive | Negative | | D | >= 1.4 | 0 | 5 | | < 1.4 | 8 | 3 | | Information Gain of D= 0.41189 | | | | |

From the above Information Gain calculations, we can build a decision tree. We should place the attributes on the tree according to their values.

An Attribute with better value than other should position as root and A branch with entropy 0 should be converted to a leaf node. A branch with entropy more than 0 needs further splitting.



Tham khảo: <https://dataaspirant.com/2017/01/30/how-decision-tree-algorithm-works/>