Gini Index and Gain Ratio

Gini Index

- Mainly used in CART algorithm(Classification and Regression Tree).
- Gini(s)=1-Type equation here.
- Consider a data set containing the attributes like amphibian,fish,bird,reptile and mammal.
- Suppose No. of example with class label amphibian=3

    No. of example with class label bird=2

    No. of example with class label fish=2

    No.of example with class label mammal=2

    No.of example with class label reptile=1

Gini Split Index

- Let S be a set of example, A be the feature/attribute.
- Gini

Gain Ratio:

- A problem occurs when information gain is applies to the attribute that can take on a large no.of distinct values ,ie information gain is baised towards choosing attributes with large no.of valus

- Gain Ration is am modification of information gain that reduces its bias by taking into account the no.of branches that would result before making split.