Model Selection:
Incorporating Model Complexity

☐ Rationale: Occam's Razor

  – Given two models of similar generalization errors, one should prefer the simpler model over the more complex model

  – A complex model has a greater chance of being fitted accidentally by errors in data

  – Therefore, one should include model complexity when evaluating a model

Gen. Error(Model) = Train. Error(Model, Train. Data) +

α x Complexity(Model)

Estimating the Complexity of Decision Trees

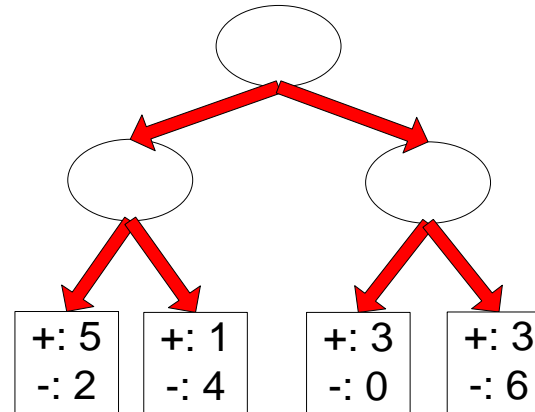☐ **Pessimistic Error Estimate** of decision tree T with k leaf nodes:

$$err_{gen}(T) = err(T) + \Omega \times \frac{k}{N_{train}}$$

– err(T): error rate on all training records

– Ω: trade-off hyper-parameter (similar to α)

◆ Relative cost of adding a leaf node

– k: number of leaf nodes

– N_{train}: total number of training records

Estimating the Complexity of Decision Trees: Example

$e(T_L) = 4/24$

$e(T_R) = 6/24$

$\Omega = 1$
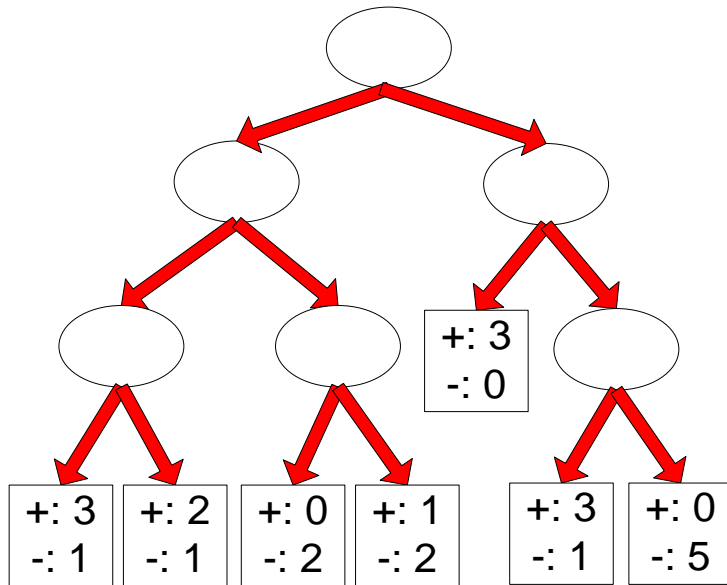
Decision Tree, T_L

Decision Tree, T_R
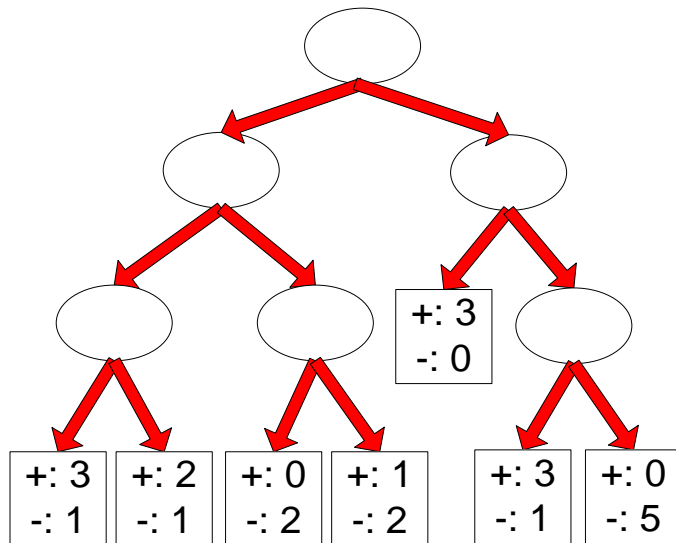
$e_{gen}(T_L) = 4/24 + 1*7/24 = 11/24 = 0.458$

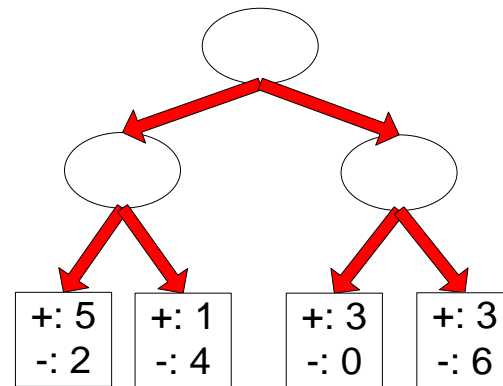$e_{gen}(T_R) = 6/24 + 1*4/24 = 10/24 = 0.417$

Estimating the Complexity of Decision Trees

☐ Resubstitution Estimate:

– Using training error as an optimistic estimate of generalization error

– Referred to as optimistic error estimate

$$e(T_L) = 4/24$$

$$e(T_R) = 6/24$$

Decision Tree, T_L

Decision Tree, T_R