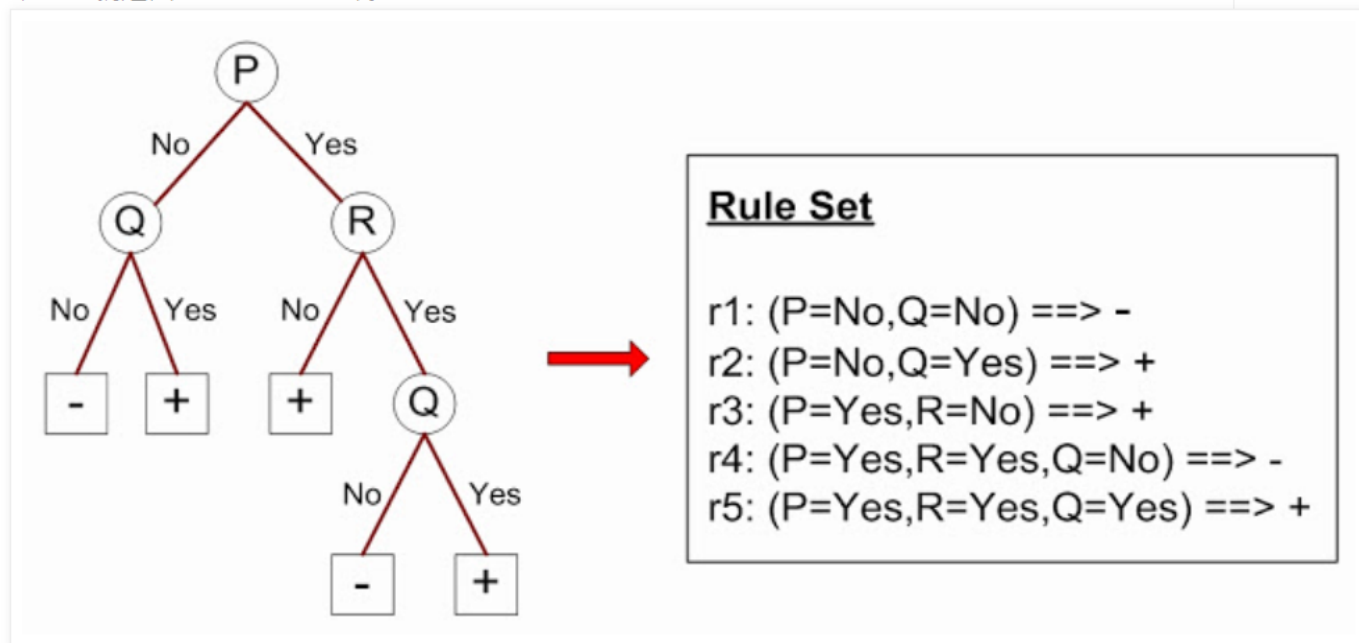


## 1. Rule-Based Classifier 規則分類的間接方法 ( Indirect Method )

上都提過，規則分類有直接與間接兩種方法，本周上課換介紹間接的方法 ( Indirect Method )。

所謂的間接方法，就不是直接拿資料去跑出規則集，而是換其他方式來生出規則集，比方說決策樹 ( Decision Tree )、類神經網路 ( Neural networks )

如下圖，就是由Decision Tree 轉 Rule-Based



像C4.5Rules就是一種間接的方法，先利用C4.5建立起**完整未修剪的決策樹**，在利用規則分類的修剪方式進行修剪動作。

而**C4.5Rules**規則產生的步驟如下：

1. 假設每個規則都是  $r: A \rightarrow y$
2. 經過條件篩檢後的規則為， $r': A' \rightarrow y$
3. 比較修剪前後的**悲觀估計** (回顧)
4. 如果有比較低的悲觀估計誤差的話，則修剪掉
5. 重複上述步驟，直到不再改善悲觀估計誤差。

而C4.5Rules的規則生成後，再來就要考慮規則的排序了。主要以規則**為排序** ( **Rule-based ordering** ) 的方式 (回顧)，相同分類的規則會放在同一個規則子集合中，在分別計算每個子集合的**總描述長度** ( **Description length** )，**以遞增方式做排序**。

規則總長度的計算方式如下：

$$\text{Description Length} = L_{\text{exception}} + g \times L_{\text{model}}$$

其中， $L$  ( exception ) 是指錯誤分類的資料 ( Misclassified example ) 所需的編碼長度； $g$ 是一個調整參數 ( Tuning parameter )，一般預設是0.5；而  $L$  ( model ) 整個模型的編碼長度。