Exercício 3: Aproximação Polinomial

Rúbia Reis Guerra
2013031143

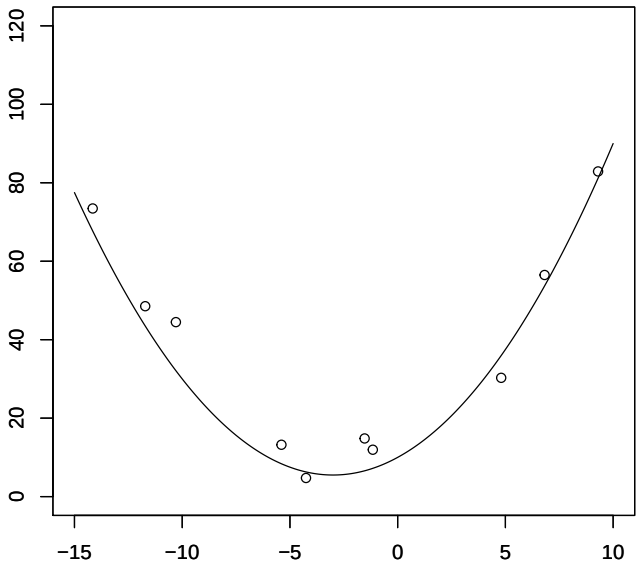
25 de abril de 2017

1 Introdução

Nesta atividade, foi implementada a aproximação polinomial de uma função geradora $f_g(x)$. Em seguida, a partir da modificação do grau do polinômio $p(x)$, foram observados os fenômenos de *underfitting* e *overfitting* nos modelos resultantes.

2 N = 10

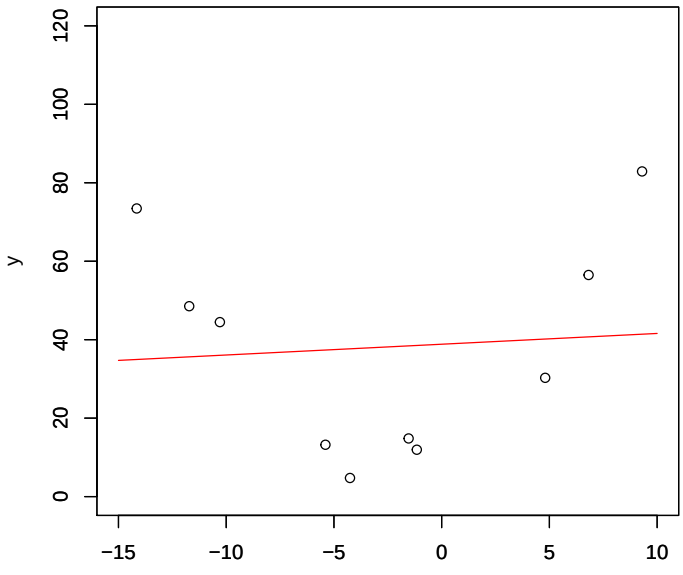
```
> rm(list=ls())
> library('corpcor')
> ######################
> N <- 10
> cores <- rainbow(8)
> X <- runif(n = N, min=-15, max=10)
> Y <- (0.5*(X^2)+3*X+10) + rnorm(length(X), 0, 4)
> xgrid <- seq(-15, 10, 0.1)
> ygrid <- (0.5*(xgrid^2)+3*xgrid+10)
> plot(X, Y, type='p', xlim=c(-15,10), ylim=c(0,120), xlab='', ylab='')
> par(new=T)
> plot(xgrid, ygrid, type='l', xlim=c(-15,10), ylim=c(0,120),
+      col='black', xlab='', ylab='', sub = 'Função geradora, 10 amostras')
```



Função geradora, 10 amostras

2.1 $\mathbf{p = 1}$

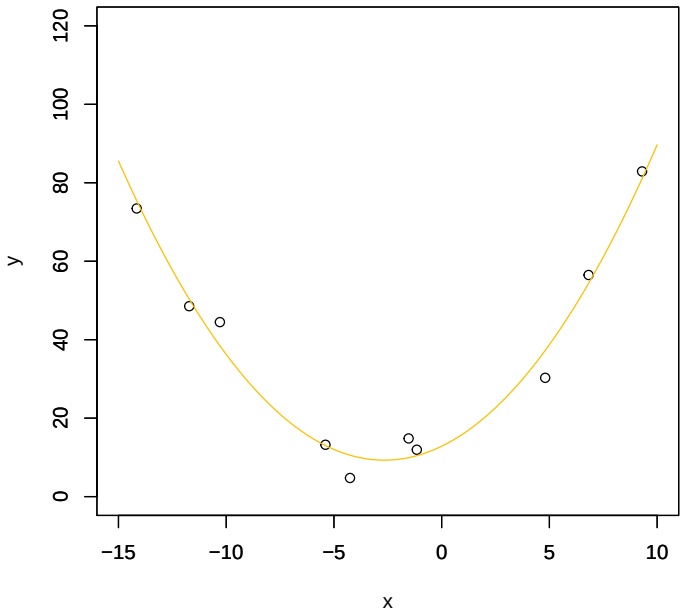
```
> H <- cbind(X, 1)
> w <- pseudoinverse(H) %*% Y
> Hgrid <- cbind(xgrid, 1)
> yhat <- H %*% w
> yhatgrid <- Hgrid %*% w
> plot(X, Y, type='p', xlim=c(-15,10), ylim=c(0,120), xlab='', ylab='')
> par(new=T)
> plot(xgrid, yhatgrid, type='l', xlim=c(-15,10), ylim=c(0,120),
+      col=cores[1], xlab='x', ylab='y', sub='10 amostras, p = 1, underfitting')
```



10 amostras, p = 1, underfitting

2.2 p = 2

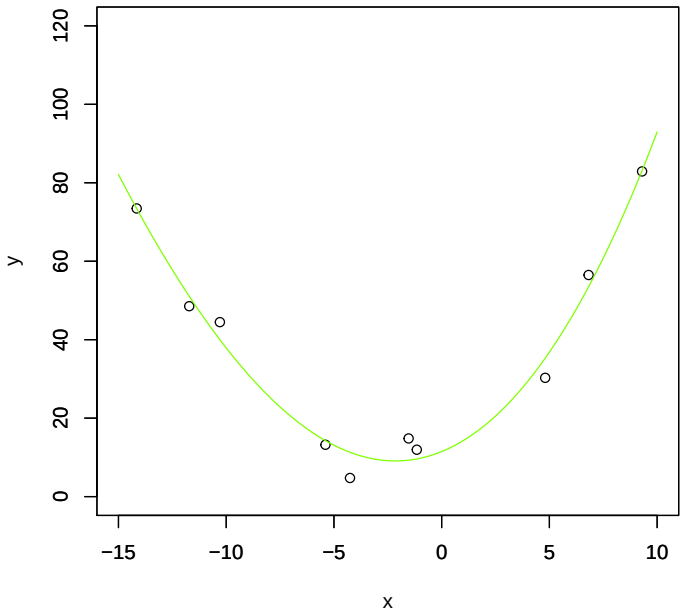
```
> H <- cbind(X^2, X, 1)
> w <- pseudoinverse(H) %*% Y
> Hgrid <- cbind(xgrid^2, xgrid, 1)
> yhat <- H %*% w
> yhatgrid <- Hgrid %*% w
> plot(X, Y, type='p', xlim=c(-15,10), ylim=c(0,120), xlab='', ylab='')
> par(new=T)
> plot(xgrid, yhatgrid, type='l', xlim=c(-15,10), ylim=c(0,120),
+      col=cores[2], xlab='x', ylab='y', sub='10 amostras, p = 2, modelo balanceado')
```



10 amostras, p = 2, modelo balanceado

2.3 p = 3

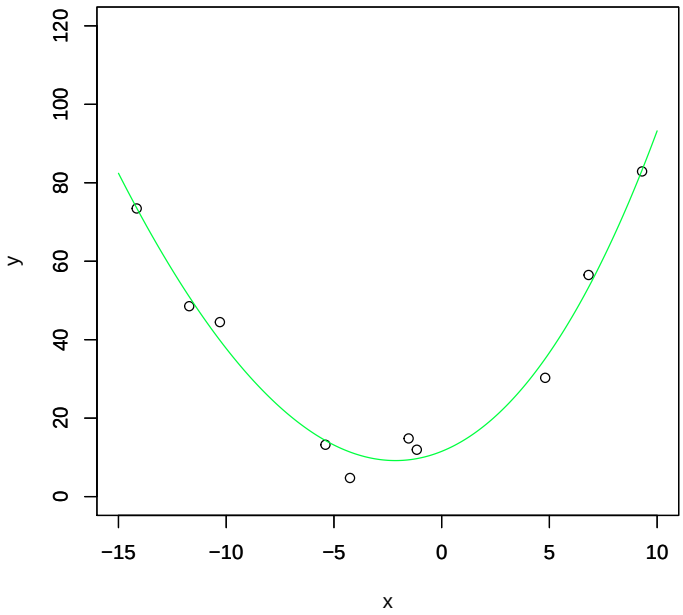
```
> H <- cbind(X^3, X^2, X, 1)
> w <- pseudoinverse(H) %*% Y
> Hgrid <- cbind(xgrid^3, xgrid^2, xgrid, 1)
> yhat <- H %*% w
> yhatgrid <- Hgrid %*% w
> plot(X, Y, type='p', xlim=c(-15,10), ylim=c(0,120), xlab='', ylab='')
> par(new=T)
> plot(xgrid, yhatgrid, type='l', xlim=c(-15,10), ylim=c(0,120),
+      col=cores[3], xlab='x', ylab='y', sub='10 amostras, p = 3, modelo balanceado')
```



10 amostras, p = 3, modelo balanceado

2.4 p = 4

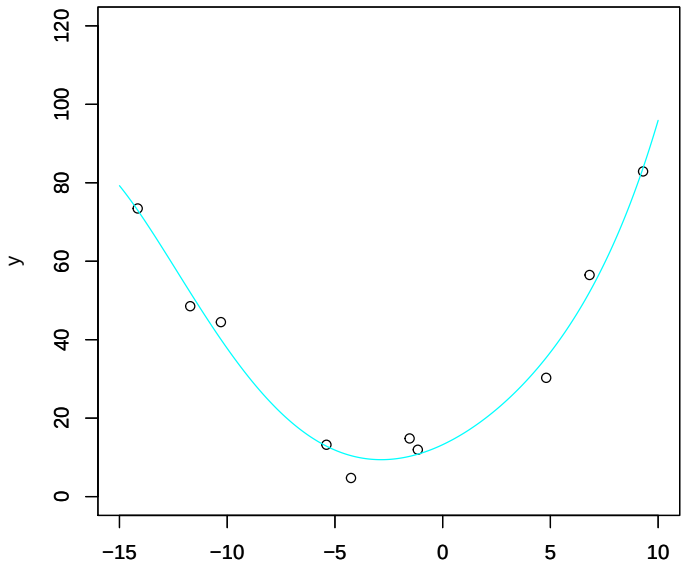
```
> H <- cbind(X^4, X^3, X^2, X, 1)
> w <- pseudoinverse(H) %*% Y
> Hgrid <- cbind(xgrid^4, xgrid^3, xgrid^2, xgrid, 1)
> yhat <- H %*% w
> yhatgrid <- Hgrid %*% w
> plot(X, Y, type='p', xlim=c(-15,10), ylim=c(0,120), xlab='', ylab='')
> par(new=T)
> plot(xgrid, yhatgrid, type='l', xlim=c(-15,10), ylim=c(0,120),
+      col=cores[4], xlab='x', ylab='y', sub='10 amostras, p = 4, modelo balanceado')
```



10 amostras, p = 4, modelo balanceado

2.5 p = 5

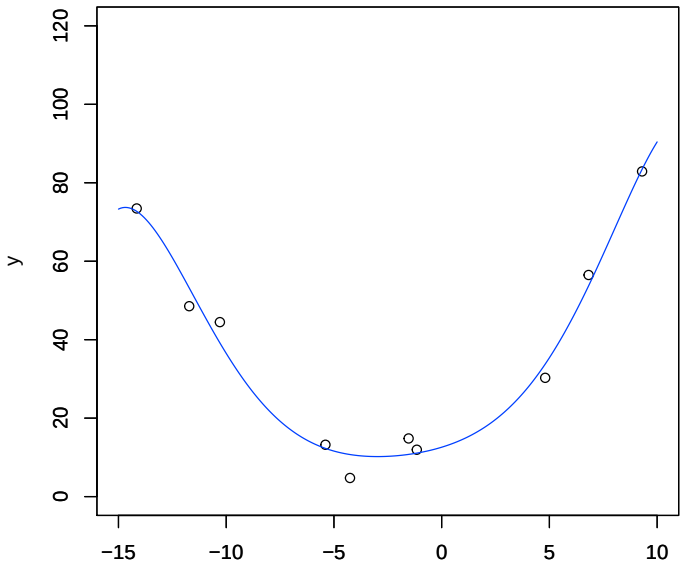
```
> H <- cbind(X^5, X^4, X^3, X^2, X, 1)
> w <- pseudoinverse(H) %*% Y
> Hgrid <- cbind(xgrid^5, xgrid^4, xgrid^3, xgrid^2, xgrid, 1)
> yhat <- H %*% w
> yhatgrid <- Hgrid %*% w
> plot(X, Y, type='p', xlim=c(-15,10), ylim=c(0,120), xlab='', ylab='')
> par(new=T)
> plot(xgrid, yhatgrid, type='l', xlim=c(-15,10), ylim=c(0,120),
+      col=cores[5], xlab='x', ylab='y', sub='10 amostras, p = 5, overfitting')
```



10 amostras, p = 5, overfitting

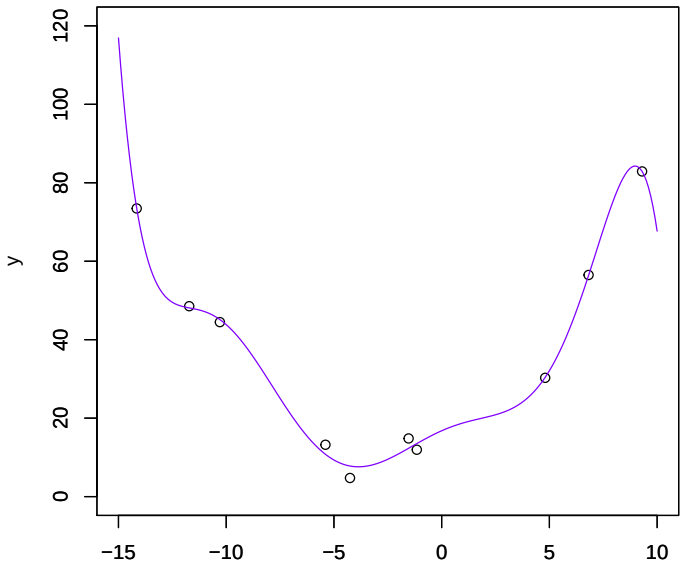
2.6 p = 6

```
> H <- cbind(X^6, X^5, X^4, X^3, X^2, X, 1)
> w <- pseudoinverse(H) %*% Y
> Hgrid <- cbind(xgrid^6, xgrid^5, xgrid^4, xgrid^3, xgrid^2, xgrid, 1)
> yhat <- H %*% w
> yhatgrid <- Hgrid %*% w
> plot(X, Y, type='p', xlim=c(-15,10), ylim=c(0,120), xlab='', ylab='')
> par(new=T)
> plot(xgrid, yhatgrid, type='l', xlim=c(-15,10), ylim=c(0,120),
+      col=cores[6], xlab='x', ylab='y', sub='10 amostras, p = 6, overfitting')
```



2.7 p = 7

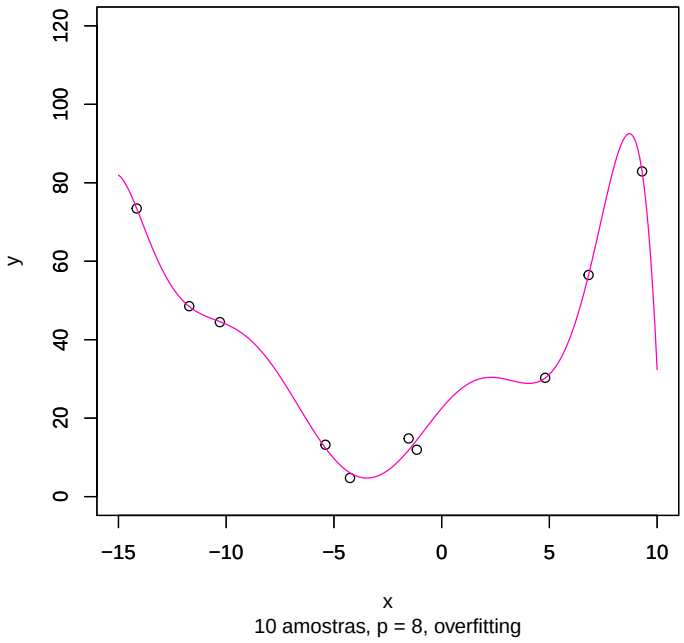
```
> H <- cbind(X^7, X^6, X^5, X^4, X^3, X^2, X, 1)
> w <- pseudoinverse(H) %*% Y
> Hgrid <- cbind(xgrid^7, xgrid^6, xgrid^5, xgrid^4, xgrid^3, xgrid^2, xgrid, 1)
> yhat <- H %*% w
> yhatgrid <- Hgrid %*% w
> plot(X, Y, type='p', xlim=c(-15,10), ylim=c(0,120), xlab='', ylab='')
> par(new=T)
> plot(xgrid, yhatgrid, type='l', xlim=c(-15,10), ylim=c(0,120),
+      col=cores[7], xlab='x', ylab='y', sub='10 amostras, p = 7, overfitting')
```



10 amostras, p = 7, overfitting

2.8 p = 8

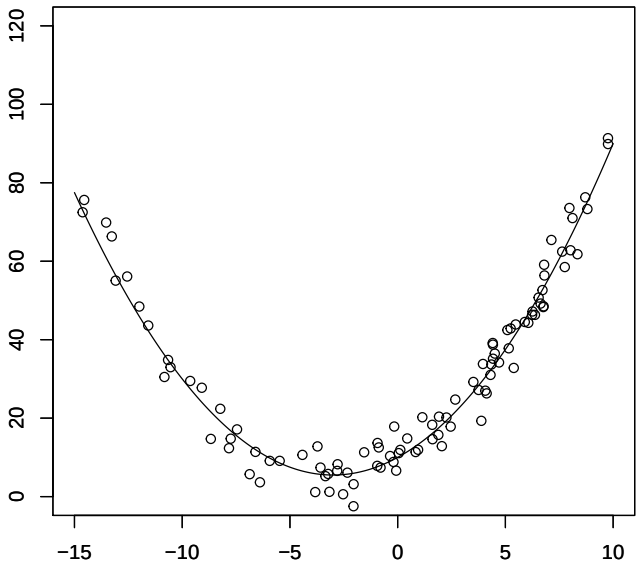
```
> H <- cbind(X^8, X^7, X^6, X^5, X^4, X^3, X^2, X, 1)
> w <- pseudoinverse(H) %*% Y
> Hgrid <- cbind(xgrid^8, xgrid^7, xgrid^6, xgrid^5, xgrid^4, xgrid^3, xgrid^2, xgrid, 1)
> yhat <- H %*% w
> yhatgrid <- Hgrid %*% w
> plot(X, Y, type='p', xlim=c(-15,10), ylim=c(0,120), xlab='', ylab='')
> par(new=T)
> plot(xgrid, yhatgrid, type='l', xlim=c(-15,10), ylim=c(0,120),
+      col=cores[8], xlab='x', ylab='y', sub='10 amostras, p = 8, overfitting')
```



3 N = 100

Com um maior número de amostras, o modelo é menos sensível a flutuações nos dados (menor variância) e, como consequência, é menos propenso a overfitting.

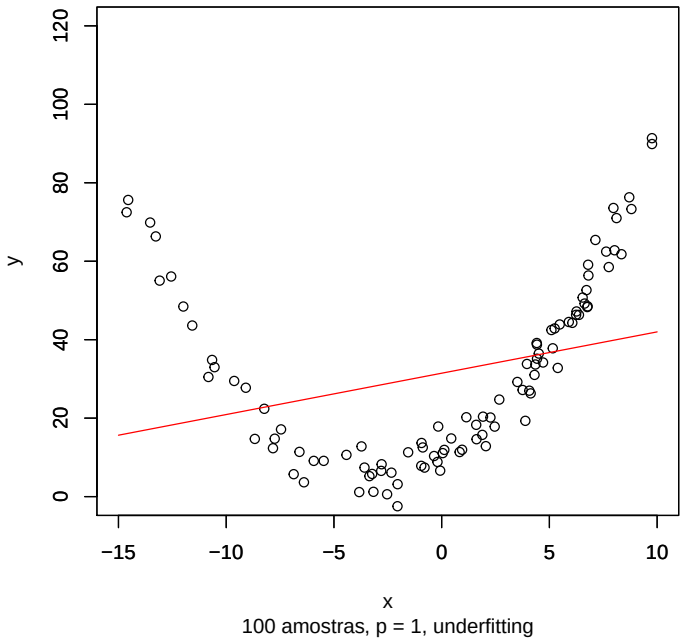
```
> rm(list=ls())
> library('corpcor')
> ######################
> N <- 100
> cores <- rainbow(8)
> X <- runif(n = N, min=-15, max=10)
> Y <- (0.5*(X^2)+3*X+10) + rnorm(length(X), 0, 4)
> xgrid <- seq(-15, 10, 0.1)
> ygrid <- (0.5*(xgrid^2)+3*xgrid+10)
> plot(X, Y, type='p', xlim=c(-15,10), ylim=c(0,120), xlab='', ylab='')
> par(new=T)
> plot(xgrid, ygrid, type='l', xlim=c(-15,10), ylim=c(0,120),
+      col='black', xlab='', ylab='', sub = 'Função geradora, 100 amostras')
```



Função geradora, 100 amostras

3.1 p = 1

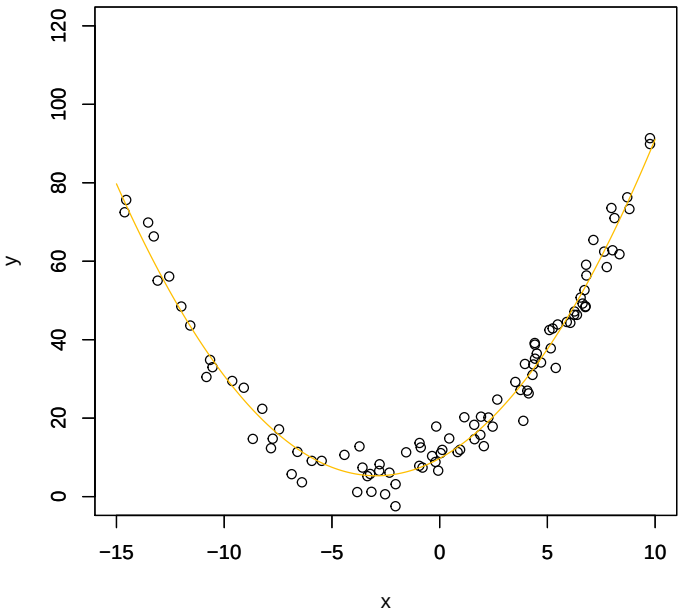
```
> H <- cbind(X, 1)
> w <- pseudoinverse(H) %*% Y
> Hgrid <- cbind(xgrid, 1)
> yhat <- H %*% w
> yhatgrid <- Hgrid %*% w
> plot(X, Y, type='p', xlim=c(-15,10), ylim=c(0,120), xlab='', ylab='')
> par(new=T)
> plot(xgrid, yhatgrid, type='l', xlim=c(-15,10), ylim=c(0,120),
+      col=cores[1], xlab='x', ylab='y', sub='100 amostras, p = 1, underfitting')
```



100 amostras, p = 1, underfitting

3.2 p = 2

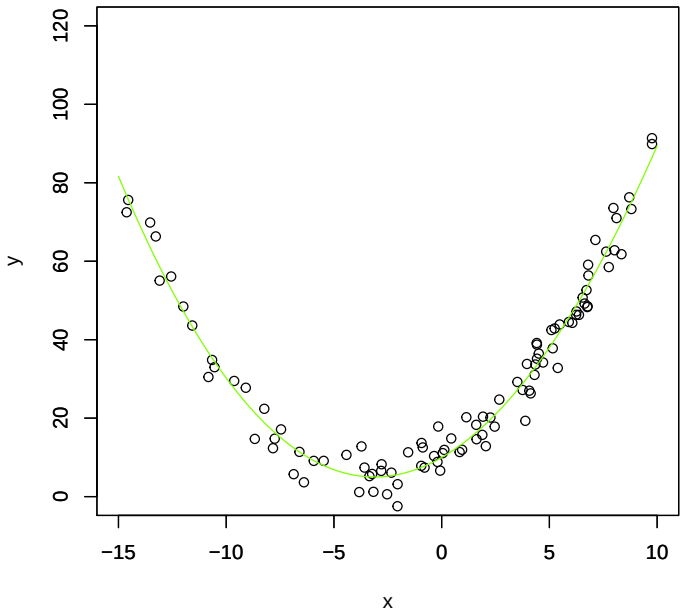
```
> H <- cbind(X^2, X, 1)
> w <- pseudoinverse(H) %*% Y
> Hgrid <- cbind(xgrid^2, xgrid, 1)
> yhat <- H %*% w
> yhatgrid <- Hgrid %*% w
> plot(X, Y, type='p', xlim=c(-15,10), ylim=c(0,120), xlab='', ylab='')
> par(new=T)
> plot(xgrid, yhatgrid, type='l', xlim=c(-15,10), ylim=c(0,120),
+      col=cores[2], xlab='x', ylab='y', sub='100 amostras, p = 2, modelo balanceado')
```



100 amostras, p = 2, modelo balanceado

3.3 p = 3

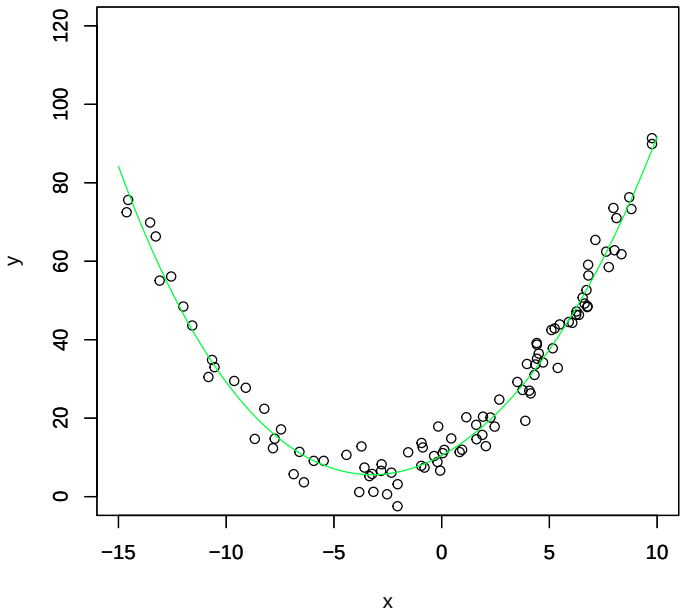
```
> H <- cbind(X^3, X^2, X, 1)
> w <- pseudoinverse(H) %*% Y
> Hgrid <- cbind(xgrid^3, xgrid^2, xgrid, 1)
> yhat <- H %*% w
> yhatgrid <- Hgrid %*% w
> plot(X, Y, type='p', xlim=c(-15,10), ylim=c(0,120), xlab='', ylab='')
> par(new=T)
> plot(xgrid, yhatgrid, type='l', xlim=c(-15,10), ylim=c(0,120),
+      col=cores[3], xlab='x', ylab='y', sub='100 amostras, p = 3, modelo balanceado')
```



100 amostras, p = 3, modelo balanceado

3.4 p = 4

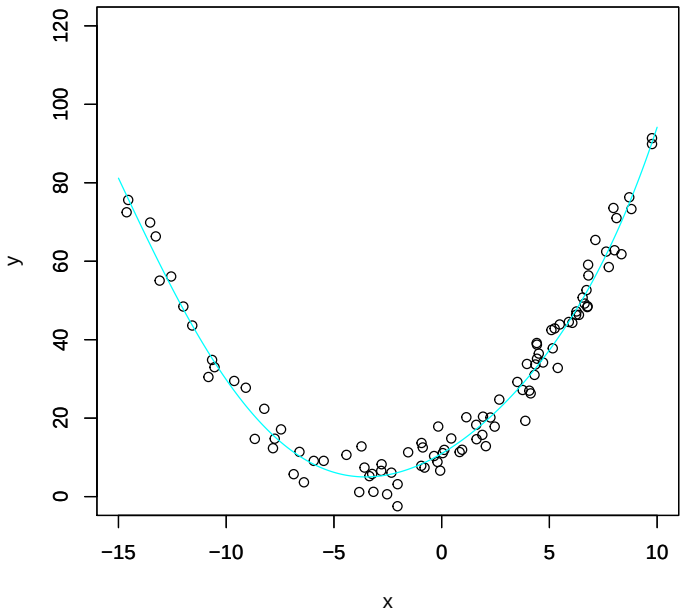
```
> H <- cbind(X^4, X^3, X^2, X, 1)
> w <- pseudoinverse(H) %*% Y
> Hgrid <- cbind(xgrid^4, xgrid^3, xgrid^2, xgrid, 1)
> yhat <- H %*% w
> yhatgrid <- Hgrid %*% w
> plot(X, Y, type='p', xlim=c(-15,10), ylim=c(0,120), xlab='', ylab='')
> par(new=T)
> plot(xgrid, yhatgrid, type='l', xlim=c(-15,10), ylim=c(0,120),
+      col=cores[4], xlab='x', ylab='y', sub='100 amostras, p = 4, modelo balanceado')
```



100 amostras, p = 4, modelo balanceado

3.5 p = 5

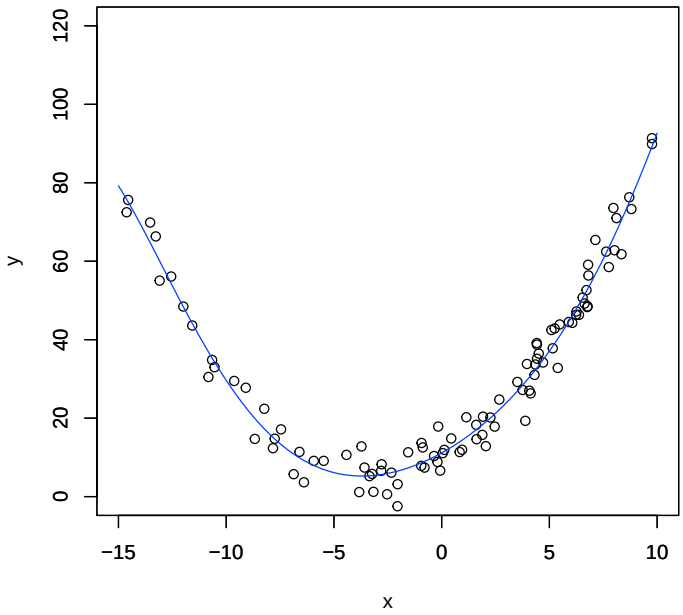
```
> H <- cbind(X^5, X^4, X^3, X^2, X, 1)
> w <- pseudoinverse(H) %*% Y
> Hgrid <- cbind(xgrid^5, xgrid^4, xgrid^3, xgrid^2, xgrid, 1)
> yhat <- H %*% w
> yhatgrid <- Hgrid %*% w
> plot(X, Y, type='p', xlim=c(-15,10), ylim=c(0,120), xlab='', ylab='')
> par(new=T)
> plot(xgrid, yhatgrid, type='l', xlim=c(-15,10), ylim=c(0,120),
+      col=cores[5], xlab='x', ylab='y', sub='100 amostras, p = 5, modelo balanceado')
```



100 amostras, p = 5, modelo balanceado

3.6 p = 6

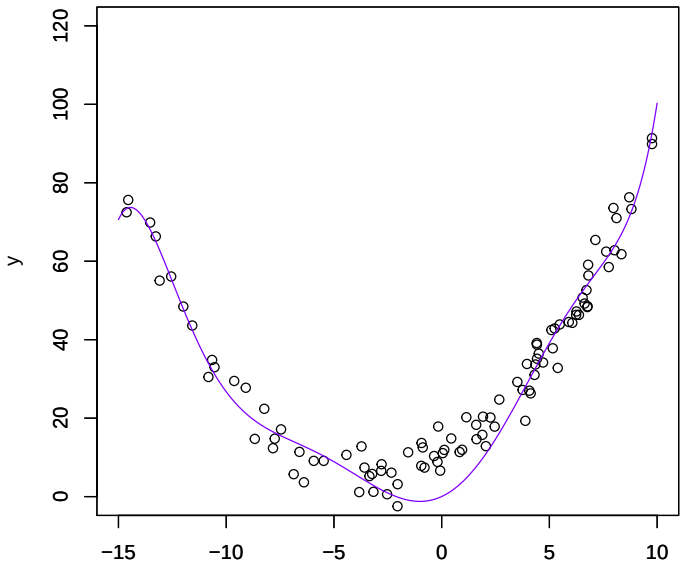
```
> H <- cbind(X^6, X^5, X^4, X^3, X^2, X, 1)
> w <- pseudoinverse(H) %*% Y
> Hgrid <- cbind(xgrid^6, xgrid^5, xgrid^4, xgrid^3, xgrid^2, xgrid, 1)
> yhat <- H %*% w
> yhatgrid <- Hgrid %*% w
> plot(X, Y, type='p', xlim=c(-15,10), ylim=c(0,120), xlab='', ylab='')
> par(new=T)
> plot(xgrid, yhatgrid, type='l', xlim=c(-15,10), ylim=c(0,120),
+      col=cores[6], xlab='x', ylab='y', sub='100 amostras, p = 6, modelo balanceado')
```



100 amostras, p = 6, modelo balanceado

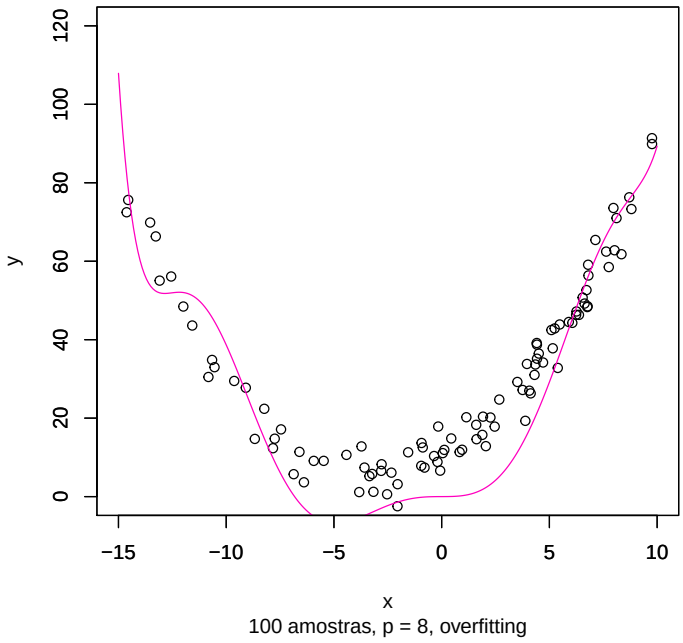
3.7 p = 7

```
> H <- cbind(X^7, X^6, X^5, X^4, X^3, X^2, X, 1)
> w <- pseudoinverse(H) %*% Y
> Hgrid <- cbind(xgrid^7, xgrid^6, xgrid^5, xgrid^4, xgrid^3, xgrid^2, xgrid, 1)
> yhat <- H %*% w
> yhatgrid <- Hgrid %*% w
> plot(X, Y, type='p', xlim=c(-15,10), ylim=c(0,120), xlab='', ylab='')
> par(new=T)
> plot(xgrid, yhatgrid, type='l', xlim=c(-15,10), ylim=c(0,120),
+      col=cores[7], xlab='x', ylab='y', sub='100 amostras, p = 7, overfitting')
```



3.8 p = 8

```
> H <- cbind(X^8, X^7, X^6, X^5, X^4, X^3, X^2, X, 1)
> w <- pseudoinverse(H) %*% Y
> Hgrid <- cbind(xgrid^8, xgrid^7, xgrid^6, xgrid^5, xgrid^4, xgrid^3, xgrid^2, xgrid, 1)
> yhat <- H %*% w
> yhatgrid <- Hgrid %*% w
> plot(X, Y, type='p', xlim=c(-15,10), ylim=c(0,120), xlab='', ylab='')
> par(new=T)
> plot(xgrid, yhatgrid, type='l', xlim=c(-15,10), ylim=c(0,120),
+      col=cores[8], xlab='x', ylab='y', sub='100 amostras, p = 8, overfitting')
```



100 amostras, p = 8, overfitting