

SVM: Support Vector Machine サポートベクターマシーン

- ・分類には、境界部分のデータポイントに注目
- ・できるだけ距離を置いて分けられる方が良いという発想

サンプリングの方法 (トレーニングデータとテストデータの二種類のデータセットの作成)

SVM 関数

```
ksvm( 分類カテゴリ 説明変数, data= データフレーム, kernel=" 使用するカーネル関数 ", kpar= パラメタ )
```

- ・これでモデルを作って、predict でテストデータを使って試す

```
predict( モデル, テストデータ )
```

- ・もしくは、オプションで交差検証法を使う
 - ・ cross= 交差検証法の回数
- ・カーネル関数
 - ・ vanilladot 線形カーネル
 - ・ polydot 多項式カーネル
 - ・ kpar=list(degree=2) で 2 乗の多項式
 - ・ rbfdot ガウシアンカーネル
 - ・ kpar=list(sigma=1) sigma が大きいと細かく複雑になる
- ・ kernel も kpar も指定しなければ、自動で最適なものを適用してくれる！

```
ksvm( 分類カテゴリ 説明変数, data= データフレーム )
```

だけで OK

分析例

- ・ 誤判別率 3.6 % (判別率あまりよくない)

confusion matrix から誤判別率の出し方

- ・ diag() で対角線の値を取り出すのがポイント
 - ・ つまり、confusion matrix で、対角線のところが正しく判別できたところになる。
 - ・ 合ってた数の合計をテーブル全体を足した数で割れば、正判別率になる。
 - ・ それを 1 から引けば、誤判別率

References

<https://www.cis.doshisha.ac.jp/mjin/R/31/31.html>

<http://yut.hatenablog.com/entry/20120827/1346024147>

<https://shohei-doi.github.io/notes/posts/2019-05-20-classification/#SVM>