

Lightweight Language Spirit
Language Update
R言語

大阪府立大学生命環境科学研究科
樋口 千洋

2007/08/04 於日本教育会館

R言語の概要 1

- SのシンタックスとSchemeのセマンティックスの**一部**を融合
 - S言語 ベル研究所が開発した統計処理言語
 - 強力なグラフィックス機能、動的スコープ
 - Scheme ここで説明する必要はないでしょう ^_^
 - **静的スコープ**、遅延評価
- R 0.49 リリース(1997/4/23)
 - 現在の最新バージョンは2.5.1(2007/6/28)
- マルチプラットフォーム(<http://www.r-project.org/>で配布)
 - Windows95以降、Mac OS X、UNIX(Linux、BSD、etc)
- 対話環境(バッチ処理も可能)
 - オブジェクトの中身を確認しながら操作ができる
- 多数の解析パッケージが公開(CRAN、Bioconductor他)
 - ネットワークで依存関係にある他パッケージをも自動インストール
- オープンソース
 - 個々のパッケージはライセンスが異なる(アカデミックフリーなど)

間違えて.net
と入力すると
大変なことに
なります!!

R言語の概要 2

- R言語の特徴
 - バージョンアップが頻繁
 - バージョンアップの際に整合性が失われることがある(欠点)
 - データはベクトルが基本
 - オブジェクト指向
 - 総称関数で様々なさまざまなデータ構造を一貫的に操作
 - 強力なヘルプシステム
 - さまざまな組み込みデータ
 - 意味のあるデータを適当に用意するのは困難
 - ソースの参照が可能
- R言語が活躍する場面
 - データマイニング、バイオインフォマティクス
 - バイオインフォマティクス論文でRでの実装例が非常に多い

R言語はベクトルデータが基本

```
R Console
ファイル 編集 その他 パッケージ ヘルプ

> (a <- 1:100)
 [1]  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18
[19] 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36
[37] 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54
[55] 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72
[73] 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90
[91] 91 92 93 94 95 96 97 98 99 100
> is.vector(a)
[1] TRUE
> (a <- 1)
[1] 1
> is.vector(a)
[1] TRUE
> is.double(a)
[1] TRUE
> rnorm(4)
[1]  1.014334  1.123060 -0.953888  1.229495
> exp(rnorm(4))
[1] 2.1525906 2.4193886 1.7072813 0.4410082
> exp(matrix(rnorm(4), nrow=2))
      [,1] [,2]
[1,] 0.8210919 1.7972235
[2,] 0.2584302 0.2221933
>
> ■
```

R言語のオブジェクト指向

```
> class(iris)
[1] "data.frame"
> dim(iris)
[1] 150 5
> summary(iris)
```

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
Min.	:4.300	Min. :2.000	Min. :1.000	Min. :0.100	setosa :50
1st Qu.:	5.100	1st Qu.:2.800	1st Qu.:1.600	1st Qu.:0.300	versicolor:50
Median :	5.800	Median :3.000	Median :4.350	Median :1.300	virginica :50
Mean :	5.843	Mean :3.057	Mean :3.758	Mean :1.199	
3rd Qu.:	6.400	3rd Qu.:3.300	3rd Qu.:5.100	3rd Qu.:1.800	
Max. :	7.900	Max. :4.400	Max. :6.900	Max. :2.500	

```
> iris.svm <- svm(Species ~ ., data = iris)
> class(iris.svm)
[1] "svm.formula" "svm"
> summary(iris.svm)
```

Call:
svm(formula = Species ~ ., data = iris)

Parameters:
SVM-Type: C-classification
SVM-Kernel: radial
cost: 1

iris、iris.svmはそれぞれ異なるクラスオブジェクトだが
summaryメソッドはそれぞれに応じた挙動をする

Speciesを目的変数、
それ以外を説明変数
として学習

irisデータセット
アヤメをがくと花卉それぞれの
幅と長さで三種類の品種に分類

サポートベクタマシン(SVM)
教師あり学習を行う識別手法の
一つ、今回は品種が教師データ

R言語でのドキュメント探索

```
R Console
ファイル 編集 その他 パッケージ ヘルプ

> apropos("svm")   ワーキングメモリの名前空間上のオブジェクトを検索(正規表現利用可)
[1] "iris.svm" "best.svm" "svm"      "tune.svm" "write.svm"
> help.search("svm")
> help(plot.svm)
> head(getS3method("plot", "svm"))
```

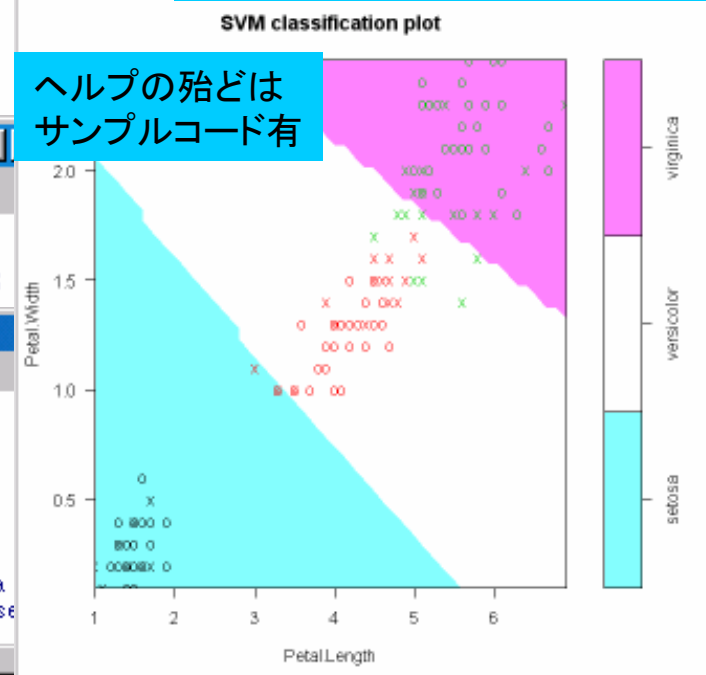
処理系内部に豊富なドキュメントがある

```
1 function (x, data, formula = NULL, fill = TRUE, grid = 50, slice = list(),
2   symbolPalette = palette(), svSymbol = "x", dataSymbol = "o",
3   ...)
4 {
5   if (x$type < 3) {
6     if (is.null(formula) && ncol(data) == 3) {
7       > example(plot.svm)
```

上記以外のキーワード引数に
マッチ、後の関数の引数として使用

example(plot.svm)の結果

ヘルプの殆どは
サンプルコード有



help.search("svm")の結果

```
R Information
ファイル 編集

plot.svm(e1071)
predict.svm(e1071)
svm(e1071)
tune.wrapper(e1071)
write.svm(e1071)
```

Plot SVM Objects
Predict Method for Support Vector Machin\$

```
R Help on 'plot.svm'
ファイル 編集

plot.svm      package:e1071

Plot SVM Objects

Description:

Generates a scatter plot of the input data of a
classification models by highlighting the classe
```

help(plot.svm)の結果

```
plt.sv> ## mor
```

R言語とバイオインフォマティクス

```
R Console
ファイル 編集 その他 パッケージ ヘルプ Vignettes

> data(affybatch.example)
> show(affybatch.example)
AffyBatch out-of-date; consider 'updateObject(<AffyBatch>)'
AffyBatch object
size of arrays=100x100 features (7 kb)
cdf=cdfenv.example (150 affyids)
number of samples=3
number of genes=150
annotation=
notes=
> image(affybatch.example)
ページ変更を確認するまで待機中...
```

DNAマイクロアレイ解析の
場でRは必須

```
> hist(affybatch.example)
> MAplot(affybatch.example)
```

