

*ビッグデータとは？

名前の通り莫大なデータのこと。
インターネットの普及とIT機器の性能の向上化によってデータはデジタル化が急速に進み、ビッグデータはとも注目されている。



*R言語とは？

- ・R言語はオープンソース・フリーソフトウェアの統計解析向けのプログラミング言語。
- ・ビッグデータに注目が置かれている現在、R言語が使われることが多くなってきている。



*偽札-真札データ

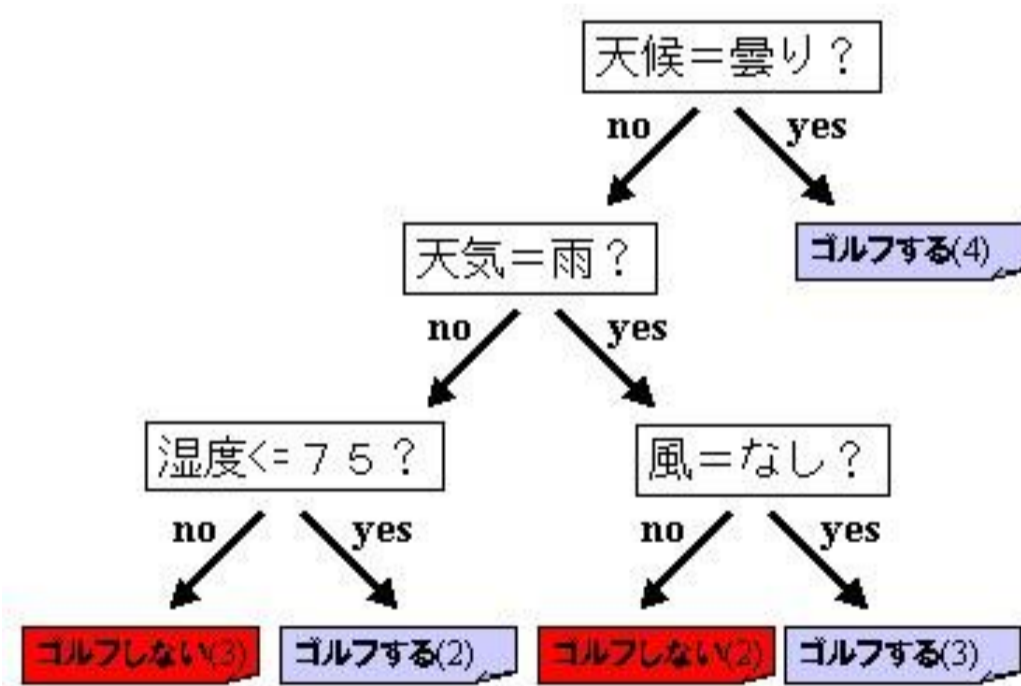
あるデータから偽札と真札の特徴を見分けてみよう！
今回のデータには「下部マージン」と「対角線」という特徴に注目します。



真偽	下部マージン	対角線
偽札	10.1	142
偽札	11	141.8
偽札	7.2	140.2
偽札	7.8	140
真札	9	141
真札	8.1	141.7
真札	8.7	142.2

*決定木とは？

決定木とは予測モデルであり、ある事柄に対する観察の結果から、その事柄の目標値についての結論を導くものである。



*ソース

- > お札データ=read.csv("お札.csv")
- > library(mvpart)
- > お札木=rpart(真偽~,data=お札データ)
- > print(お札木)
- > mvpart(真偽~,data=お札データ)



*実行結果

この実行結果から真札は対角線の長さが140.4より大きく、下部マージンが9.45より小さいと分かる。

