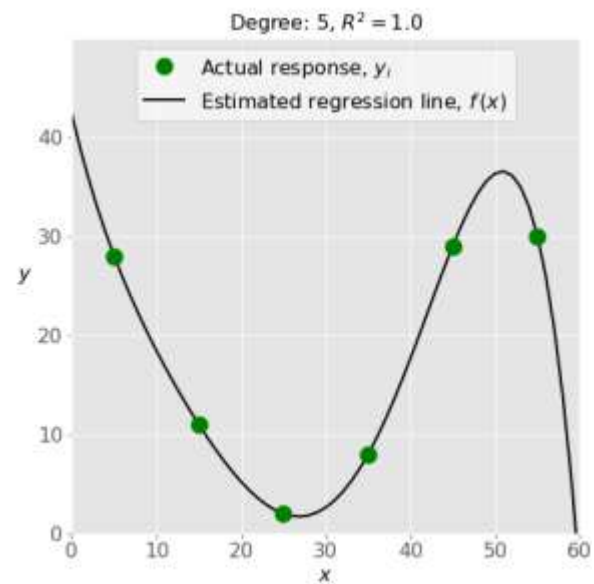
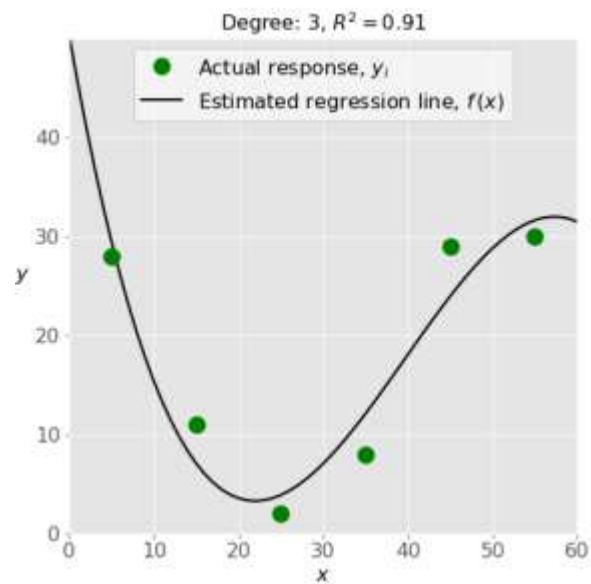
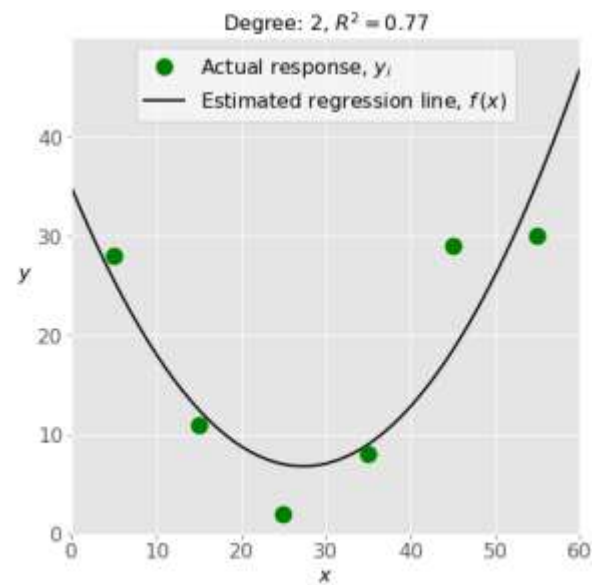
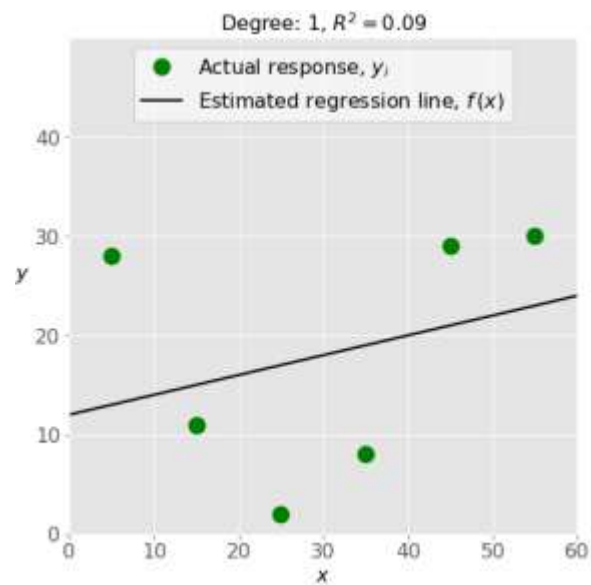


Regresión Polinomial





- La Lineal:

$$y = w_1x + w_0$$

- Polinomial de grado 3

$$y = w_3x^3 + w_2x^2 + w_1x + w_0$$

$$y = w_3x^3 + w_2x^2 + w_1x^1 + w_0x^0$$

Es igual...

- Los pasos no son diferentes a la regresión lineal
- La función de pérdida
- Se calcula la derivada
- Se usa la forma matricial
- Gradient Descent, SGD, OLS, etc...

Con Algunos cambios

Nuestra entrada:

$$(x, x^2, x^3)$$

Podemos mapearlo a otra variable

$$\Phi(x) = \begin{bmatrix} 1 \\ x \\ x^2 \\ x^3 \end{bmatrix}$$

- 1 variable de grado 1: $y = w_0 + w_1x$
 - $\Phi(x) = [1 \quad x]$
 - $y = w_0\Phi_0 + w_1\Phi_1$
-
- 2 variables de grado 1 : $y = w_0 + w_1x_1 + w_2x_2$
 - $\Phi(x) = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \end{bmatrix}$
 - $y = w_0 + w_1\Phi_{1,1} + w_2\Phi_{1,2}$

- 1 variable de grado 2: $y = w_0 + w_1x + w_2x^2$

- $\Phi(x) = [1 \quad x \quad x^2]$

- $y = w_0\Phi_0 + w_1\Phi_1 + w_2\Phi_2$

- 2 variables de grado 2 : $y = w_0 + w_1x_1 + w_2x_2 + w_3x_1^2 + w_4x_2^2$

- $\Phi(x) = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \end{bmatrix}$

- $y = w_0 + w_1\Phi_{1,1} + w_2\Phi_{1,2} + w_3\Phi_{2,1} + w_4\Phi_{2,2}$

- Para múltiples observaciones y múltiples grados...

$$\Phi(X) = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \end{bmatrix}$$

La predicción:

En lugar de

$$y = XW = W^T X$$

Se vuelve:

$$y = \Phi(x)W = W^T \Phi(x)$$

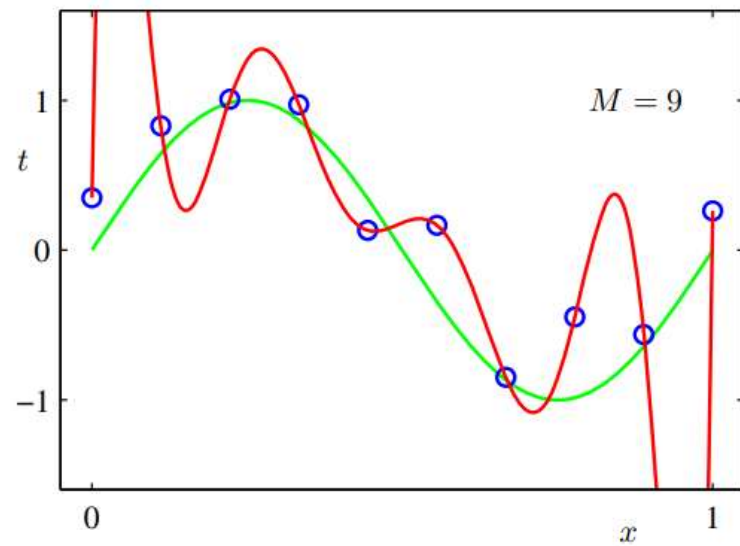
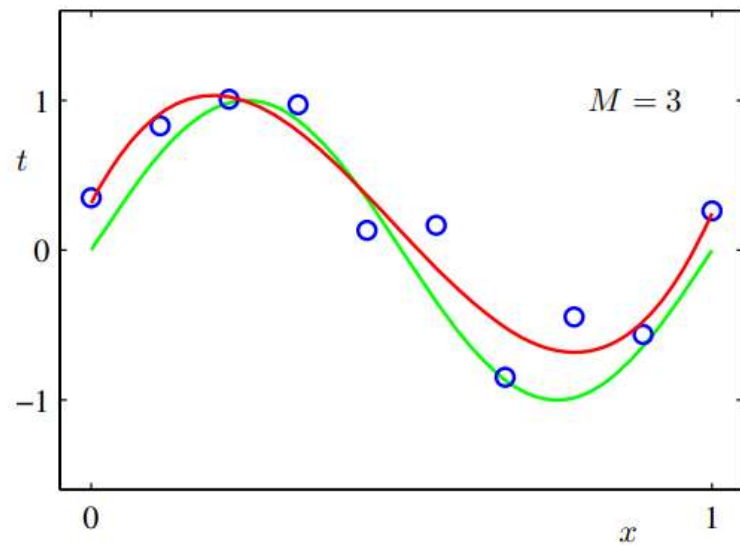
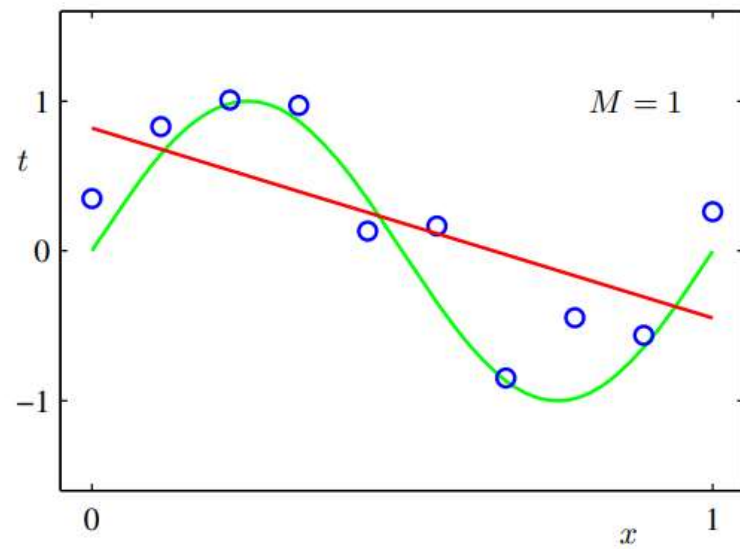
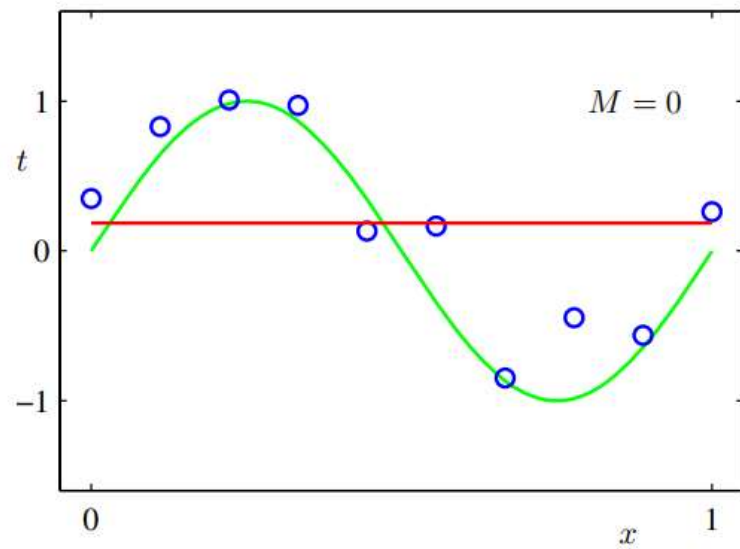
Sigue siendo Regresión Lineal

Con respecto a W

La complejidad de x queda absorbida en la función Φ

Modelo potente = bueno?

- Mientras más dimensiones mejor?
- Mientras mas complejo mejor?
- Generalización: habilidad del modelo de predecir data no vista





Nuestro modelo con $M=9$ overfit

Está modelando el ruido

Los pesos se vuelven grandes

Tenemos que motivar al modelo a
usar pesos pequeños

Regularización

Problema: Muchos input features pueden complicar el modelo rápidamente

Solución: Regularizar para mantenerlo simple

- Modificamos la función de pérdida:

$$\mathcal{L}(W) = \sum_{i=0}^n (y_i - w^T x_i)^2 + \lambda w^2$$

$$\lambda \|w\|$$

- Modificamos el Update Rule

$$w \leftarrow w - \alpha \left[\sum_{i=1}^n (x_i (wx_i - y_i)) - \lambda w \right]$$

Tipos de regularización

Ridge regression
usa la norma L2
→ λw^2

Lasso regression
usa la norma L1
→ $\lambda ||w||$

- Hay que elegir lambda con cuidado

