# Data Classification with Decision Trees

**Structure of classifier**

* Subspaces of classifier do not have to be connected
* Border can be much more complex than single line

**Attribute selection**

* Fewer attrivutes decrease cost of learning
* A good attribute is one who carries a lot of “information”
* Find attributes with high entropy / high mutual information

## Information theory

Information content depends on log of probability distribution

**Entropy**

Diagram

Description automatically generated

**Mutual information**

Measue of how much of the information in two random variablesis sjared

*I(X,Y) = H(X) + H(Y) -H(X,Y)*

## Decision Trees

Text

Description automatically generated with low confidence