REDES NEURONALES 2024

Clase 23 parte 2

Lunes 11 de noviembre 2024

FAMAF, Universidad Nacional de Córdoba

Instituto de Física Enrique Gaviola (UNC-CONICET)

M7/11/23 c21p3

EL PROBLEMA DE LA REGRESIÓN POLINOMIAL

Volvamos al problema de regresión pero asumiendo como modelo un polinomio de grado arbitrario. Mostraremos un conjunto de puntos medidos generados con cierto modelo y trataremos de ver cómo el grado del polinomio interpolante impacta en la capacidad de hacer buenas estimaciones en regiones en las cuales no disponemos de mediciones.

Supongamos que tenemos una única variable x ($N=1$) y una única variable de salida y. Por cierto esto se puede generalizar a dimensiones arbitrarias.

Sea f(x) el modelo que explica el comportamiento de (x_i, y_i). Como todo proceso de medición, involucra aleatoriedad. Esto quiere decir que cuando medimos:

$$y_i = f(x_i) + \eta_i$$

no en la región de aprendizaje

donde η_i es una variable aleatoria a la cual supondremos que podemos describir por una distribución normal (gaussiana). Suponemos que no hay sesgo ni correlaciones:

$$\langle \eta_i \rangle = 0, \qquad \langle \eta_i \eta_j \rangle = \delta_{ij} \sigma^2$$

Vamos a llamar

$$g_\alpha(x; w_\alpha)$$

a nuestro modelo, que será un polinomio. Notemos que:

- α es el orden del polinomio modelo;
- x es la variable independiente y
- w_α representa el conjunto de parámetros del polinomio (del modelo)

Nosotros usaremos solo tres modelos:

1. Polinomio de grado uno

$$g_1(x; \{w\}_1) = g_1(x; \{w\}_1) = w_0 + w_1 x$$

$$\alpha = 1$$

$$\{w\}_1 = \{w_0, w_1\}$$

2. Polinomio de grado tres

$$g_3(x; \{w\}) = g_3(x; \{w\}_1) = w_0 + w_1 x + w_2 x^2 + w_3 x^3$$

$$\alpha = 3$$

$$\{w\}_1 = \{w_0, w_1, w_2, w_3\} \qquad \alpha + 1 = 4$$

3. Polinomio de grado diez

$$g_3(x;\{w\}) = g_3(x;\{w\}_1) = w_0 + w_1 x + \qquad + w_9 x^9 + w_{10} x^{10}$$

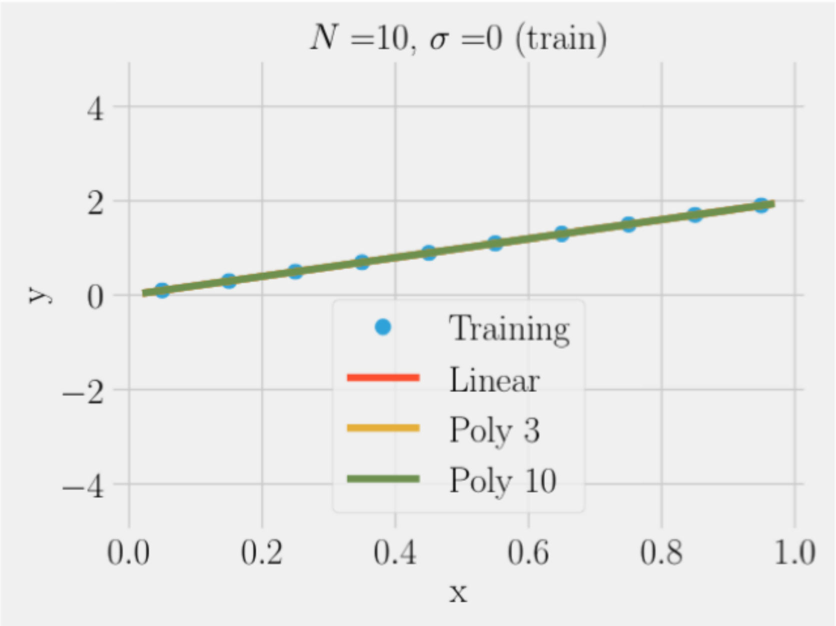
$$\alpha = 10$$

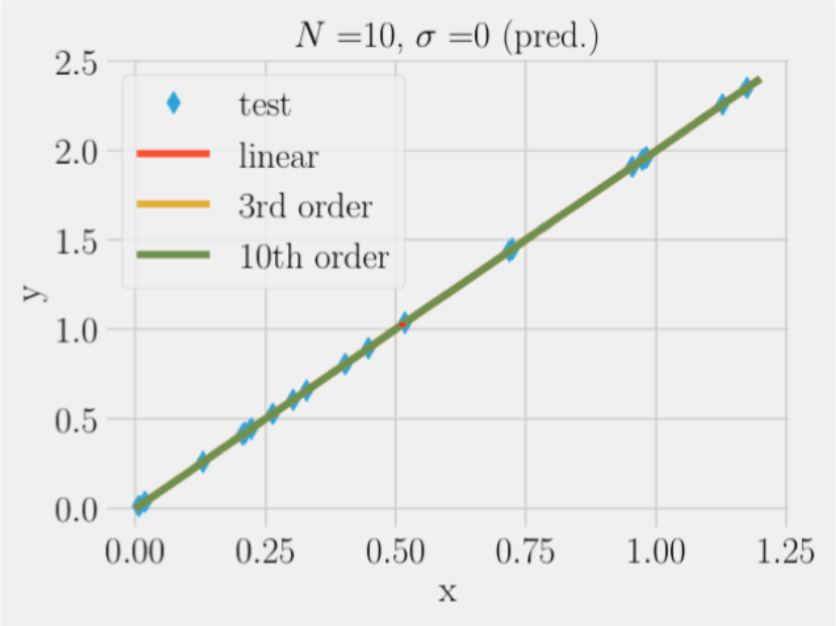
$$\{w\}_1 = \{w_0, w_1, w_2, w_3, w_4, w_5, w_6, w_7, w_8, w_9, w_{10}\}$$

$$\alpha + 1 = 11$$

El caso $\sigma = 0$ y $N_{train} = 100$

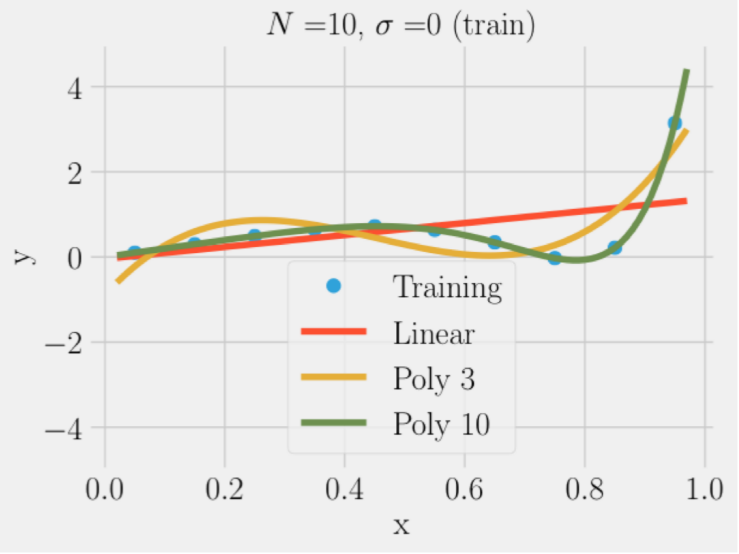
$$g_1(x; w_0, w_1) = g_1(x; \{w\}_1) \qquad = w_0 + w_1 x$$

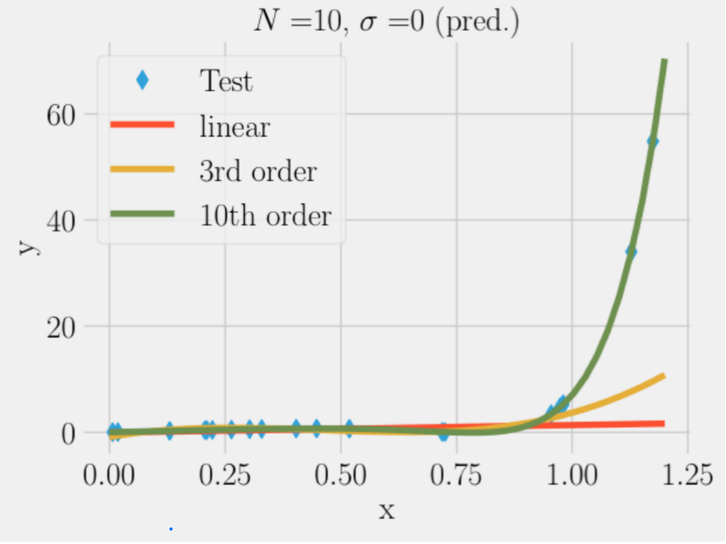




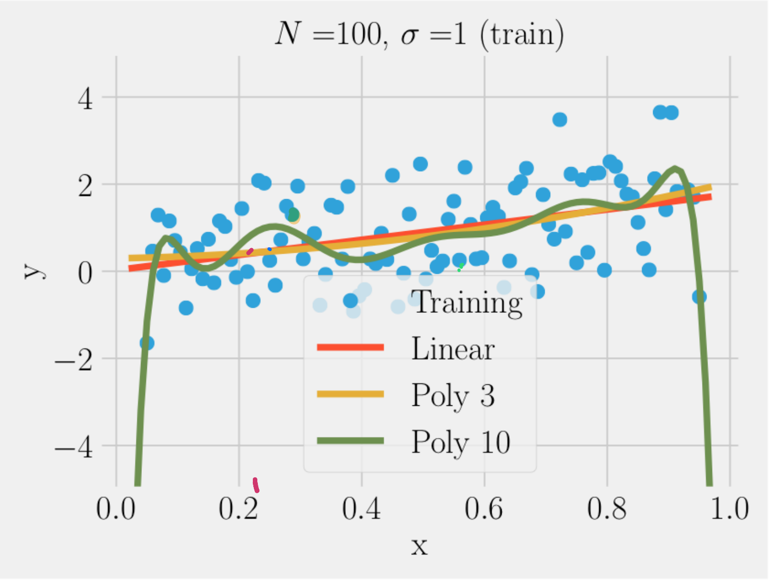
$$g_{10}(x, \{w_0, w_1, \dots w_{10}\}) = g(x; \{w\}_{10})$$

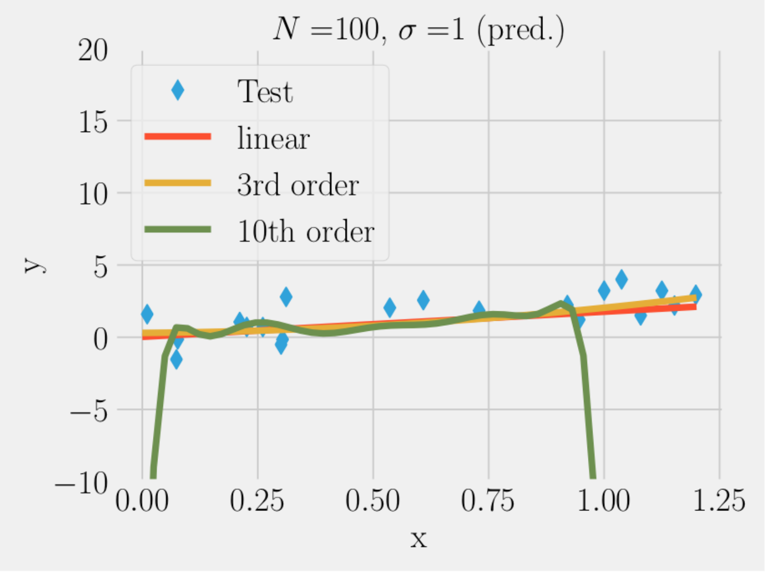
$$= w_0 + w_1 x + w_2 x^2 + \dots + w_{10} x^{10}$$

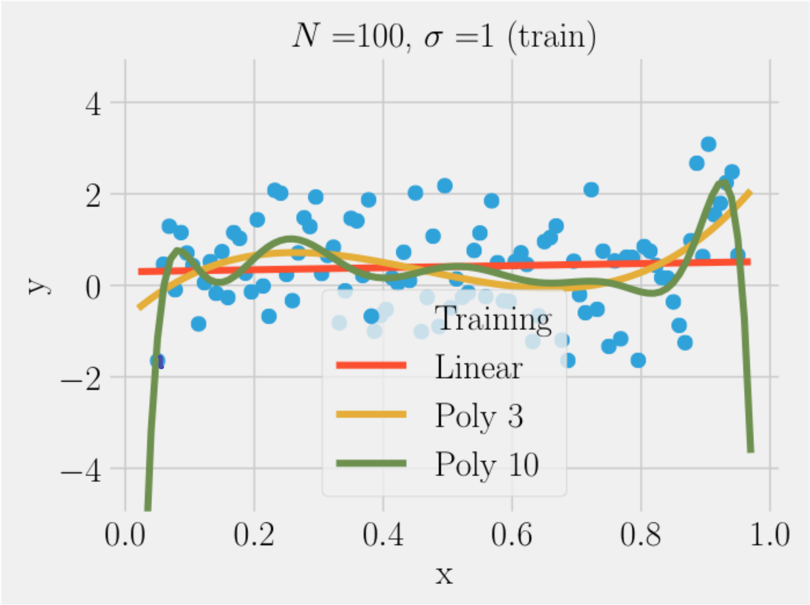




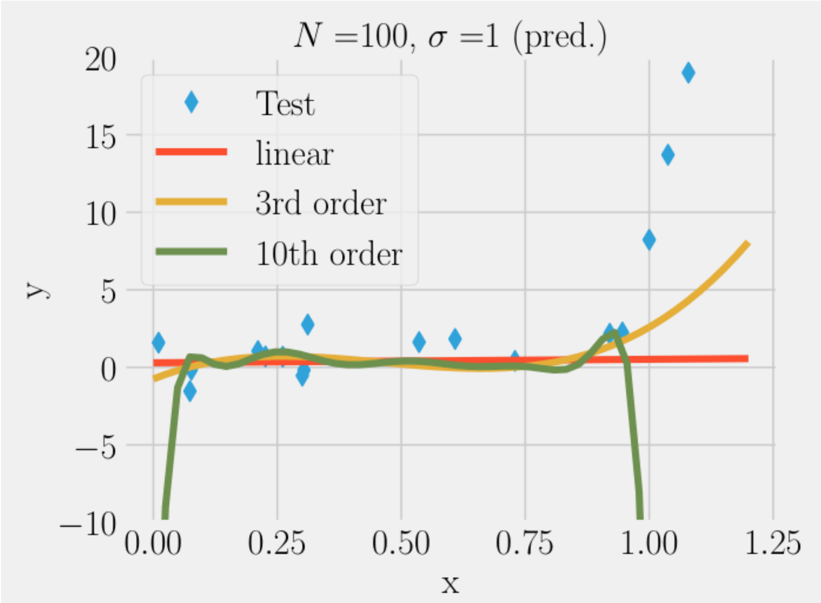
El caso $\sigma = 1$ y $N_{train} = 100$





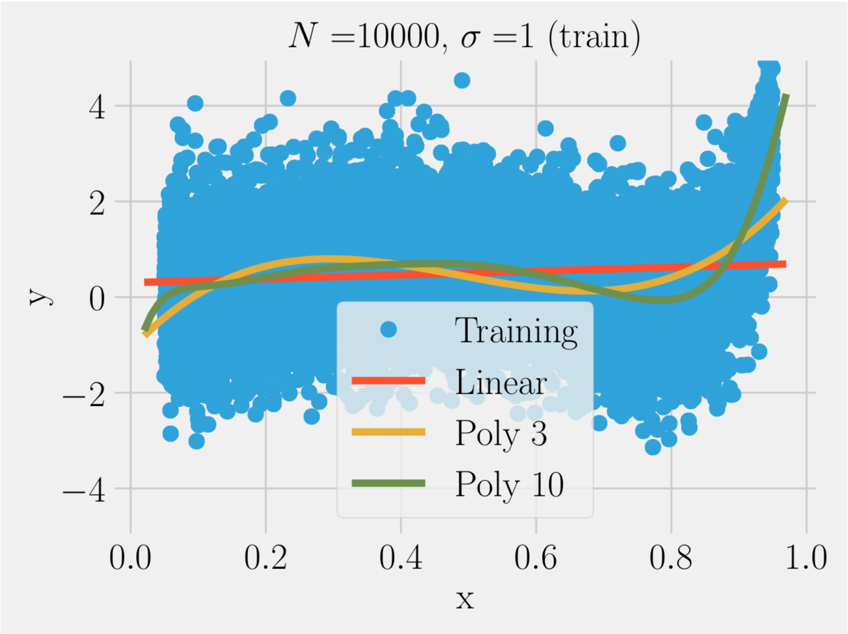


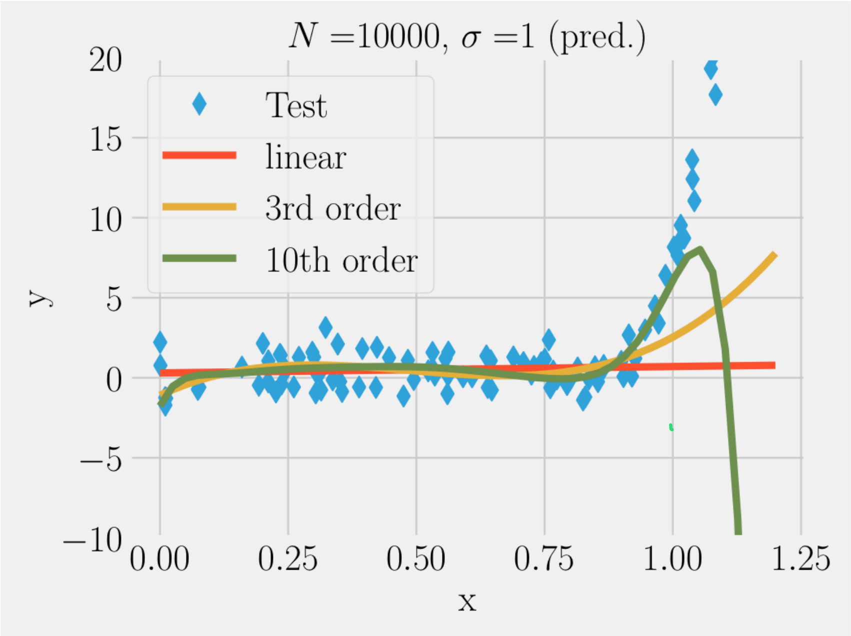
N =100, σ =1 (train)
Training
Linear
Poly 3
Poly 10
y
x



N =100, σ =1 (pred.)
Test
linear
3rd order
10th order
y
x

El caso $\sigma = 1$ y $N_{train} = 1000$





Cuanto más simple es el modelo, o sea, cuanto menos parámetros (sinapsis y umbrales) debemos ajustar (determinar), tendremos más error de aprendizaje E_{in} o "BIAS" (sesgo), pero dependerá menos de la realización particular del conjunto de entrenamiento que nos haya tocado en suerte.

Aun con un número muy grande de puntos de entrenamiento, la capacidad de predecir más allá del conjunto de entrenamiento se puede degradar rápidamente si elegimos un modelo con un número inadecuado de parámetros (porque son muy pocos o porque son demasiados).