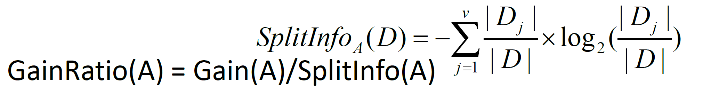
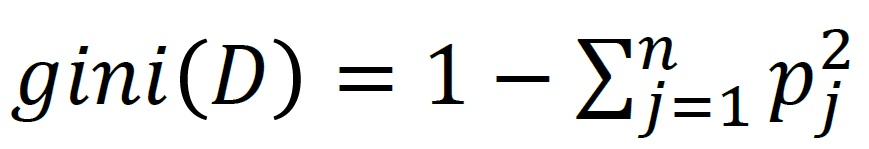
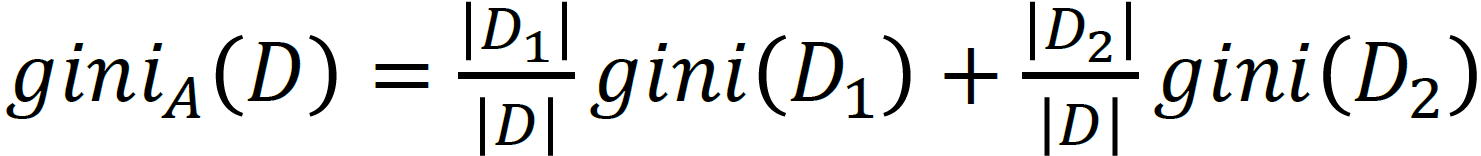
**Entropy:** Is biased toward multivalued attributes

**Gain Ratio:** Tends to prefer unbalanced splits in which one partition is much smaller than the others

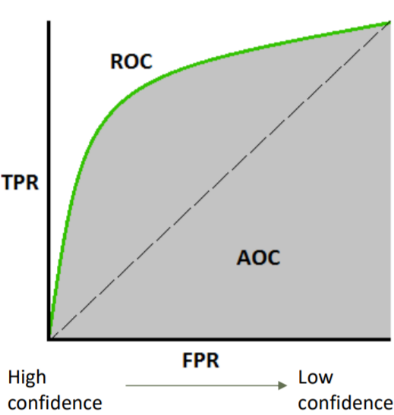
**Gini Index:** Is biased to multivalued attributes, Has difficulty when # of classes is large, Tends to favor tests that result in equal-sized partitions and purity in both partitions

* **Naïve Bayes Strength**: Easy to implement, Good results obtained in most of the cases
* **Naïve Bayes Weakness**: Assumption: Attributes conditional independence, therefore loss of accuracy,

**Bayesian Network:**

* A → B → C (Cascade): A and C are dependent, but (conditionally) independent if the value of B is given
* A ← B → C (Common Parent): A and C are dependent, but (conditionally) independent if the value of B is given
* A → B ← C (Common Child): A and C are by themselves independent, but conditioning on B will make A and C dependent

**Metrics**

* Accuracy = (TP + TN) / All <- Depends on positives
* Specificity = TN / N <- Not depends on positives
* Sensitivity/Recall = TP / P <- Depends on positives
* Precision = TP / (TP + FP) <- Depends on positives
* Precision on the negative label = TN / (TN + FN) <- Not depends on positives
* F1-measure = 2PR/(P + R) F\_B = (B^2 + 1)PR/(B^2\*P + R)

**ROC Curve**

* The ROC curve cannot be read from a single confusion matrix.
* One ROC curve can only be used to describe binary classification. For multi-class classification, can plot one VS one or one VS all ROC curves.

**Issues Affecting Model Selection:** Accuracy, Speed, Robustness, Scalability, Interpretability, goodness of rules

**Classification of Class-Imbalanced Data Sets:** Oversampling, Under-sampling, Threshold-moving, Ensemble techniques

**Random Forecast**: Comparable in accuracy to Adaboost, but more robust to errors and outliers. Insensitive to the number of attributes selected for consideration at each split, and faster than typical bagging or boosting

**Boosting**: Boosting tends to have greater accuracy, but it also risks overfitting the model to misclassified data

**Weak labeling**:

* Weak labels from crowd workers, output of heuristic rules, or the result of distant supervision (from KBs), or the output ofother classifiers, etc.
* Constraints and invariances (e.g., from physics, logic, or other experts)
* Probability distributions (e.g.,from weak or biased classifiers, user-providedlabels, feature expectations, or measurements)

**Classical Linear Classifiers:** Linear Discriminant Analysis (LDA), Logistic Regression, Perceptron, SVM

* **A generative classifier models p(Y,X)**: It models how the data was "generated,“ and what is the likelihood this or that class generated this instance, and pick the one with higher probability.
* **A discriminative classifier models p(Y|X)**: It uses the data to create a decision boundary
  + **Strength**: Prediction accuracy is generally high, Robust, works when training examples contain errors, Fast evaluation of the learned target function
  + **Criticism**: Long training time, Difficult to understand the learned function (weights), Not easy to incorporate domain knowledge

**SVM**

* **Strength**: SVM is effective on high dimensional data; Accuracy is high owing to their ability to model complex nonlinear decision boundaries; SVM can also be used for classifying multiple (> 2) classes and for regression analysis (with additional parameters)
* **Weakness**: SVM is not scalable to the number of data objects in terms of training time and memory usage. parameters of a solved model are difficult to interpret, Training can be slow

**CNN**: Three properties of CNN: local connectivity, parameter sharing and subsampling.

**Major obstacles of RNN:** Vanishing and exploding gradients. It is difficult to model long-range dependencies (10 timestamps or more)

**Why pattern-based classification?**

* Feature construction: Higher order; compact; discriminative
* Complex data modeling: Graphs, Sequences, Semi-structured/unstructured data

**DPClass framework**. Highlights: 1) only need a few discriminative patterns (e.g. 20) 2) able to handle categorical and numerical variables