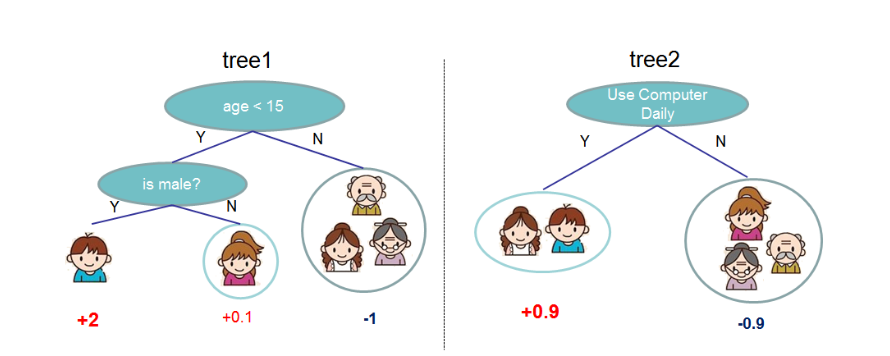
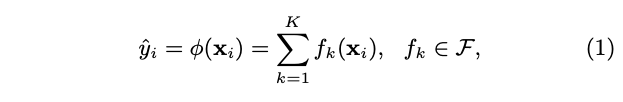
XGboost (eXtreme Gradient Boosting) 極限梯度提升

XGBoost即是**數顆**「分類與回歸樹(Classification and Regression Tree；CART)」組合成一個高準確率的模型。每加入一次新的函數至原模型中，修正上一顆樹的錯誤，不斷的迭代以提升整體的模型。如圖一所示，Tree1根據年紀與性別分類，藍衣小男孩所得的葉子分數為+2；Tree2為每日使用電腦習慣，小男孩所得的葉子分數為+0.9，因此最後的模型結果，小男孩總共獲得2.9分，反之，阿公的分數為-1.9，而這個累加結果則會作為預測結果。



圖一、數集成模型（Chen and Guestrin, 2016）。

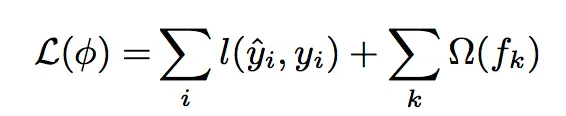
每棵樹的模型可被定義為：



(1)

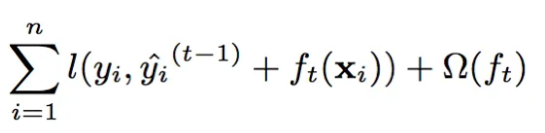
其中，，為權重、q為這棵樹中該葉子的index（這棵樹共有片葉子；）；為樹的總數量。換句話說，對於第i個樣本而言，在第k顆樹的中，所對應的葉子權重為；在第k+1顆樹的中，所對應的葉子權重為，而這個樣本所得的分數即為。

為了修正上顆樹的錯誤，則需要定義目標函數來優化，該函數可表示成：



(2)

前項為損失函數（loss function），用來量化預測與真實的差異；後項為懲罰項，其描述樹的複雜度，也用來避免迭代有overfitting的狀況出現。而XGB是基於「修正上一顆樹的錯誤」來求得最佳模型，因此，在第t次迭代中，都在修正第t-1次的迭代情形，故2式可寫成：



(3)

為了