



1



2

Regressão

- Dada uma base de dados **ROTULADA**, com valores reais
- Exemplo: Mercado imobiliário (*real state*)
 - <https://archive.ics.uci.edu/ml/datasets/Real+estate+valuation+data+set>
- Dados de Taiwan: Data da transação, Idade, Distância de serviços deslocamento (MRT), Distância de serviços de conveniência, Latitude, Longitude, Preço da Unidade de Área (3,3 m²)
 - 7 atributos
 - 414 instâncias
- Quer-se treinar um modelo que **APRENDE** esses rótulos
 - De forma que se um dado desconhecido seja apresentado, ele dê o valor
 - Ex:
 - Data = 2014.500
 - Idade= 3.2
 - etc
 - **QUAL É O VALOR DA UNIDADE DE ÁREA???**

Prof. Dr. Razer A N R Montaña

SEPT / UFPR

3

3

Regressão

• Summary

```
> summary(df)
      No      X1.transaction.date  X2.house.age  X3.distance.to.the.nearest.MRT.station  X4.number.of.convenience.stores
Min.   : 1.0      Min.   :2013      Min.   : 0.000      Min.   : 23.38      Min.   : 0.000
1st Qu.:104.2    1st Qu.:2013      1st Qu.: 9.025    1st Qu.: 289.32    1st Qu.: 1.000
Median :207.5    Median :2013      Median :16.100    Median : 492.23    Median : 4.000
Mean   :207.5    Mean   :2013      Mean   :17.713    Mean   :1083.89    Mean   : 4.094
3rd Qu.:310.8    3rd Qu.:2013      3rd Qu.:28.150    3rd Qu.:1454.28    3rd Qu.: 6.000
Max.   :414.0    Max.   :2014      Max.   :43.800    Max.   :6488.02    Max.   :10.000

      X5.latitude  X6.longitude  Y.house.price.of.unit.area
Min.   :24.93     Min.   :121.5     Min.   : 7.60
1st Qu.:24.96     1st Qu.:121.5     1st Qu.: 27.70
Median :24.97     Median :121.5     Median : 38.45
Mean   :24.97     Mean   :121.5     Mean   : 37.98
3rd Qu.:24.98     3rd Qu.:121.5     3rd Qu.: 46.60
Max.   :25.01     Max.   :121.6     Max.   :117.50
```

- Objetivo: treinar um modelo de classificação
 - Usar 80% da base para treino do modelo
 - Usar 20% da base para teste de acurácia do modelo treinado

Prof. Dr. Razer A N R Montaña

SEPT / UFPR

4

4

Regressão

- No R
 - Os dados já estão disponíveis, basta usá-los
 - Usaremos o pacote "caret" para treinar modelos
- 1. Instalação e uso dos pacotes necessários

```
> install.packages("e1071")
> install.packages("randomForest")
> install.packages("kernlab")
> install.packages("caret")
> install.packages("RSNNS")
> library("caret")
> library(RSNNS)
```

Prof. Dr. Razer A N R Montão

SEPT/UFPR

5

5

Regressão

2. Carga da base de dados
 - Fazer o download da base
 - Depois carregar o arquivo .csv

```
> df <- read.csv("real-state.csv", sep=";", dec=",")
```

Prof. Dr. Razer A N R Montão

SEPT/UFPR

6

6

Regressão

3. Particionar a bases em treino (80%) e teste (20%)

```
> set.seed(7)

> indices <- createDataPartition(df$Y.house.price.of.unit.area, p=0.80,
list=FALSE)

> treino <- df[indices,]

> teste <- df[-indices,]
```

Prof. Dr. Razer A N R Montão

SEPT/UFRP

7

7

Regressão

4. Treinar um modelo Random Forest com a base de treino e fazer previsões

```
> rf <- train(Y.house.price.of.unit.area~., data=treino, method="rf")
> predicoes.rf <- predict(rf, teste)
```

5. Treinar um modelo SVM com a base de treino e fazer previsões

```
> svm <- train(Y.house.price.of.unit.area~., data=treino, method="svmRadial")
> predicoes.svm <- predict(svm, teste)
```

Prof. Dr. Razer A N R Montão

SEPT/UFRP

8

8

Regressão

6. Treinar um modelo RNA (NNET) com a base de treino e fazer previsões

```
> rna <- train(Y.house.price.of.unit.area~., data=treino, method="nnet")
> predicoes.rna <- predict(rna, teste)
```

7. Treinar um modelo MLP com a base de treino e fazer previsões

```
> model.mlp <- mlp(treino[,1:7], treino[,8], linOut=T)
> predicoes.mlp <- predict(model.mlp, teste[,1:7])
```

Prof. Dr. Razer A N R Montão

SEPT/UFRP

9

9

Regressão

8. Obter métricas de cada modelo

```
> rmse.rf <- RMSE(predicoes.rf, teste$Y.house.price.of.unit.area)
> rmse.svm <- RMSE(predicoes.svm, teste$Y.house.price.of.unit.area)
> rmse.rna <- RMSE(predicoes.rna, teste$Y.house.price.of.unit.area)
> rmse.mlp <- RMSE(predicoes.mlp, teste$Y.house.price.of.unit.area)
```

```
> cat("RMSE RF: ", rmse.rf, "\n")
> cat("RMSE SVM: ", rmse.svm, "\n")
> cat("RMSE RNA: ", rmse.rna, "\n")
> cat("RMSE MLP: ", rmse.mlp, "\n")
```

```
> cat("RMSE RF: ", rmse.rf, "\n")
RMSE RF: 10.59227
> cat("RMSE SVM: ", rmse.svm, "\n")
RMSE SVM: 10.66894
> cat("RMSE RNA: ", rmse.rna, "\n")
RMSE RNA: 40.24188
> cat("RMSE MLP: ", rmse.mlp, "\n")
RMSE MLP: 15.90258
```

Prof. Dr. Razer A N R Montão

SEPT/UFRP

10

10

Classificação.

9. Comparar resultados – Qual o melhor?
 - RMSE : *Root Mean Squared Error*
 - Raiz do erro quadrático médio
 - Quanto menor o RMSE, menos erro foi cometido

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2}$$



Exercícios.

1. Efetuar o exercício de regressão apresentado, usando a base Real State.
 - a) Apresente os resultados dos modelos
 - b) Apresente o modelo que deu o melhor resultado
 - c) Tente obter modelos melhores, procure na internet como configurar os modelos para obter resultados diferentes