

Atividade Modelo de Regressão Não Linear

Danilo A C Souto

Atividade

A empresa possui um sistema cujo uso é muito custoso pois é pago por utilização e o contrato em vigência está vencendo. Pretende-se substituir o serviço por uma tecnologia mais barata.

Mas existem clientes com contratos ativos e que irão vencer em até 3 anos. Portanto a solução terá que ser mantida por no mínimo esse período até ser substituída completamente de maneira que estes clientes possam migrar de uma maneira tranquila.

Para manter a o uso do sistema atual, sem ter blackout do serviço, tiveram as seguintes propostas:

1. plano A + 35 centavos por uso
2. plano B + Assinatura com custo de RS 10.000,00 mensais + 30 centavos por uso
3. plano C + Contrato com duração de 3 anos por RS 1.000.000,00 + 17 centavos por uso

Para isso o seu diretor quer saber:

Qual plano deve ser contratado para os próximos 3 Anos para que a empresa tenha o menor gasto?

Quanto será gasto com cada um dos planos?

Os dados para a resolução do exercício estão no arquiv `servico-atividade-final-2.csv` nele há o uso diário do serviço. Faça a soma acumulada do serviço para poder estimar.

O modelo a ser estimado segue um padrão sigmoidal. Onde a utilização no início tem um crescimento rápido mas, como o serviço irá ser substituído, haverá um decréscimo até que no fim do período não haja mais utilização escolha uma função de crescimento sugestão:

$$f(x) = \frac{\beta_0}{1 + e^{-\beta_1 \cdot (x - \beta_2)}}$$

onde β_0 é o valor máximo da curva, β_1 é o valor da inclinação da curva, e β_2 é o ponto médio da curva no valor de x, são os parâmetros a serem obtidos

Importante! Para que haja convergência entre os valores, é necessário que os valores estejam na mesma escala, portanto vamos dividir o valor acumulado por 1.000

Resolução

Carregando os dados do arquivo

```
dados <- read.csv2(file="../servico-atividade-final-2.csv")
```

Analisando os dados

```
head(dados)
```

```
##           data uso  ano mes
## 1 2012-11-13   1 2012  11
## 2 2012-11-20   1 2012  11
## 3 2012-11-21   6 2012  11
## 4 2012-11-22  12 2012  11
## 5 2012-11-23   3 2012  11
## 6 2012-11-26   5 2012  11
```

```
tail(dados)
```

```
##           data  uso  ano mes
## 1944 2018-06-07 18905 2018   6
## 1945 2018-06-08 15419 2018   6
## 1946 2018-06-09  9801 2018   6
## 1947 2018-06-10  8707 2018   6
## 1948 2018-06-11 32241 2018   6
## 1949 2018-06-12 35187 2018   6
```

Ajustando a data inicial

```
data <- as.Date(dados$data, origin = "1970-01-01")
dados$data <- data
dados$dia <- as.numeric(dados$data)
dados$dia <- dados$dia - min(dados$dia)
head(dados)
```

```
##           data uso  ano mes dia
## 1 2012-11-13   1 2012  11   0
## 2 2012-11-20   1 2012  11   7
## 3 2012-11-21   6 2012  11   8
## 4 2012-11-22  12 2012  11   9
## 5 2012-11-23   3 2012  11  10
## 6 2012-11-26   5 2012  11  13
```

Gerando uma coluna com o somatório acumulado da utilização

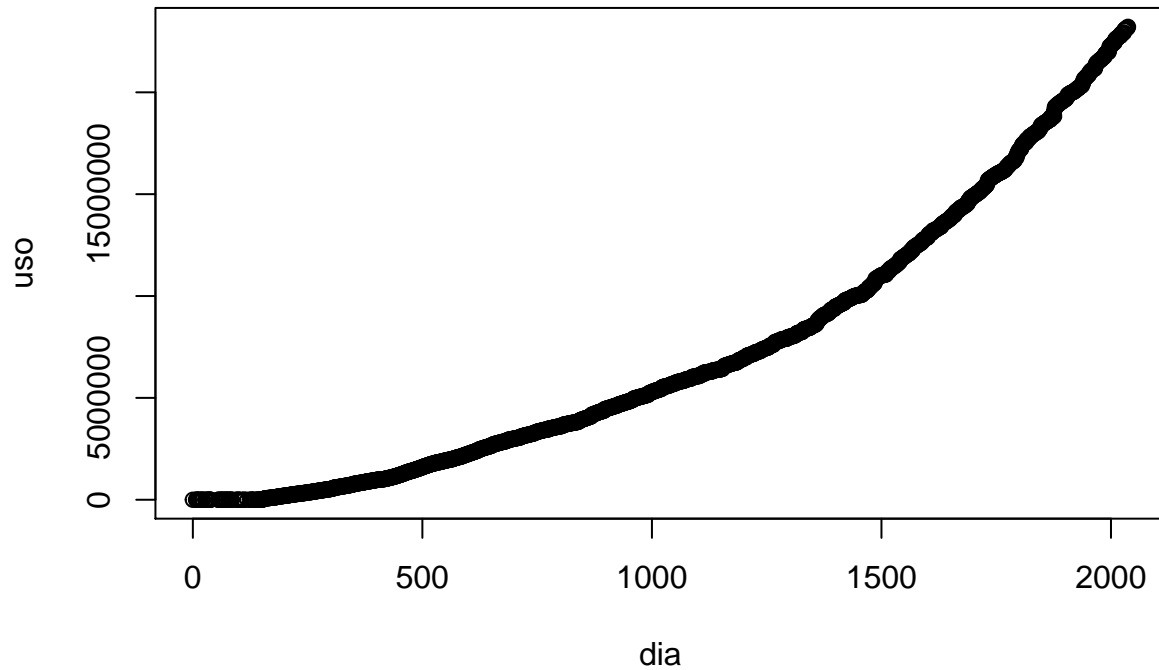
```
acumulado <- cumsum(dados$uso)
dados <- cbind(dados,acumulado)
head(dados)
```

```
##           data uso  ano mes dia acumulado
## 1 2012-11-13   1 2012  11   0           1
## 2 2012-11-20   1 2012  11   7           2
## 3 2012-11-21   6 2012  11   8           8
## 4 2012-11-22  12 2012  11   9          20
## 5 2012-11-23   3 2012  11  10          23
## 6 2012-11-26   5 2012  11  13          28
```

Ajustando para exibir no gráfico

```
## -- Attaching packages ----- tidyverse 1
## v ggplot2 3.3.2      v purrr 0.3.4
## v tibble 3.0.1       v dplyr 1.0.0
## v tidyr 1.1.0        v stringr 1.4.0
## v readr 1.3.1        v forcats 0.5.0
```

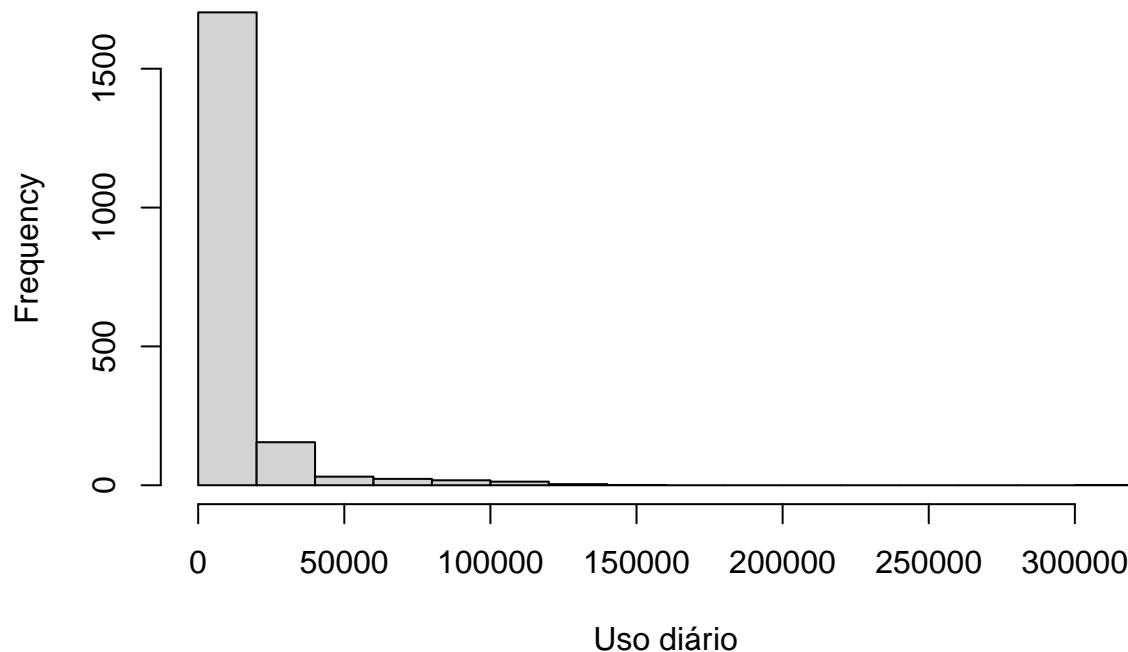
```
## -- Conflicts ----- tidyverse_conflic
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()      masks stats::lag()
plot(dados$acumulado ~ dados$dia, xlab="dia", ylab="uso")
```



Pela assimetria do histograma do uso podemos ver que há um grande desequilíbrio na utilização diária do serviço. Muitos dias com pouco uso e alguns com uma utilização bastante alta

```
hist(dados$uso, main="Histograma", xlab="Uso diário")
```

Histograma



Modelos não lineares pois o programa tem um fim previsto. Ou seja, vai ter um momento que não crescerá mais podendo-se escolher pela sigmóide.

escolha uma função de crescimento sugestão:

$$f(x) = \frac{\beta_0}{1 + e^{-\beta_1 \cdot (x - \beta_2)}}$$

onde β_0 é o valor máximo da curva, β_1 é o valor da inclinação da curva, e β_2 é o ponto médio da curva no valor de x, são os parâmetros a serem obtidos

Para que haja convergência entre os valores, é necessário que os valores estejam na mesma escala, portanto vamos dividir o valor acumulado por 1.000

```
dados$ac_reduzido <- dados$acumulado/1000
```

Ajustando o dia inicial

```
#x1 <- min(dados$dia)
#xf <- max(dados$dia)+1500
#dia2 <- x1:xf
```

Função escolhida

```
curva1 <- NULL
curva1$formula <- y ~ (b0) / ( 1 + exp(-b1 * ( x - b2 )))
```

Definindo os valores iniciais para otimização

```
b0 <- max(dados$ac_reduzido) * 10
b1 <- 0.001
b2 <- max(dados$dia)
```

```
curva1$parametros <- list(b0=b0, b1=b1, b2=b2)
```

Realizando a regressão com a função `nls()`

```
modelo <- nls( formula = y ~ b0 / ( 1 + exp(-b1 * ( x - b2 ))) ,
              start = list(b0=b0, b1=b1, b2=b2) ,
              data = data.frame(y = dados$ac_reduzido, x = dados$dia) )
summary(modelo)
```

```
##
## Formula: y ~ b0/(1 + exp(-b1 * (x - b2)))
##
## Parameters:
##      Estimate      Std. Error t value      Pr(>|t|)
## b0 49314.8558708    814.3579962    60.6 <0.0000000000000002 ***
## b1   0.0020223      0.0000112   180.8 <0.0000000000000002 ***
## b2  2106.9709231    14.3768289   146.6 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 399 on 1946 degrees of freedom
##
## Number of iterations to convergence: 10
## Achieved convergence tolerance: 0.00000131
```

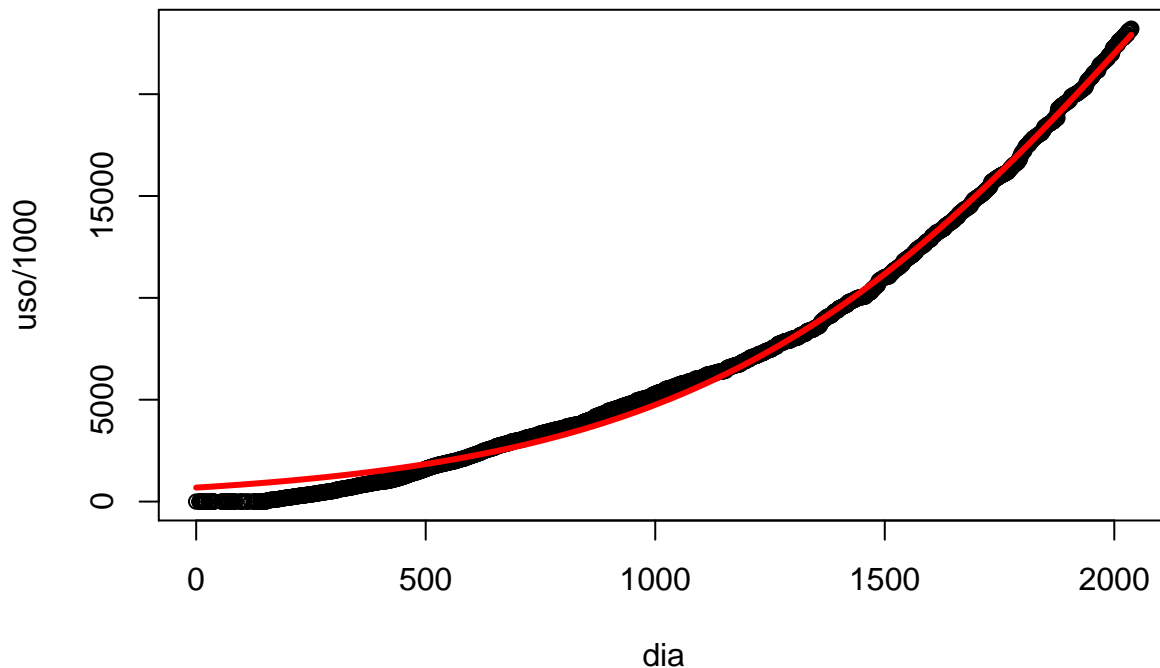
Comparando o critério de Akaike com outras regressões vemos que tem um valor menor

```
cbind(AIC_nls=AIC(modelo),
      AIC_lm=AIC(glm(uso ~ dia, data = dados )),
      AIC_glm=AIC(glm(uso ~ dia, data = dados , family = Gamma()) ))
```

```
##      AIC_nls  AIC_lm AIC_glm
## [1,] 28876.3 43352.7 39707.7
```

Vamos ver graficamente se o ajuste se adequa bem aos dados

```
plot(dados$ac_reduzido ~ dados$dia, xlab="dia", ylab="uso/1000")
points(dados$dia, predict(modelo), type = "l" , col = 'red', lwd=3)
```



#Plote um grafico de uso ~ dia e veja se o modelo está parecido com o observado

Agora o que é necessário fazer é prever o consumo para os proximos anos. Ajuste o gráfico quanto os dados que estão em dias

considere que hoje é o último dia exibido nos dados

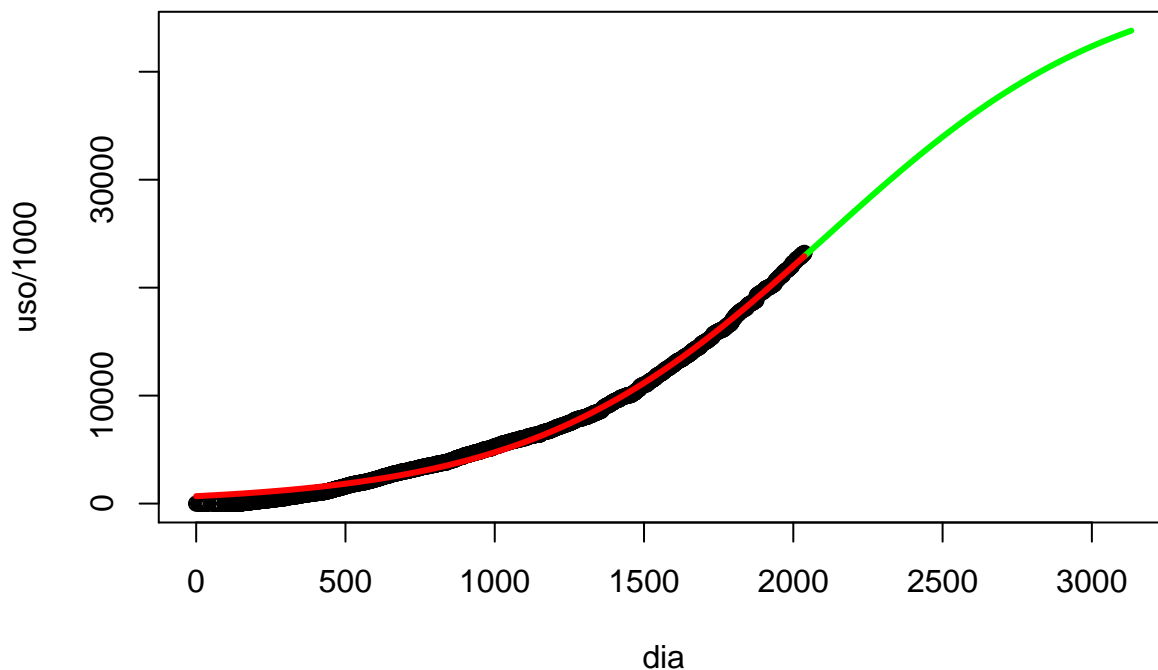
```
hoje <- max(dados$dia)
consumido <- max(dados$acumulado)
```

Agora resta fazer a previsão para os próximos 3 anos (lembre-se que estamos trabalhando em dias)

```
dia3anos <- hoje + 3*365
previsao <- data.frame(dia = hoje:dia3anos )
previsao$consumo <- predict(modelo, newdata = data.frame(x = previsao$dia ))
tail(previsao)
```

```
##      dia consumo
## 1091 3127 43754.0
## 1092 3128 43763.9
## 1093 3129 43773.9
## 1094 3130 43783.8
## 1095 3131 43793.8
## 1096 3132 43803.7
```

```
plot(dados$ac_reduzido ~ dados$dia,
     xlab="dia",
     ylab="uso/1000",
     xlim = c(min(dados$dia),dia3anos) ,
     ylim = c(0, max(previsao)))
lines(x = previsao$dia, y = previsao$consumo , lwd=3, col = "green")
points(dados$dia, predict(modelo), type = "l", lwd=3 , col = 'red')
```



ex-

ibindo os valores encontrados para b0, b1, b2

```
coefficients(modelo)
```

```
##          b0          b1          b2
## 49314.85587079  0.00202232 2106.97092315
```

Agora determinando a função de custo para cada plano pode-se estimar o custo final de cada um ao final do contrato e escolher o mais economico para a empresa

```
#plano A
# 35 centavos por uso
planoA <- function(x){
  x * 0.35
}

#planoB
# Assinatura
# 10.000 mes
# 30 centavos por uso
planoB <- function(x, meses){
  x * 0.3 + 10000 * meses
}

#planoC
# Contrato 3 anos
# 1.000.000 por contrato
# 17 centavos por uso
planoC <- function(x){
  x * 0.175 + 1000000
}
```

Para estimar o custo final basta utilizar a estimativa final de consumo e subtrair do que foi consumido até hoje

```
consumo3anos <- max(previsao$consumo) * 1000 - consumido
consumo3anos
```

```
## [1] 20588886
```

Custos de cada plano

```
custos <- cbind(
  CustoA = planoA(consumo3anos),
  CustoB = planoB(consumo3anos, 36),
  CustoC = planoC(consumo3anos))
```

```
custos
```

```
##          CustoA  CustoB  CustoC
## [1,] 7206110 6536666 4603055
```

Como pode ser visto, ao final dos 3 anos teremos consumidos 47.812.000 utilizações com o intervalo de confiança 95% de

```
confint(modelo)
```

```
## Waiting for profiling to be done...
```

```
##          2.5%          97.5%
## b0 47812.97834056 50941.94701958
## b1    0.00200138    0.00204336
## b2  2080.13633069  2135.09588585
```

e, apesar do plano C ter uma penalização inicial muito alta, 1 milhão já na assinatura do contrato, ele apresenta o menor custo final 4.603.055,00 contra mais de 7 milhões e 6,5 milhões dos planos B e C, respectivamente.

Se considerarmos o valor da assintótica β_0 e não um corte em 3 anos, ainda assim o plano C continua vantajoso com seu intervalo de confiança de 95%

```
intervalo <- confint(modelo)[1,]*1000
```

```
## Waiting for profiling to be done...
```

```
CustosAssintotica <- cbind(
  CustoA = planoA(intervalo - consumido),
  CustoB = planoB(intervalo - consumido, 36),
  CustoC = planoC(intervalo - consumido))
```

```
CustosAssintotica
```

```
##          CustoA  CustoB  CustoC
## 2.5%  8609372 7739461 5304686
## 97.5% 9704511 8678152 5852255
```