Fundamentos de Aprendizaje Automático y Reconocimiento de Patrones

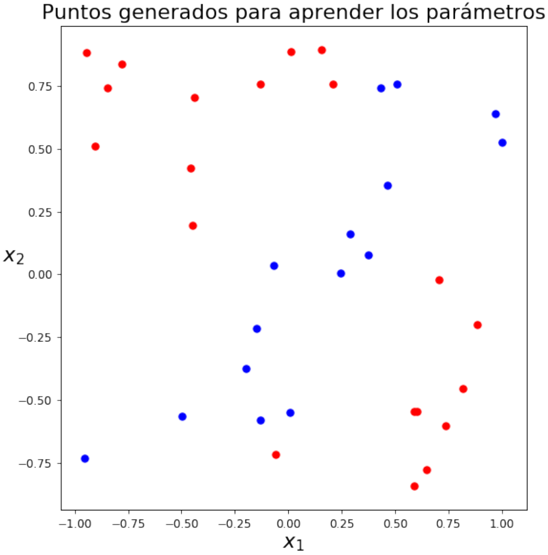
Práctico 4

Instituto de Ingeniería Eléctrica
Facultad de Ingeniería

UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

1. Ejercicio 1: regularización con *weight decay*
2. Ejercicio 2: Curvas de aprendizaje
3. Ejercicio 3: Regresión polinómica regularizada
4. Ejercicio 4: Prediccion del valor medio de las casas

Ejercicio 1: regularización con *weight decay*



Ejercicio 1: regularización con *weight decay*

La función de costo que se minimiza mediante el criterio de mínimos cuadrados es:

$$\mathrm{w}^* = \underset{\mathrm{w}\in\mathbb{R}^{d+1}}{\operatorname{argmin}} \quad E_{\mathrm{in}}(\mathrm{w}) = \underset{\mathrm{w}\in\mathbb{R}^{d+1}}{\operatorname{argmin}} \quad \frac{1}{N}\|\mathrm{Xw}-\mathrm{y}\|^2$$

- Solución cerrada

$$\mathrm{w}^* = \mathrm{X}^\dagger \mathrm{y} = \left(\mathrm{X}^{\mathrm{T}}\mathrm{X}\right)^{-1}\mathrm{X}^{\mathrm{T}}\mathrm{y}$$

La función de costo que se minimiza para implementar *weight decay*:

$$\mathrm{w}^* = \underset{\mathrm{w}\in\mathbb{R}^{d+1}}{\operatorname{argmin}} \quad E_{\mathrm{aug}}(\mathrm{w}) = \underset{\mathrm{w}\in\mathbb{R}^{d+1}}{\operatorname{argmin}} \quad E_{\mathrm{in}}(\mathrm{w}) + \frac{\lambda}{N}\mathrm{w}^{\mathrm{T}}\mathrm{w}$$

Consigna: Hallar la solución cerrada en este caso e implementarla (*regresion_Ridge()*)

Ejercicio 1: regularización con *weight decay*

La función de costo que se minimiza mediante el criterio de mínimos cuadrados es:

$$\mathrm{w}^* = \underset{\mathrm{w}\in\mathbb{R}^{d+1}}{\operatorname{argmin}} \quad E_{\mathrm{in}}(\mathrm{w}) = \underset{\mathrm{w}\in\mathbb{R}^{d+1}}{\operatorname{argmin}} \quad \frac{1}{N}\|\mathrm{Xw}-\mathrm{y}\|^2$$

- Solución cerrada

$$\mathrm{w}^* = \mathrm{X}^\dagger \mathrm{y} = \left(\mathrm{X}^{\mathrm{T}}\mathrm{X}\right)^{-1}\mathrm{X}^{\mathrm{T}}\mathrm{y}$$

La función de costo que se minimiza para implementar *weight decay*:

$$\mathrm{w}^* = \underset{\mathrm{w}\in\mathbb{R}^{d+1}}{\operatorname{argmin}} \quad E_{\mathrm{aug}}(\mathrm{w}) = \underset{\mathrm{w}\in\mathbb{R}^{d+1}}{\operatorname{argmin}} \quad E_{\mathrm{in}}(\mathrm{w}) + \frac{\lambda}{N}\mathrm{w}^{\mathrm{T}}\mathrm{w}$$

Consigna: Hallar la solución cerrada en este caso e implementarla (*regresion_Ridge()*)

Ejercicio 1: regularización con *weight decay*

La función de costo que se minimiza mediante el criterio de mínimos cuadrados es:

$$\mathrm{w}^* = \underset{\mathrm{w}\in\mathbb{R}^{d+1}}{\operatorname{argmin}} \quad E_{\mathrm{in}}(\mathrm{w}) = \underset{\mathrm{w}\in\mathbb{R}^{d+1}}{\operatorname{argmin}} \quad \frac{1}{N}\|\mathrm{X}\mathrm{w}-\mathrm{y}\|^2$$

- Solución cerrada

$$\mathrm{w}^* = \mathrm{X}^{\dagger}\mathrm{y} = \left(\mathrm{X}^{\mathrm{T}}\mathrm{X}\right)^{-1}\mathrm{X}^{\mathrm{T}}\mathrm{y}$$

La función de costo que se minimiza para implementar *weight decay*:

$$\mathrm{w}^* = \underset{\mathrm{w}\in\mathbb{R}^{d+1}}{\operatorname{argmin}} \quad E_{\mathrm{aug}}(\mathrm{w}) = \underset{\mathrm{w}\in\mathbb{R}^{d+1}}{\operatorname{argmin}} \quad E_{\mathrm{in}}(\mathrm{w}) + \frac{\lambda}{N}\mathrm{w}^{\mathrm{T}}\mathrm{w}$$

Consigna: Hallar la solución cerrada en este caso e implementarla (*regresion_Ridge()*)

Ejercicio 1: regularización con *weight decay*

La función de costo que se minimiza mediante el criterio de mínimos cuadrados es:

$$\mathrm{w}^* = \underset{\mathrm{w}\in\mathbb{R}^{d+1}}{\operatorname{argmin}} \; E_{\mathrm{in}}(\mathrm{w}) = \underset{\mathrm{w}\in\mathbb{R}^{d+1}}{\operatorname{argmin}} \; \frac{1}{N}\|\mathrm{Xw} - \mathrm{y}\|^2$$

- Solución cerrada

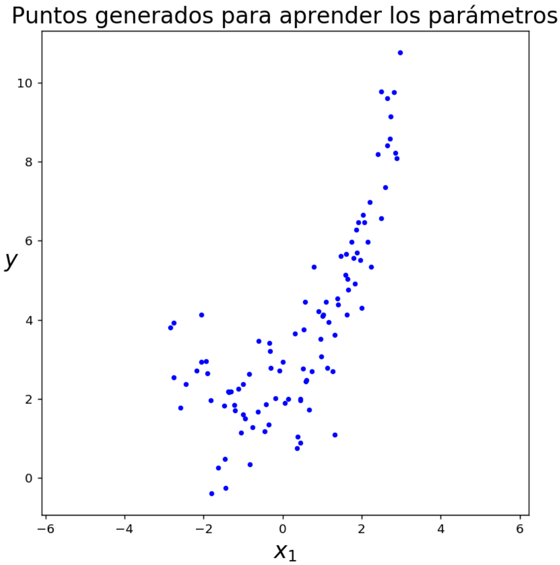
$$\mathrm{w}^* = \mathrm{X}^{\dagger}\mathrm{y} = \left(\mathrm{X}^{\mathrm{T}}\mathrm{X}\right)^{-1}\mathrm{X}^{\mathrm{T}}\mathrm{y}$$

La función de costo que se minimiza para implementar *weight decay*:

$$\mathrm{w}^* = \underset{\mathrm{w}\in\mathbb{R}^{d+1}}{\operatorname{argmin}} \; E_{\mathrm{aug}}(\mathrm{w}) = \underset{\mathrm{w}\in\mathbb{R}^{d+1}}{\operatorname{argmin}} \; E_{\mathrm{in}}(\mathrm{w}) + \frac{\lambda}{N}\mathrm{w}^{\mathrm{T}}\mathrm{w}$$

Consigna: Hallar la solución cerrada en este caso e implementarla (*regresion_Ridge()*)

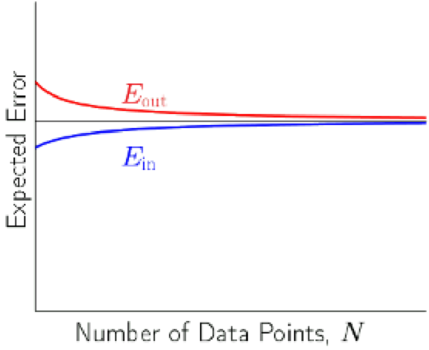
Ejercicio 2: curvas de aprendizaje

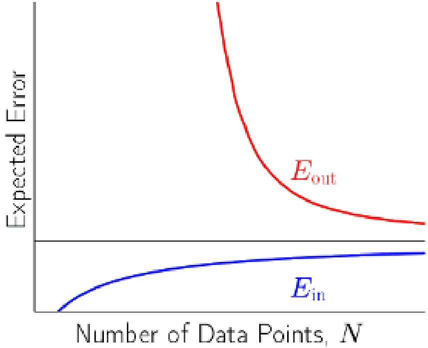


Ejercicio 2: curvas de aprendizaje

- Implementar *transformacion_polinomica()*
- Utilizar *regresion_lineal()* del **Ejercicio 1**
- Mostrar el resultado de aplicar el modelo a un conjunto de datos de validación

Ejercicio 2: curvas de aprendizaje





- Implementar *generar_curva_aprendizaje()*
- Implementar *mse()*

$$mse = \frac{1}{N}\sum_{n=1}^{N}\left(y_n - \mathrm{w}^T\mathrm{x}\right)^2$$

Ejercicio 2: curvas de aprendizaje

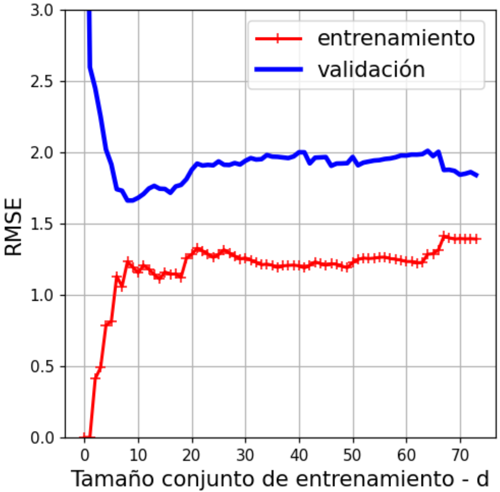


Figure: Curva de aprendizaje modelo lineal

Ejercicio 2: curvas de aprendizaje

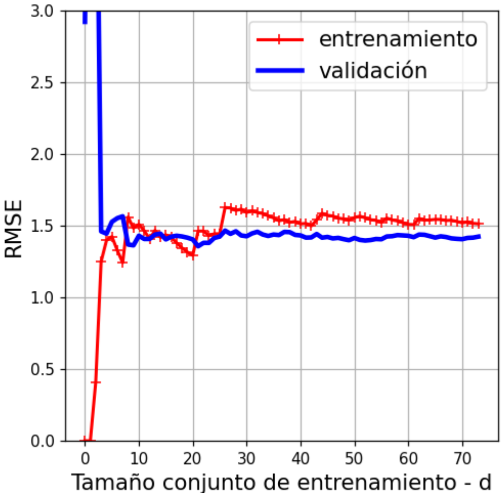


Figure: Curva de aprendizaje modelo lineal

Ejercicio 2: curvas de aprendizaje

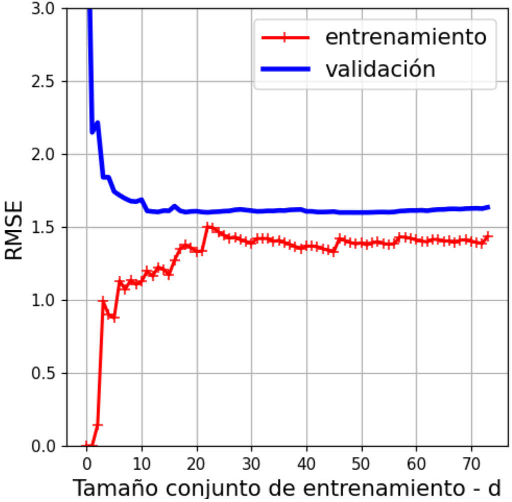


Figure: Curva de aprendizaje modelo lineal

Ejercicio 2: curvas de aprendizaje

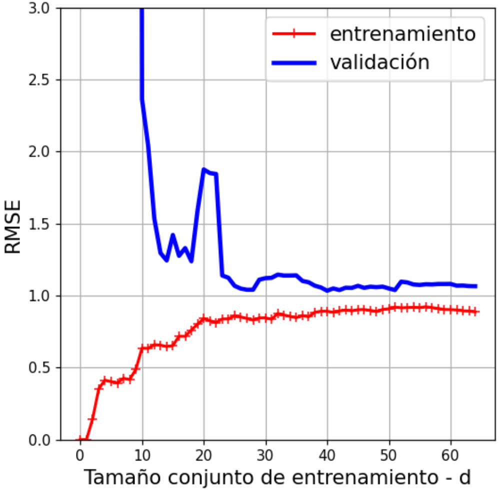


Figure: Curva de aprendizaje modelo orden 10

Ejercicio 2: curvas de aprendizaje

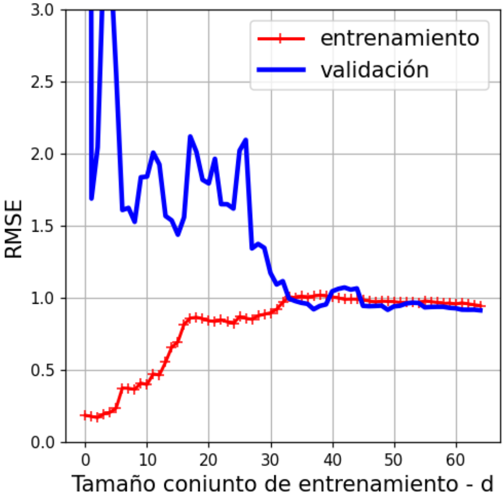


Figure: Curva de aprendizaje modelo orden 10

Ejercicio 2: curvas de aprendizaje

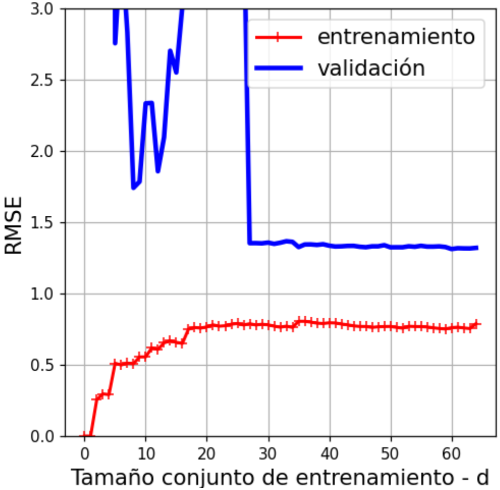
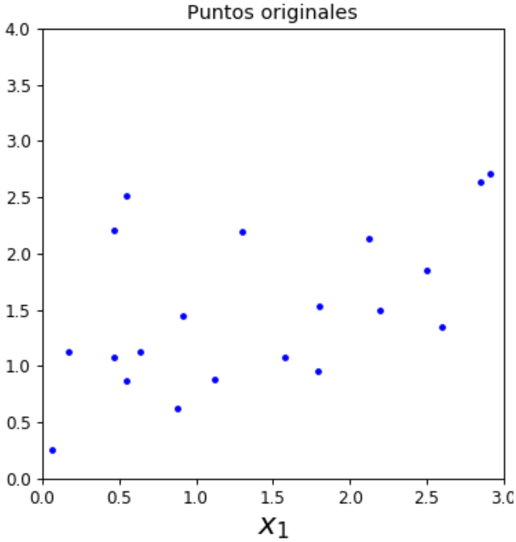


Figure: Curva de aprendizaje modelo orden 10

Ejercicio 3: regresión polinómica regularizada



Ejercicio 3: regresión polinómica regularizada

- Utilizar *regresion_Ridge()* del **Ejercicio 1**
- Implementar *comparar_regularizaciones()*
  - *transformacion_polinomica()* (ya implementado)
  - *regresion_Ridge()* (ya implementado)

Ejercicio 3: regresión polinómica regularizada

- En la regresión de Ridge se minimiza el error aumentado

$$
\begin{aligned}
E_{\text{aug}}(\mathrm{w}) &= E_{\text{in}}(\mathrm{w}) + \frac{\lambda}{N}\mathrm{w}^{\intercal}\mathrm{w} \\
&= \frac{1}{N}\sum_{n=1}^{N}(\mathrm{w}^{\intercal}\mathrm{z}_n - y_n)^2 + \frac{\lambda}{N}\mathrm{w}^{\intercal}\mathrm{w} \\
&= \frac{1}{N}(\mathrm{Zw} - \mathrm{y})^{\intercal}(\mathrm{Zw} - \mathrm{y}) + \frac{\lambda}{N}\mathrm{w}^{\intercal}\mathrm{w} \\
&= \frac{1}{N}(\mathrm{w}^{\intercal}\mathrm{Z}^{\intercal}\mathrm{Zw} - 2\mathrm{w}^{\intercal}\mathrm{Z}^{\intercal}\mathrm{y} + \mathrm{y}^{\intercal}\mathrm{y}) + \frac{\lambda}{N}\mathrm{w}^{\intercal}\mathrm{w}
\end{aligned}
$$

- Usando que

$$\nabla_{\mathrm{w}}(\mathrm{w}^{\intercal}\mathrm{Aw}) = (\mathrm{A} + \mathrm{A}^{\intercal})\mathrm{w}, \quad \nabla_{\mathrm{w}}(\mathrm{w}^{\intercal}\mathrm{b}) = \mathrm{b}, \quad \nabla_{\mathrm{w}}(\mathrm{w}^{\intercal}\mathrm{w}) = 2\mathrm{w}$$

- Se obtiene

$$\mathrm{w}^* = (\mathrm{Z}^{\intercal}\mathrm{Z} + \lambda \mathrm{I})^{-1}\mathrm{Z}^{\intercal}\mathrm{y}$$

Ejercicio 3: regresión polinómica regularizada

```
def regresion_Ridge(X, y, reg, regularizar_bias=False):

    '''
    Implementa la solución cerrada de la regresión de Ridge. En esta variante se
    permite elegir si se desea regularizar el bias (coeficiente w_0) o no.
    Entrada:
        X: matriz de tamaño Nx(d+1)
        y: valores objetivo
        reg: coeficiente que multiplica al término de regularización
        regularizar_bias: variable booleana que indica si se regulariza el término
                          de bias o no.
    Salida:
        w_reg: parámetros encontrados mediante la regularización de Ridge
    '''

    # Si no se regulariza el coeficiente w_0 se pone el elemento [0,0]
    # de la matríz identidad a cero
    reg_matrix = np.identity(X.shape[1])
    if not regularizar_bias:
        reg_matrix[0,0]=0

    ######################################################################
    ############# EMPIEZA ESPACIO PARA COMPLETAR CÓDIGO  ##################
    ######################################################################

    # w_reg =

    ######################################################################
    ############# TERMINA ESPACIO PARA COMPLETAR CÓDIGO  ##################
    ######################################################################


    return w_reg
```

Ejercicio 3: regresión polinómica regularizada

```
def regresion_Ridge(X, y, reg, regularizar_bias=False):

    '''
    Implementa la solución cerrada de la regresión de Ridge. En esta variante se
    permite elegir si se desea regularizar el bias (coeficiente w_0) o no.
    Entrada:
        X: matriz de tamaño Nx(d+1)
        y: valores objetivo
        reg: coeficiente que multiplica al término de regularización
        regularizar_bias: variable booleana que indica si se regulariza el término
                          de bias o no.
    Salida:
        w_reg: parámetros encontrados mediante la regularización de Ridge
    '''

    # Si no se regulariza el coeficiente w_0 se pone el elemento [0,0]
    # de la matríz identidad a cero
    reg_matrix = np.identity(X.shape[1])
    if not regularizar_bias:
        reg_matrix[0,0]=0

    ######################################################################
    ############# EMPIEZA ESPACIO PARA COMPLETAR CÓDIGO  ##################
    ######################################################################

    # w_reg =
    w_reg = np.linalg.inv(X.T @ X + reg * reg_matrix) @ X.T @ y

    ######################################################################
    ############# TERMINA ESPACIO PARA COMPLETAR CÓDIGO  ##################
    ######################################################################


    return w_reg
```

Efecto de regularizar el bias

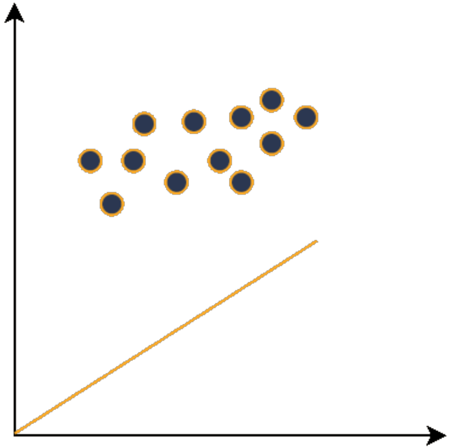


Figure: Regularizando fuertemente el bias

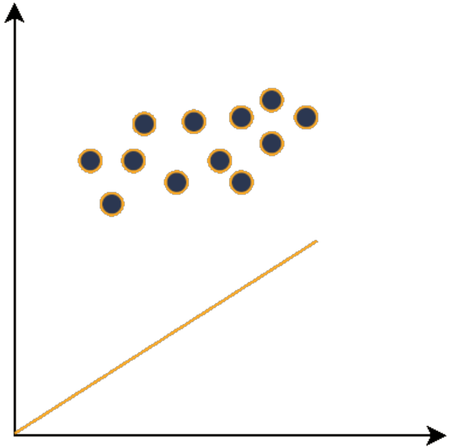


Figure: Sin regularizar el bias

Efecto de regularizar el bias

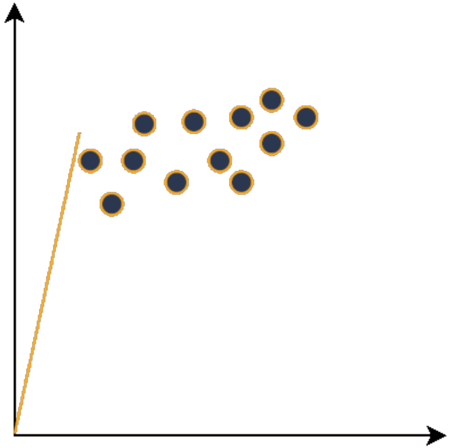


Figure: Regularizando fuertemente el bias

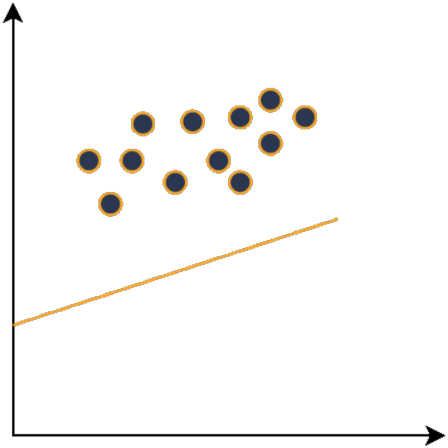


Figure: Sin regularizar el bias

Efecto de regularizar el bias

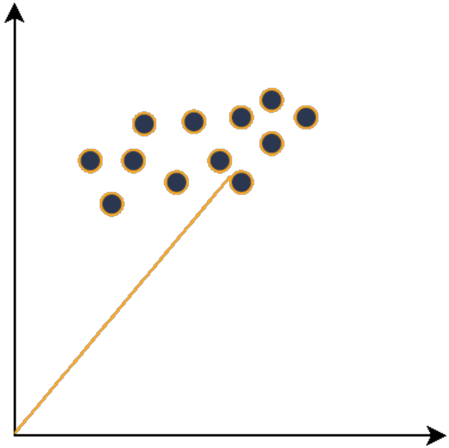


Figure: Regularizando fuertemente el bias

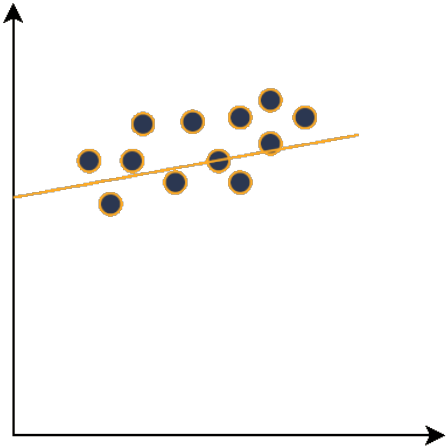


Figure: Sin regularizar el bias

Efecto de regularizar el bias

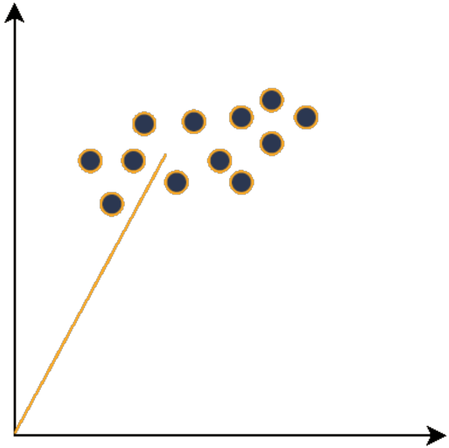


Figure: Regularizando fuertemente el bias

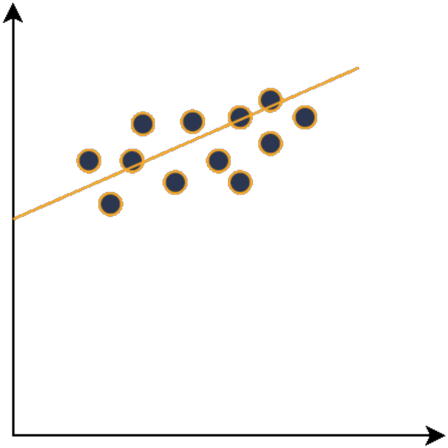


Figure: Sin regularizar el bias

Efecto de regularizar el bias

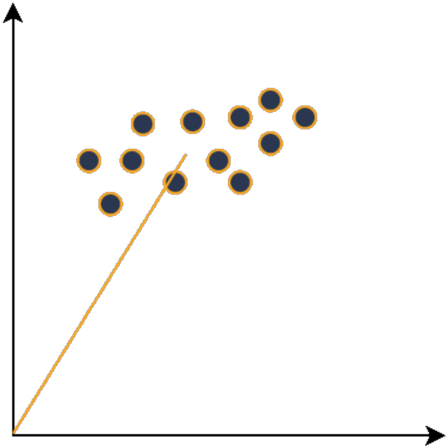


Figure: Regularizando fuertemente el bias

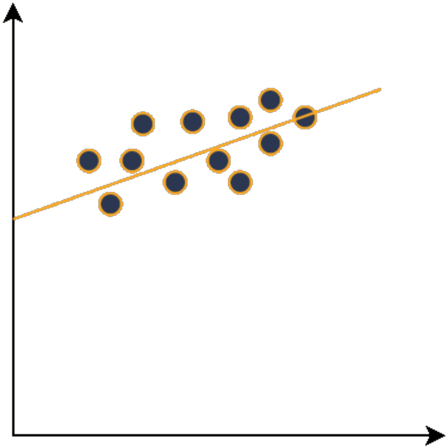
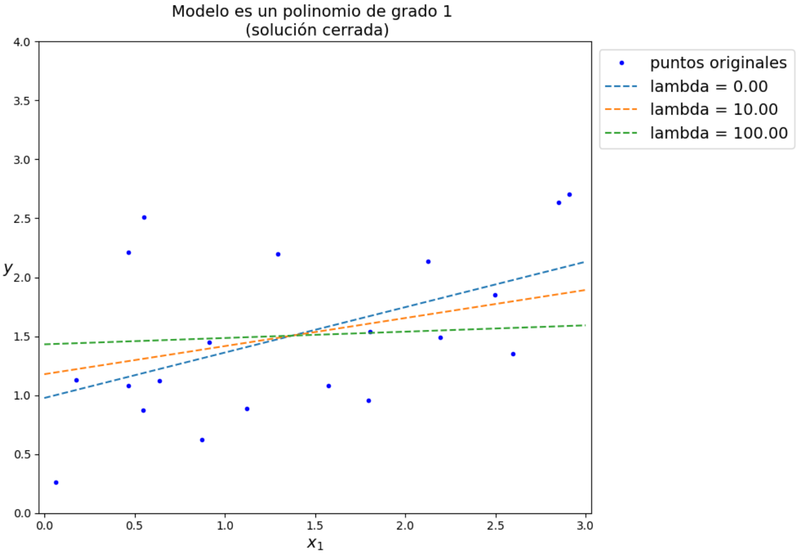
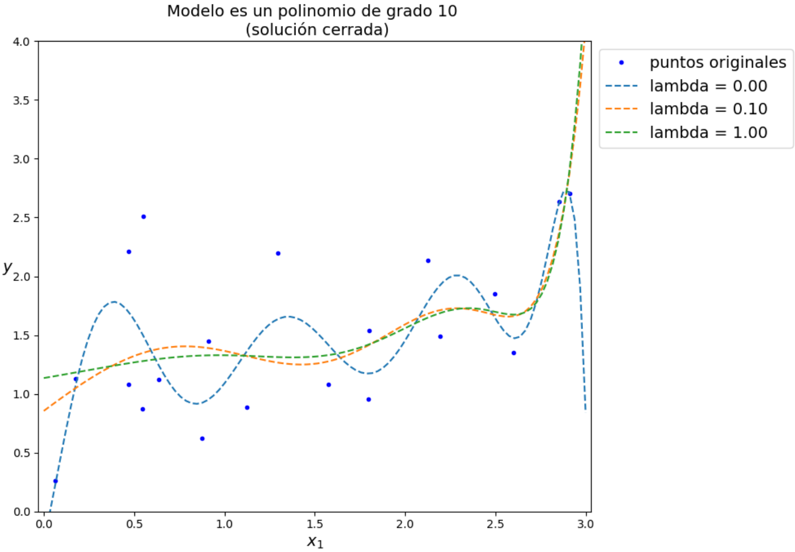


Figure: Sin regularizar el bias

Ejercicio 3: regresión polinómica regularizada



Ejercicio 3: regresión polinómica regularizada



Ejercicio 3: regularización mediante descenso por gradiente estocástico

- Implementar *regresion_lineal_SGD()*
- Se deberán contemplar los siguientes casos:
  - *Sin regularización*

$$E_{in} = \frac{1}{2N}\sum_{n=1}^{N}\left(y_n - \mathrm{w}^T\mathrm{x_n}\right)^2$$

  - *Regularización de Ridge*

$$\begin{aligned} E_{aug} &= E_{in} + E_{Ridge} \\ &= E_{in} + \frac{\lambda}{2}\sum_{i=1}^{d} w_i^2 \end{aligned}$$

  - *Regularización de Lasso*

$$\begin{aligned} E_{aug} &= E_{in} + E_{Lasso} \\ &= E_{in} + \lambda\sum_{i=1}^{d} |w_i| \end{aligned}$$

Ejercicio 3: regularización mediante descenso por gradiente estocástico

- Implementar *regresion_lineal_SGD()*
- Se deberán contemplar los siguientes casos:
  - *Sin regularización*

$$E_{in} = \frac{1}{2N} \sum_{n=1}^{N} \left(y_n - \mathrm{w}^T \mathrm{x_n} \right)^2$$

  - *Regularización de Ridge*

$$\begin{aligned} E_{aug} &= E_{in} + E_{Ridge} \\ &= E_{in} + \frac{\lambda}{2} \sum_{i=1}^{d} w_i^2 \end{aligned}$$

  - *Regularización de Lasso*

$$\begin{aligned} E_{aug} &= E_{in} + E_{Lasso} \\ &= E_{in} + \lambda \sum_{i=1}^{d} |w_i| \end{aligned}$$

Ejercicio 3: regularización mediante descenso por gradiente estocástico

- Implementar *regresion_lineal_SGD()*
- Se deberán contemplar los siguientes casos:
  - *Sin regularización*

$$E_{in} = \frac{1}{2N} \sum_{n=1}^{N} \left(y_n - \mathrm{w}^T \mathrm{x_n} \right)^2$$

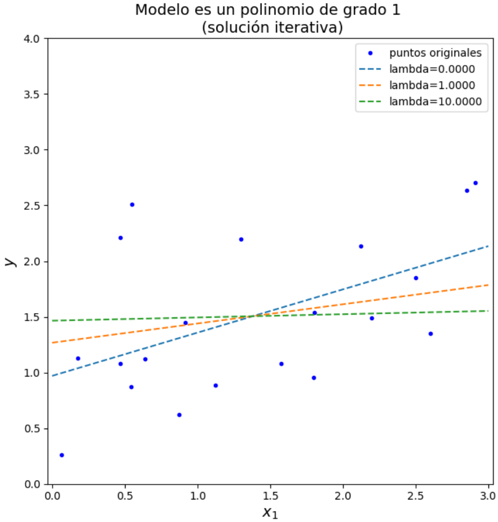
  - *Regularización de Ridge*

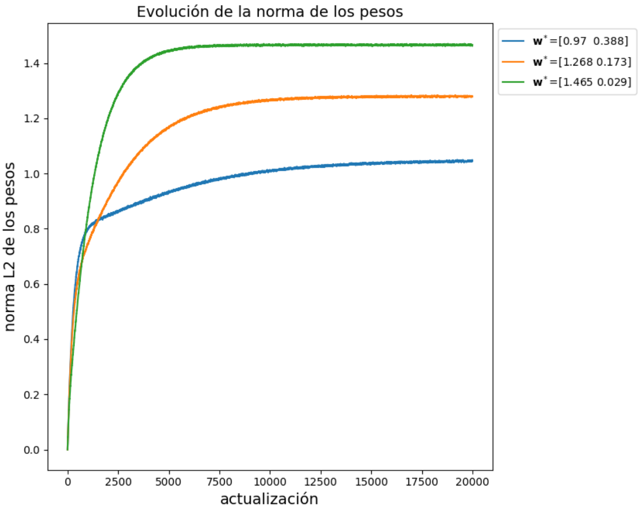
$$\begin{aligned} E_{aug} &= E_{in} + E_{Ridge} \\ &= E_{in} + \frac{\lambda}{2} \sum_{i=1}^{d} w_i^2 \end{aligned}$$

  - *Regularización de Lasso*

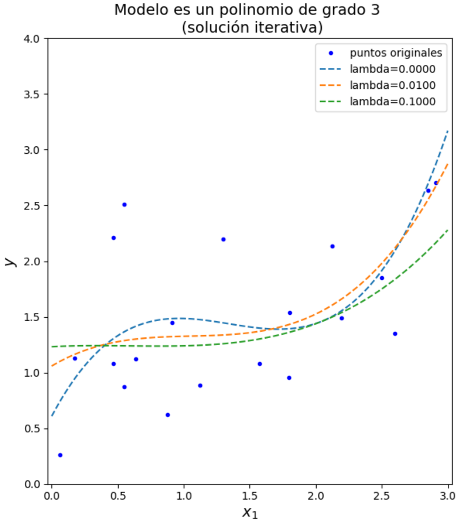
$$\begin{aligned} E_{aug} &= E_{in} + E_{Lasso} \\ &= E_{in} + \lambda \sum_{i=1}^{d} |w_i| \end{aligned}$$

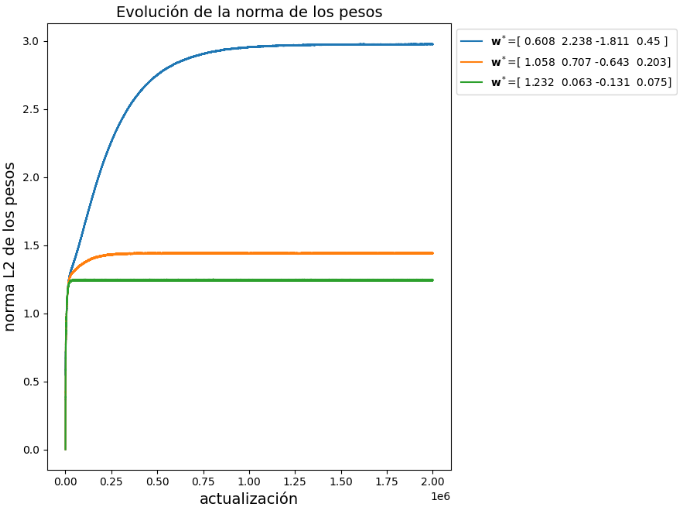
Ejercicio 3: regularización mediante descenso por gradiente estocástico



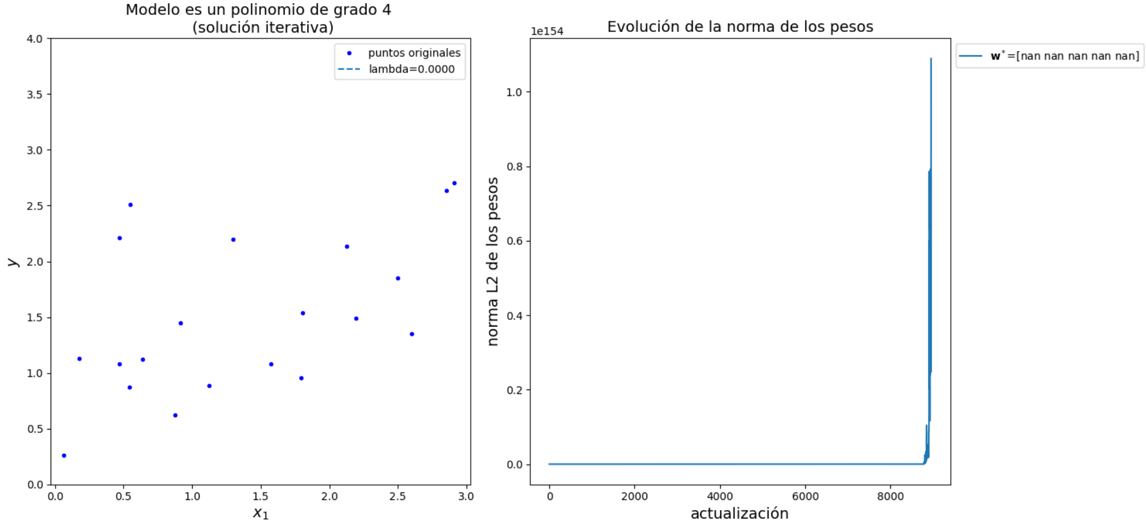


Ejercicio 3: regularización mediante descenso por gradiente estocástico

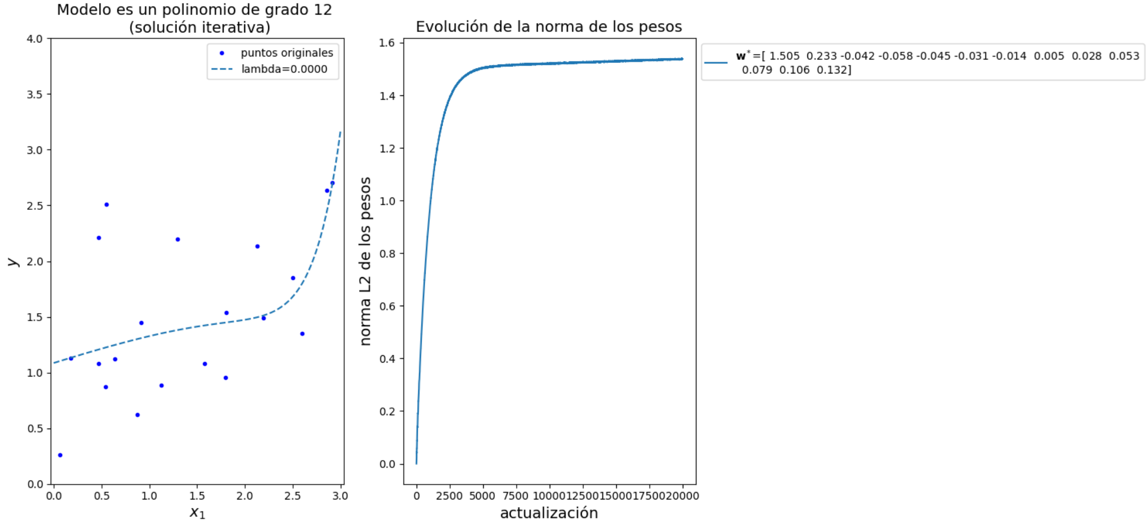




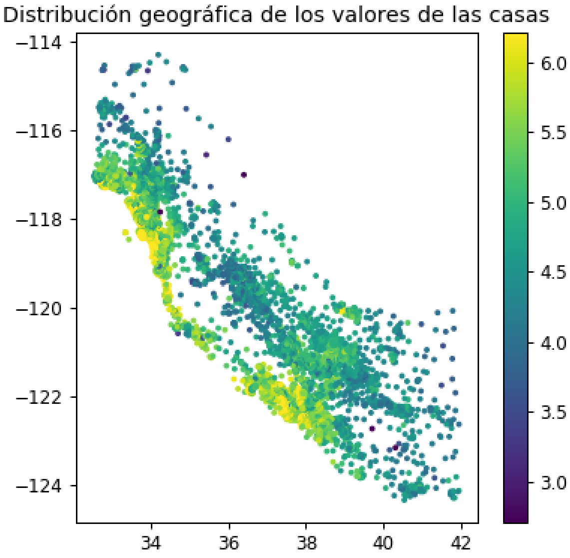
Ejercicio 3: regularización mediante descenso por gradiente estocástico



Ejercicio 3: regularización mediante descenso por gradiente estocástico



Ejercicio 4: predicción del valor medio de las casas



Ejercicio 4: predicción del valor medio de las casas