Universidad de Santiago de Chile
Facultad de Ingeniería
Depto. de Ingeniería Informática

Taller de minería de datos avanzada

Capítulo V

"Máquinas de vectores soporte"

Profesores: Max Chacón y Felipe-Andrés Bello Robles

Ejemplos

- SVM en clasificación
- Estimación con SVM
- Tunning con SVM
  - Kernel Lineal
  - Kernel RBF

Librería

e1071

- Librería de procesamiento y análisis estadístico que entre otras funciones permite trabajar con maquinas de vectores soporte

- require(e1071)
- data(iris)
- attach(iris)
- formula=Species ~ .
- model <- svm(formula, data = iris)

```
> print(model)

Call:
svm(formula = formula, data = iris)


Parameters:
   SVM-Type:  C-classification
 SVM-Kernel:  radial
       cost:  1
      gamma:  0.25

Number of Support Vectors:  51
```

```
> summary(model)

Call:
svm(formula = formula, data = iris)


Parameters:
   SVM-Type:  C-classification
 SVM-Kernel:  radial
       cost:  1
      gamma:  0.25

Number of Support Vectors:  51

 ( 8 22 21 )


Number of Classes:  3

Levels:
 setosa versicolor virginica
```

Librería

- x <- subset(iris, select = -Species)
- y <- Species
- model <- svm(x, y)

test with train data

- pred <- predict(model, x)
- table(pred, y)

```
> table(pred, y)
            y
pred         setosa versicolor virginica
  setosa         50          0         0
  versicolor      0         48         2
  virginica       0          2        48
```

compute decision values and probabilities:

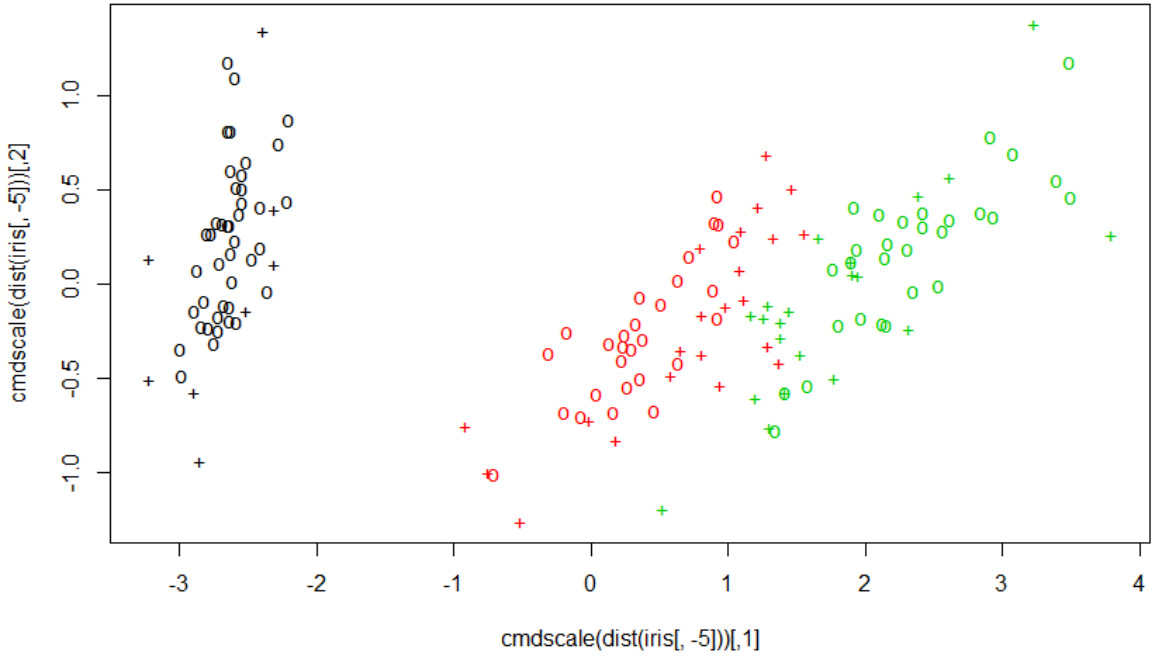
- pred <- predict(model, x, decision.values = TRUE)
- attr(pred, "decision.values")[1:4,]

```
> attr(pred, "decision.values")[1:4,]
  setosa/versicolor setosa/virginica versicolor/virginica
1          1.196152         1.091757            0.6708810
2          1.064621         1.056185            0.8483518
3          1.180842         1.074542            0.6439798
4          1.110699         1.053012            0.6782041
```

Librería

visualize (classes by color, SV by crosses):

- plot(cmdscale(dist(iris[,-5])), col = as.integer(iris[,5]), pch = c("o","+")[1:150 %in% model$index + 1])



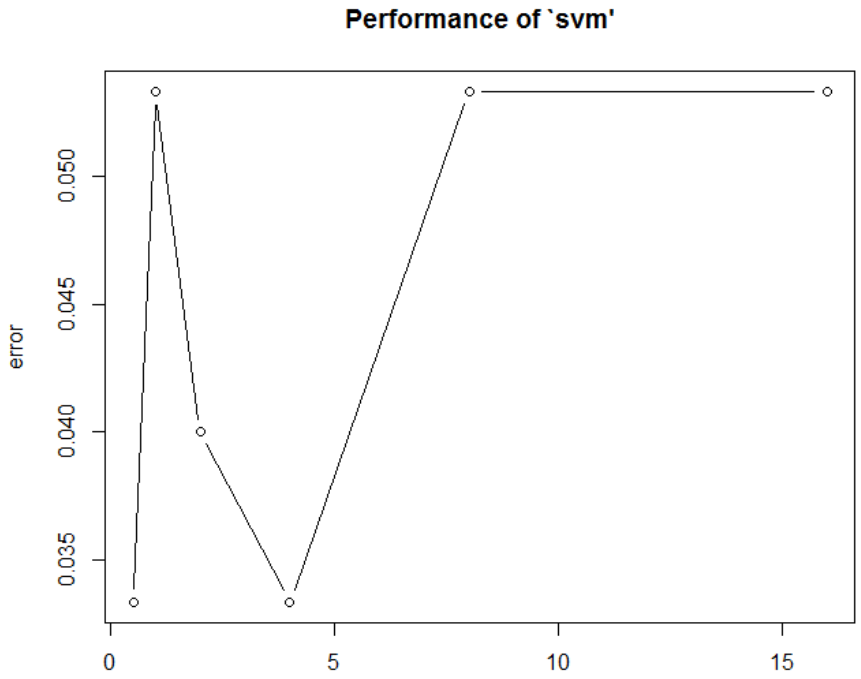
Librería

- obj <- tune(svm, Species~., data = iris, kernel = "linear", ranges = list(cost = 2^(-1:4)), tunecontrol = tune.control(sampling = "cross", cross = 2))
- summary(obj)

```
> summary(obj)

Parameter tuning of 'svm':

- sampling method: 2-fold cross validation

- best parameters:
 cost
  0.5

- best performance: 0.03333333

- Detailed performance results:
  cost      error dispersion
1  0.5 0.03333333 0.00942809
2  1.0 0.05333333 0.00000000
3  2.0 0.04000000 0.00000000
4  4.0 0.03333333 0.00942809
5  8.0 0.05333333 0.01885618
6 16.0 0.05333333 0.01885618
```

Librería

- plot(obj)
- summary(obj$best.model)




```
> summary(obj$best.model)

Call:
best.tune(method = svm, train.x = Species ~ ., data = iris, ranges = list(cost = 2^(-1:4)),
    tunecontrol = tune.control(sampling = "cross", cross = 2), kernel = "linear")


Parameters:
   SVM-Type:  C-classification
 SVM-Kernel:  linear
       cost:  0.5
      gamma:  0.25

Number of Support Vectors:  36

 ( 2 17 17 )


Number of Classes:  3

Levels:
 setosa versicolor virginica
```

Librería

- pred <- predict(obj$best.model, x)
- table(pred, Species)

```
> table(pred, Species)
            Species
pred         setosa versicolor virginica
  setosa         50          0         0
  versicolor      0         49         1
  virginica       0          1        49
```

Librería

- obj <- tune(svm, Species~., data = iris, kernel = "radial", ranges = list(gamma = 2^(-2:4), cost = 2^(-1:4), tunecontrol = tune.control(sampling = "cross", cross = 2)))
- summary(obj)

```
> summary(obj)

Parameter tuning of ‘svm’:

- sampling method: 2-fold cross validation

- best parameters:
 gamma cost
  0.25    2

- best performance: 0.05333333

- Detailed performance results:
   gamma cost      error dispersion
1   0.25  0.5 0.07333333 0.02828427
2   0.50  0.5 0.08000000 0.03771236
```

Librería

- pred <- predict(obj$best.model, x)
- table(pred, Species)

```
> table(pred, Species)
            Species
pred         setosa versicolor virginica
  setosa         50          0         0
  versicolor      0         48         2
  virginica       0          2        48
```

- obj <- tune(svm, Species~., data = iris, kernel = "radial", ranges = list(gamma = 2^(-7:12), cost = 2^(-7:14), tunecontrol = tune.control(sampling = "cross", cross = 2)))

```
> summary(obj)

Parameter tuning of ‘svm’:

- sampling method: 2-fold cross validation

- best parameters:
  gamma cost
 0.0625    4

- best performance: 0.01333333

- Detailed performance results:
         gamma       cost      error dispersion
1   7.8125e-03 7.8125e-03 0.72000000 0.01885618
2   1.5625e-02 7.8125e-03 0.72000000 0.01885618
```

Librería

- pred <- predict(obj$best.model, x)
- table(pred, Species)

```
> table(pred, Species)
            Species
pred         setosa versicolor virginica
  setosa         50          0         0
  versicolor      0         49         1
  virginica       0          1        49
```