

【データ】

ファイル bank-additional-full.csv は、ポルトガルの銀行による定期預金のセールスデータである。41,188件の顧客について、20種類の属性(説明変数)と、定期預金をしたかどうか(目的変数)が含まれている。10分の1程度にデータを減らした bank-additional.csv もあるので、4万件では多すぎる場合は、まずこちらを使用してもよい。

これらのデータは、R上で次のコマンドにより読み込むことができる。
(CSVと名付けられているが、区切り記号がなぜかカンマではない)

```
read.table("bank-additional-full.csv", header=TRUE, sep=";")
```

データの説明は bank-additional-names.txt (英語)を参照のこと。
注意事項として、11番目の変数(duration)は予測には使用しないことがある。この変数はセールス電話の通話時間であるが、実際の場面では電話をかける前に定期預金の成約の見込みを知りたいのであるから、目的変数の予測に使うのは適切ではない。また、この変数は目的変数と強い相関がある。予測ではなく分析に使用するのには差し支えない。

【課題】

このデータセットの目的は、定期預金するかどうかを予測することではあるが、自由に分析してデータから読み取れることを報告すること。もちろん、何らかの機械学習手法で予測を試みるのが望ましい。

1月11日または18日に、各自10分程度で分析結果を発表する。
PowerPoint等でプレゼンを作成して発表すること。

【出典】

このデータセットは、UCI Machine Learning Repository から取得した。
<http://archive.ics.uci.edu/ml/datasets/Bank+Marketing>

以上