

判別分析

```
library(MASS)
```

- ・目的変数はカテゴリーで二値（L1 vs. L2 とか男女とか、合格不合格とか）
- ・2つ以上の説明変数（二つだとグラフで表現しやすい）

■全体のデータをランダムに半分ずつに分ける `sample()`

- ・トレーニングデータ
- ・テストデータ

■トレーニングデータを使って判別モデルを作成

```
モデル <- ( カテゴリー変数 ~ 変数 A + 変数 B, トレーニングデータ )
```

- ・モデルのあてはまり具合を確認

```
予測結果1 <- predict( モデル , トレーニングデータ )
```

- ・`predict()` 内では、データの指定は「`data = データ`」としてはいけない。

```
table( 予測結果1 $class, トレーニングデータ $ カテゴリー )
```

- ・判別率

```
tbl1 <- table( 予測結果1 $class, トレーニングデータ $ カテゴリー )
```

```
sum(diag(tbl1)) / sum(tbl1)
```

■テストデータで検証

- ・モデルを使って、テストデータを予測して（あてはめて）みる

```
予測結果2 <- predict( モデル , テストデータ )
```

```
table( 予測結果2 $class, テストデータ $ カテゴリー )
```

- ・判別率

```
tbl2 <- table( 予測結果2 $class, テストデータ $ カテゴリー )
```

```
sum(diag(tbl2)) / sum(tbl2)
```

■判別率の確認

```
install.packages("caret", dependencies =T)
library(caret)
```

を使う

- ・このとき、予測結果のオブジェクトの中の class に予測値が入っている

■最初から、Leave One Out という交差検証法を使って、係数の平均を出す方法がある。

```
モデル <- lda( カテゴリー ~ ., data = データ全体 , CV = T)
```

- ・最後の CV = T で、そのオプションを使用（Cross Validation）

■判別直線の描写

- ・こんなデータを例に
- ・定数項の計算

```
teisu <- apply(lda.model$means %*% lda.model$scaling, 2, mean)
```

- ・判別式の直線の求め方（H: height, W: weight）
 - ・ H と W を使った判別式
 - ・ y が判別得点（0 より上か下かで判別する）
 - ・それを 0 として、式の変換
 - ・ y 軸が height なので、H を求めるように
- ・データをプロットして、
- ・そこに、判別直線を重ねる
- ・xc は、x 軸の範囲、つまり、weight の最低値から最大値
- ・yc が、xc に応じて計算される y 軸の値
- ・ references
 - ・ http://daas.la.coocan.jp/GLM/tahenryou_03_discrim.htm
 - ・ <http://entertainment-lab.blogspot.com/2012/12/r.html>

群馬大の青木先生の「判別分析」スクリプト

<http://aoki2.si.gunma-u.ac.jp/R/disc.html>

- ・で分析
- ・でグラフ描写
 - ・ `plot( 判別モデル , which="scatter", xpos="topleft")`

■ステップワイズで

<http://aoki2.si.gunma-u.ac.jp/R/sdis.html>

判別分析の例

判別分析の例 2

References

http://daas.la.coocan.jp/GLM/tahenryou_03_discrim.htm

<http://entertainment-lab.blogspot.com/2012/12/r.html>

<https://mikuhatsume.hatenadiary.com/entry/2021/02/12/165456>

<https://tatsukioike.com/rmult/0026/>

<http://www.kogures.com/hitoshi/javascript/r/discriminant/index.html>