

【データサイエンス基礎講座】:2014年12月度

データサイエンス基礎講座(超初級・実践編)
2014年11月26日～12月17日<全5回>演習用資料

R入門

インストールから活用まで

- ・主催:株式会社インプレス
- ・企画/製作:フューチャーブリッジパートナーズ株式会社



第4限目

- カーネル法とサポートベクターマシン
- ロジスティクス回帰と機械学習
- 決定木

カーネル法

- カーネルライブラリのインストールおよびデータの確認
 - `install.packages("kernlab")`
 - `library(kernlab)`
 - `data(iris)`
 - `head(iris)`
 - `x = as.matrix(iris[,1:4])` # 1～4列目まで抽出
- irisデータ
 - 3種類のアヤメについて、50株ずつ採取し、ガクの長さ(Sepal.Length)、幅(Sepal.Width)、花弁の長さ(Petal.Length)、幅(Petal.Width)について測定したデータ
- `head(iris)`

```
> head(iris)
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1           5.1           3.5           1.4           0.2  setosa
2           4.9           3.0           1.4           0.2  setosa
3           4.7           3.2           1.3           0.2  setosa
4           4.6           3.1           1.5           0.2  setosa
5           5.0           3.6           1.4           0.2  setosa
6           5.4           3.9           1.7           0.4  setosa
> |
```

ガクの長さ ガクの幅 花弁の長さ 花弁の幅

カーネル法

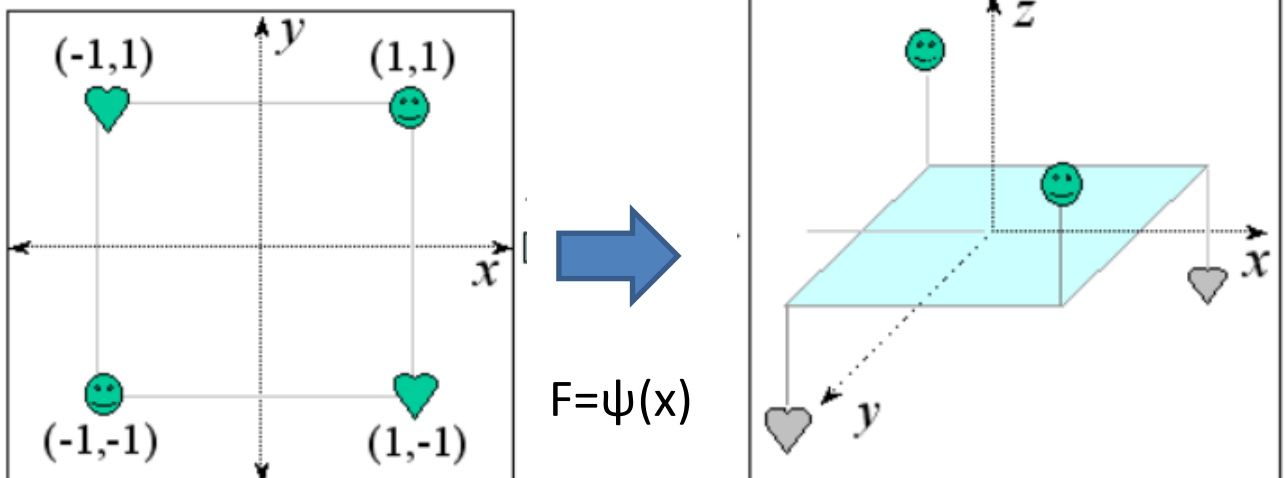
- カーネル法

- データを高次元の特徴空間に写像したときの主成分分析
- 具体的には、2次元平面座標(x,y)に、 $A1(1,1), A2(1,-1), A3(-1,-1), A4(-1,1)$ があるとする
- $A1, A3$ が一つのクラスであるとする、平面上にクラスの境界線を引けない
- 2次元平面(x,y)の4つの点を3次元空間(x,y,z)に射影すると、 $A1(1,1,1), A2(1,-1,-1), A3(-1,-1,1), A4(-1,1,-1)$ になり、両クラスは平面で切り分けることが可能になる。
- 高次元の特徴変換をカーネル法 ($\psi(x)$)と呼ぶ

参考ビデオ

<https://www.youtube.com/watch?v=3liCbRZPrZA&hl=ja&gl=JP>

データ写像



明確に境界を線引きできない

$z=0$ の平面を境界面

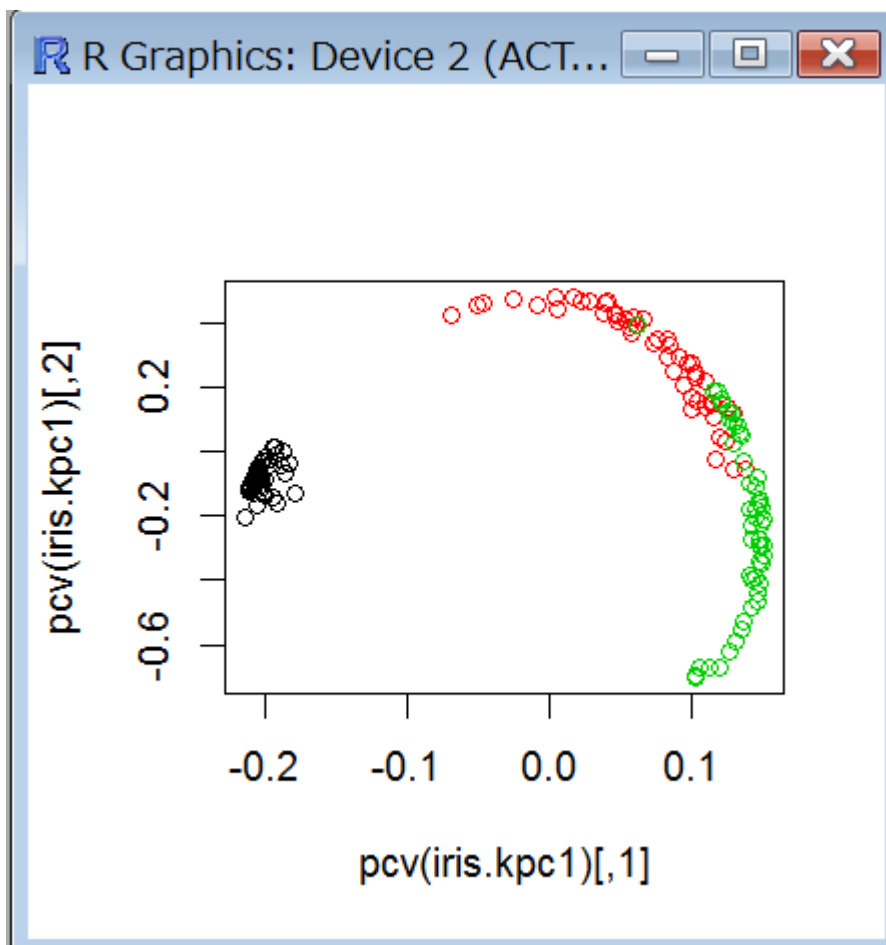
カーネル法

- カーネル主成分分析(KPC)
 - カーネル法による特徴平面に対する主成分分析
 - KPCの導出手順
 - カーネル関数 $K(X_1, X_2) = \langle \psi(X_1), \psi(X_2) \rangle$ を決める
 - データから写像行列を K_{mn} を求める
 - K_{mn} の固有値と固有ベクトルを求める (主成分分析)
- Rによるカーネル主成分分析
 - 関数 `kpca` (データ, `kernel`= カーネル関数, `features` = (求める主成分数) `kpar` = カーネル関数のパラメーターのリスト)
 - カーネル関数: `rbfdot`(ガウジアン、デフォルト)、`polydot`(多項式)
 - `vanilladot`(線形)
 - `tanhdot`(タンジェント)
 - `aplacedot`(ラプラシアン)
 - `besseldot`(ベッセル)
 - `anovadot`(ANOVA RBF)、
 - `splinedot`(スプライン)

カーネル法

- Rによるカーネル主成分分析
 - iris.kpc1 =
kpca(x, kernel="rbfdot", features=2, kpar=list(sigma=0.1))
 - summary(iris.kpc1)
 - plot(pcv(iris.kpc1), col=as.integer(iris[,5]))

```
R Console  
>  
> iris.kpc1<-kpca(x, kernel="rbfdot", features=2, kpar=list(sigma=0.1))
```



決定木

- Min+1SE法の図示
 - `plotcp(result1)`

