

## Especificación errónea del modelo

Particionemos el modelo de regresión lineal múltiple  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$  en la siguiente manera

$$\begin{aligned}\mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \\ &= (\mathbf{X}_1, \mathbf{X}_2) \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} + \boldsymbol{\epsilon} \\ &= \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}\end{aligned}$$

Si realizamos inferencia sin incluir  $\mathbf{X}_2\boldsymbol{\beta}_2$  cuando los datos provienen del modelo con  $\boldsymbol{\beta}_2 \neq \mathbf{0}$  entonces decimos que estamos subajustando (underfitting). Si incluimos  $\mathbf{X}_2\boldsymbol{\beta}_2$  cuando los datos provienen del modelo con  $\boldsymbol{\beta}_2 = \mathbf{0}$  decimos que estamos sobreajustando (overfitting).

Esto puede ser visualizado en el ejemplo de regresión polinomial.

```
In [1]: using Distributions
y = [ 1.0 + x + 3.0*x^2 for x in -1:0.05:1 ] .+ rand(Normal(),41)
x = collect(-1:0.05:1)
X = zeros(41,7) # Matriz de ceros para construir X
for i in 1:41
    X[i,:] = [ 1, x[i], x[i]^2, x[i]^3, x[i]^4, x[i]^5, x[i]^6 ] # iteramos para c
end
X_1 = X[:,1:2];

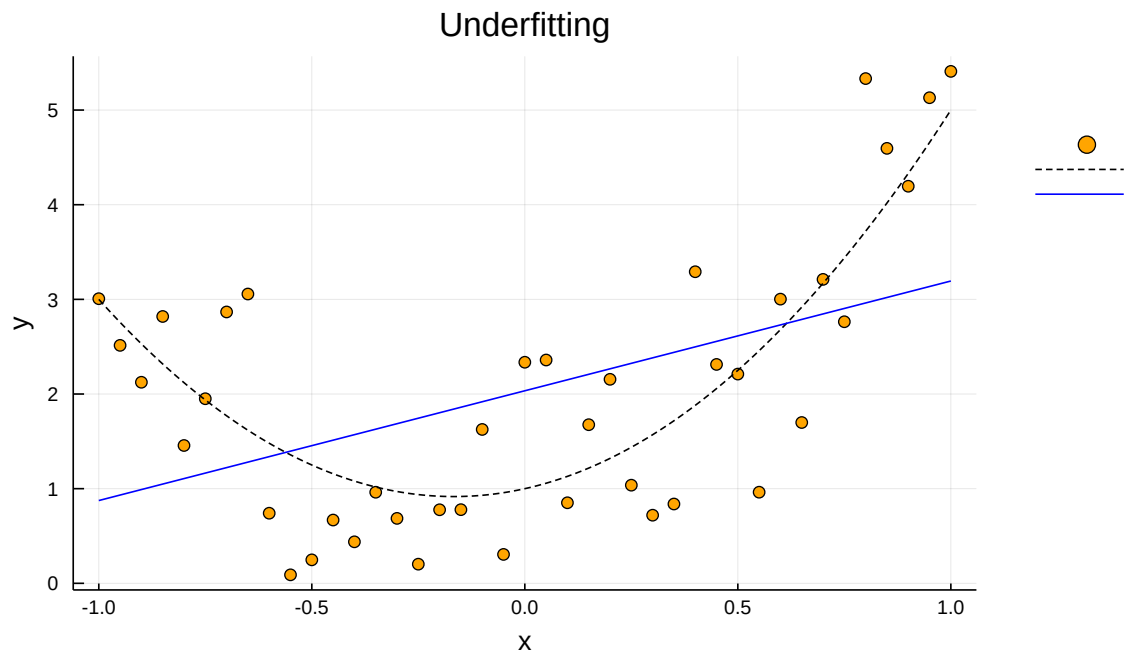
In [2]: using Plots, Measures #, Plots.PlotMeasures # Paquete para producir imágenes
default(size = (900, 400))
f(x) = 1.0 + x + 3.0*x^2 # Media cuadrática verdadera
β_u = (X_1' * X_1)^(-1) * X_1' * y
β_o = (X' * X)^(-1) * X' * y;
```

```

In [3]: # Underfitting
f_u(x) =  $\beta_u[1] + \beta_u[2]*x$  # Media lineal dada por máxima verosimilitud al subest.
mesh = collect(-1.0:1.0/100.0:1.0)
scatter(x,y,color="orange",label="Observaciones")
plot!(mesh,f.(mesh), color = :black, linestyle=:dash ,label="Media cuadrática ver")
plot!(mesh,f_u.(mesh), color = :blue, label="Media lineal estimada", legend=:oute
ylabel!("y")
xlabel!("x")
title!("Underfitting")

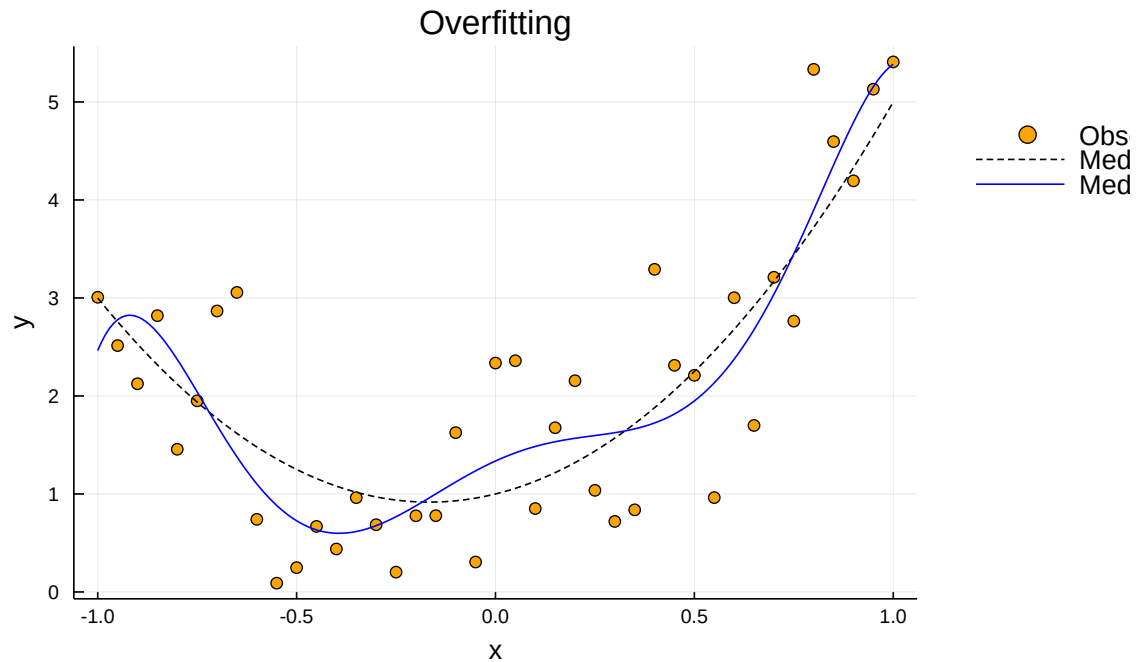
```

Out[3]:



```
In [4]: # Overfitting
f_o(x) = β_o[1] + β_o[2]*x + β_o[3]*x^2 + β_o[4]*x^3 + β_o[5]*x^4 + β_o[6]*x^5 + β_o[7]*x^6
mesh = collect(-1.0:1.0/100.0:1.0)
scatter(x,y,color="orange",label="Observaciones")
plot!(mesh,f.(mesh), color = :black, linestyle=:dash, label="Media cuadrática verdadera")
plot!(mesh,f_o.(mesh), color = :blue, label="Media polinomial grado 6 estimada",
ylabel!("y")
xlabel!("x")
title!("Overfitting")
```

Out[4]:



## Ortogonalización

Recordemos que al considerar dos modelos:

$$y = X_1 \beta_1^* + \epsilon^*$$

y

$$y = X_1 \beta_1 + X_2 \beta_2 + \epsilon$$

no se tiene necesariamente que los estimadores OLS satisfacen  $\hat{\beta}_1^* = \hat{\beta}_1$ . Observemos que si usamos el modelo con  $\beta_1^*$  cuando el modelo verdadero es el que contiene  $\beta_1$  y  $\beta_2$  en las ecuaciones anteriores, se tiene que

$$\begin{aligned} \mathbb{E}[\hat{\beta}_1^*] &= (X_1' X_1)^{-1} X_1' \mathbb{E}[y] \\ &= (X_1' X_1)^{-1} X_1' (X_1 \beta_1 + X_2 \beta_2) \\ &= \beta_1 + (X_1' X_1)^{-1} X_1' X_2 \beta_2 \end{aligned}$$

por lo que la estimación de  $\beta_1$  se ve afectada por  $X_2$  a menos que  $X_1' X_2 = 0$ . En el siguiente teorema vemos que si  $X_1' X_2 = 0$  entonces  $\hat{\beta}_1^*$  y  $\hat{\beta}_1$  más que tener el mismo valor esperado, son iguales.

**Teorema** Si  $\mathbf{X}'_1 \mathbf{X}_2 = \mathbf{0}$  entonces  $\hat{\boldsymbol{\beta}}_1$ , en el modelo completo  $\mathbf{y} = \mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{X}_2 \boldsymbol{\beta}_2 + \boldsymbol{\epsilon}$ , es igual a  $\hat{\boldsymbol{\beta}}_1^*$ , en el modelo reducido  $\mathbf{y} = \mathbf{X}_1 \boldsymbol{\beta}_1^* + \boldsymbol{\epsilon}^*$ .

Antes de proceder a la demostración enunciamos el siguiente Lemma para trabajar con la inversión de matrices expresadas en bloques

**Lemma** Si

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{1,1} & \mathbf{A}_{1,2} \\ \mathbf{A}_{2,1} & \mathbf{A}_{2,2} \end{pmatrix}$$

es no singular, simétrica y tal que  $\mathbf{A}_{1,1}^{-1}$  existe, y  $\mathbf{B} = \mathbf{A}_{2,2} - \mathbf{A}_{2,1} \mathbf{A}_{1,1}^{-1} \mathbf{A}_{1,2}$  es invertible, entonces

$$\mathbf{A}^{-1} = \begin{pmatrix} \mathbf{A}_{1,1}^{-1} + \mathbf{A}_{1,1}^{-1} \mathbf{A}_{1,2} \mathbf{B}^{-1} \mathbf{A}_{2,1} \mathbf{A}_{1,1}^{-1} & -\mathbf{A}_{1,1}^{-1} \mathbf{A}_{1,2} \mathbf{B}^{-1} \\ -\mathbf{B}^{-1} \mathbf{A}_{2,1} \mathbf{A}_{1,1}^{-1} & \mathbf{B}^{-1} \end{pmatrix}$$

**Prueba del teorema**

De  $\hat{\boldsymbol{\beta}} = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y}$  se sigue que

$$\begin{pmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{X}'_1 \mathbf{X}_1 & \mathbf{X}'_1 \mathbf{X}_2 \\ \mathbf{X}'_2 \mathbf{X}_1 & \mathbf{X}'_2 \mathbf{X}_2 \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X}'_1 \mathbf{y} \\ \mathbf{X}'_2 \mathbf{y} \end{pmatrix}$$

Usando el lemma anterior tenemos que

$$\hat{\boldsymbol{\beta}}_1 = \left( (\mathbf{X}'_1 \mathbf{X}_1)^{-1} + (\mathbf{X}'_1 \mathbf{X}_1)^{-1} \mathbf{X}'_1 \mathbf{X}_2 \mathbf{B}^{-1} \mathbf{X}'_2 \mathbf{X}_1 (\mathbf{X}'_1 \mathbf{X}_1)^{-1} \right) \mathbf{X}'_1 \mathbf{y} - (\mathbf{X}'_1 \mathbf{X}_1)^{-1} \mathbf{X}'_1 \mathbf{X}_2 \mathbf{B}^{-1} \mathbf{X}'_2 \mathbf{y}$$

Por lo que  $\mathbf{X}'_1 \mathbf{X}_2 = \mathbf{0}$  implica

$$\hat{\boldsymbol{\beta}}_1 = (\mathbf{X}'_1 \mathbf{X}_1)^{-1} \mathbf{X}'_1 \mathbf{y} = \hat{\boldsymbol{\beta}}_1^*$$

El resultado anterior será de utilidad para realizar pruebas de hipótesis.

