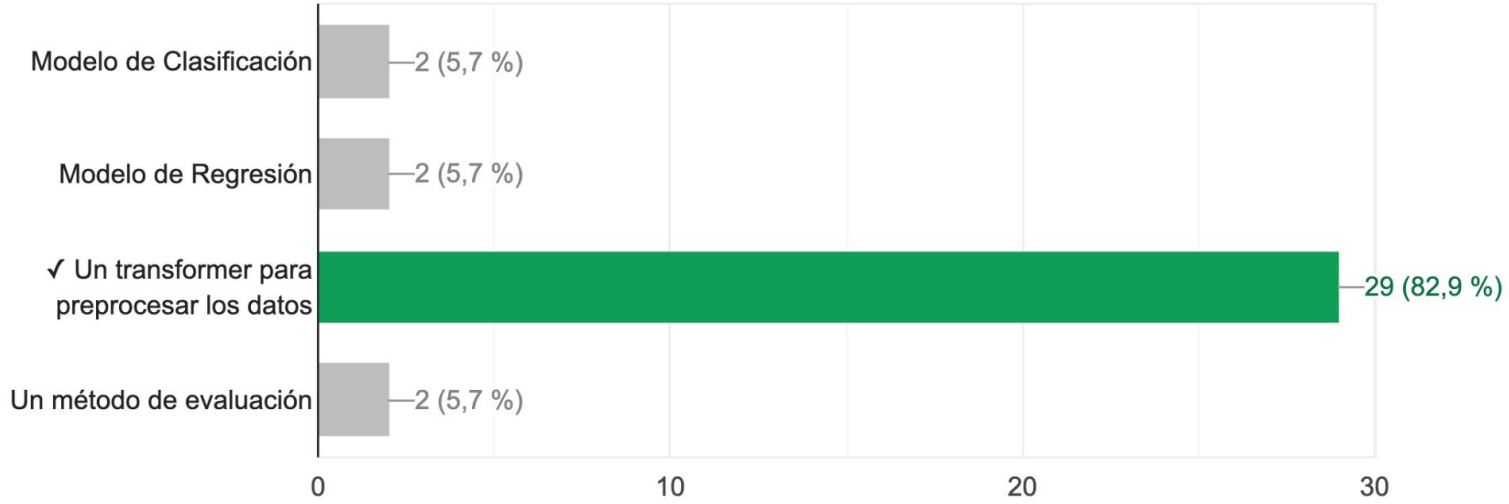
Form Polinomios

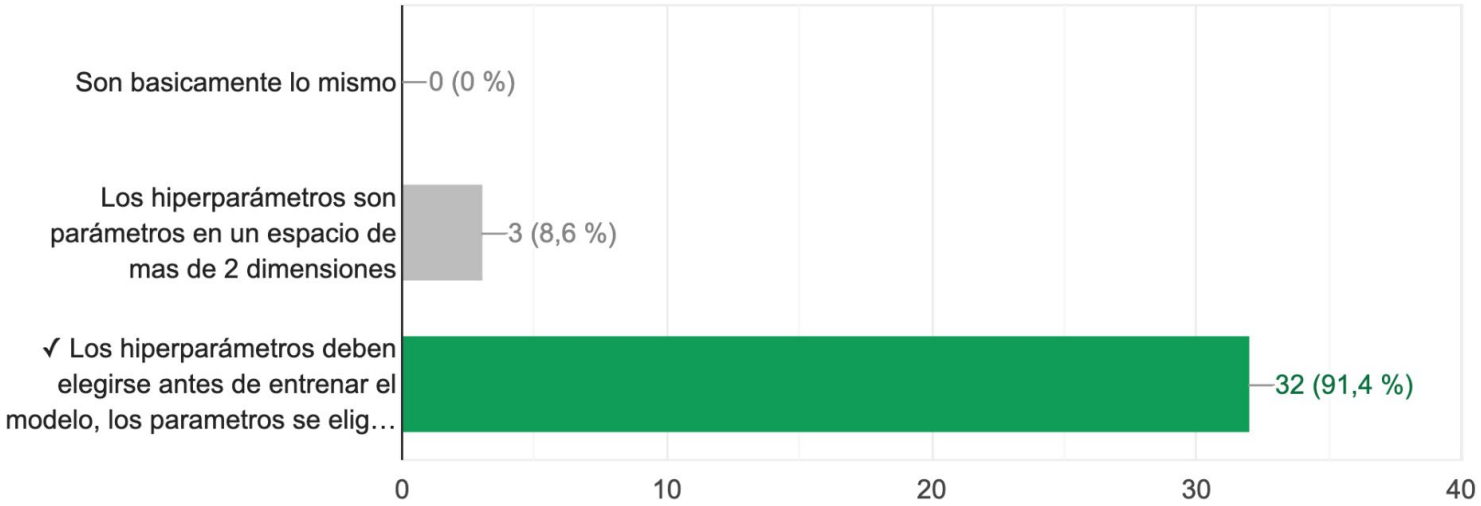
El metodo PolynomialFeatures de sklearn es un

29 de 35 respuestas correctas



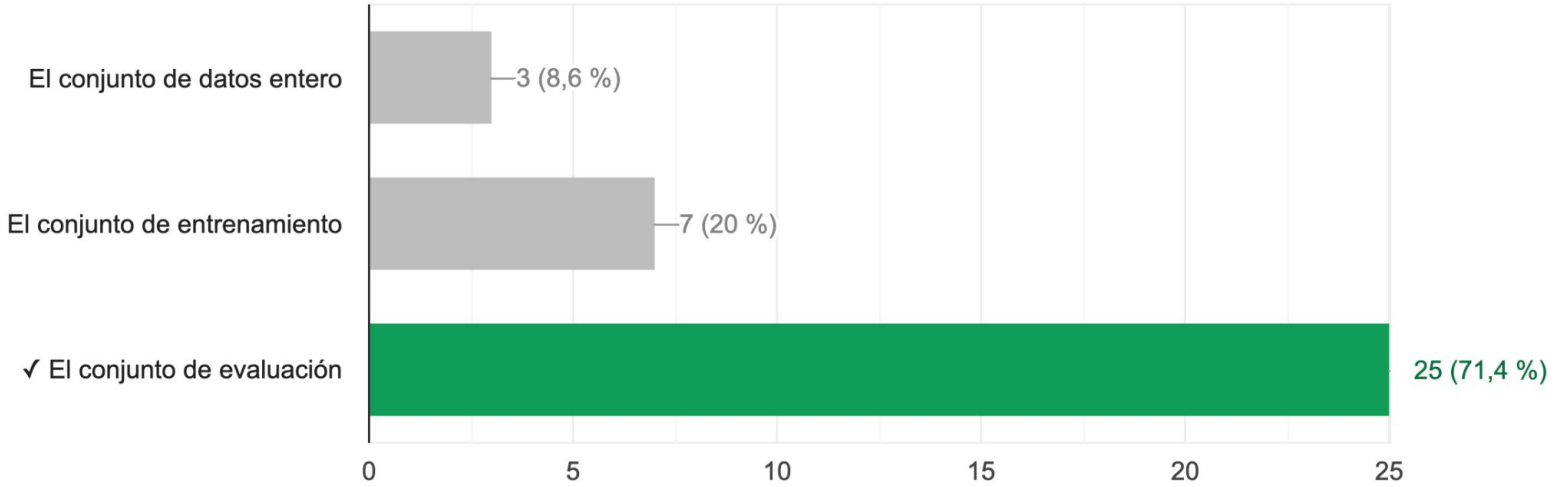
La diferencia entre parámetros e hiperparámetros es

32 de 35 respuestas correctas



Que conjunto de datos utilizamos para medir el poder de generalización?

25 de 35 respuestas correctas



IAA-2023c1
Clase 5: Regularización

UNSAM
UNIVERSIDAD
NACIONAL DE
SAN MARTÍN

Repaso: Regresión Polinómica

Unidimensional

$$x \rightarrow \vec{x} = (x, x^2, \ldots, x^M)$$

$$z = \sum_{i=0}^{M} w_i x^i$$

Hiperparámetro

parámetros

Bidimensional

$$\vec{x} = (x_1, x_2) \rightarrow \vec{x} = (x_1, x_2, x_1^2, x_1 x_2, x_2^2, \ldots, x_1^M, x_1^{M-1} x_2, \ldots, x_1 x_2^{M-1}, x_2^M)$$

F features, al grado M se añaden $\binom{M+F-1}{M} = \frac{(M+F-1)!}{M!(F-1)!} \longrightarrow \frac{M^{F-1}}{(F-1)!}$ términos

Repaso: Parámetros e Hiper-parámetros

Pero ahora tengo que elegir el M **antes** de *entrenar.*

Parámetros:

- Son elegidos por el algoritmo de *optimización* para minimizar la *función de pérdida* medida sobre el *set de datos de* ***entrenamiento.***

Hiperparámetros:

- Son elegidos *antes* de entrenar el modelo. El criterio de elección es para conseguir un modelo que *generalice* mejor.

Figura del Bishop

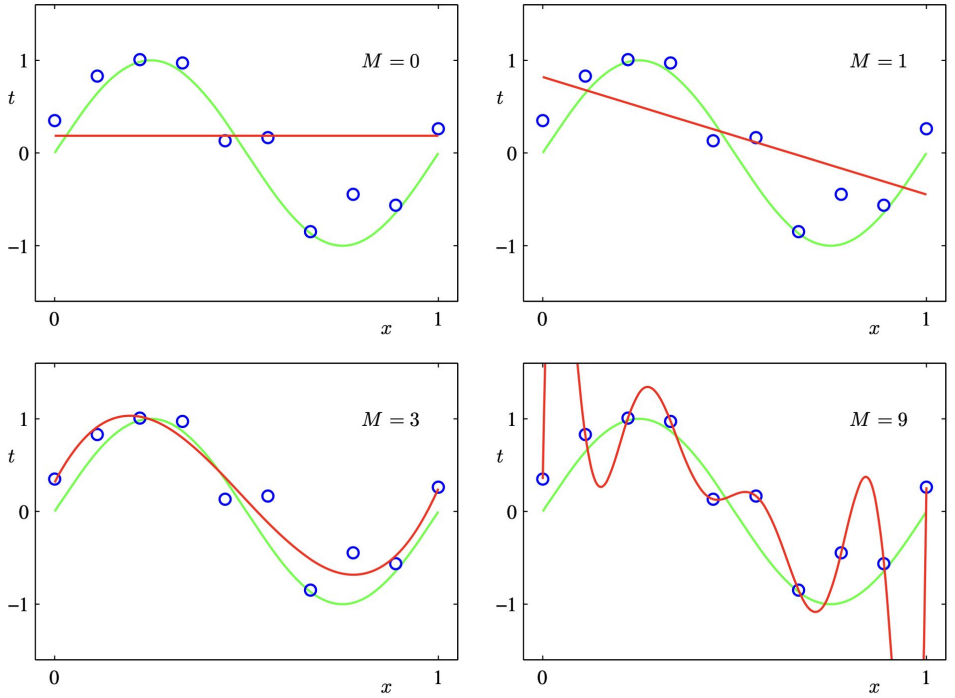


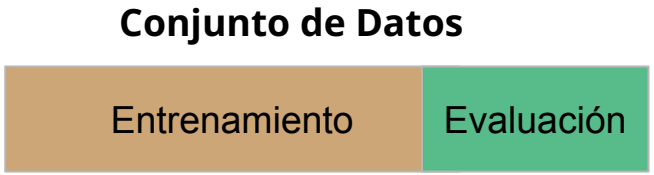
Figure 1.4 Plots of polynomials having various orders M, shown as red curves, fitted to the data set shown in Figure 1.2.

Repaso: Generalización

Poder de generalización:

- La performance esperada sobre un conjunto de datos *nuevo* (i.e. no visto durante el entrenamiento)

- Para aproximarlo, separamos un conjunto de datos y no lo usamos para entrenar: Conjunto de Evaluación.



- Criterio objetivo para elegir el mejor M:

Aquel que **maximice la métrica objetivo** sobre el **conjunto de evaluación.**

Flexibilidad de un Modelo

¿Cómo medimos la flexibilidad de un modelo?

Flexibilidad de un Modelo

¿Cómo medimos la flexibilidad de un modelo?

- Número de parámetros

1 parámetro

2 parámetros

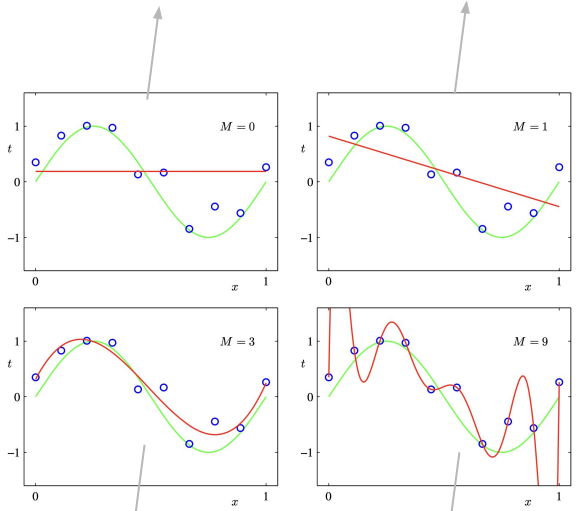


Figure 1.4 Plots of polynomials having various orders M, shown as red curves, fitted to the data set shown in Figure 1.2.

4 parámetros

9 parámetros

Flexibilidad de un Modelo

¿Cómo medimos la flexibilidad de un modelo?

- Número de parámetros

¿Cómo *limitamos* la flexibilidad de un modelo?

1 parámetro

2 parámetros

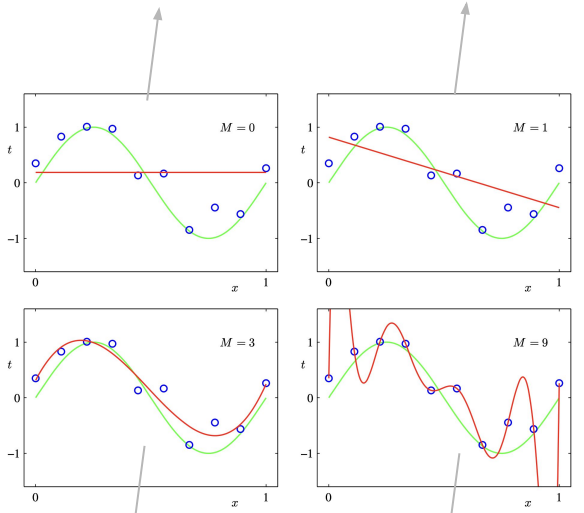


Figure 1.4 Plots of polynomials having various orders M, shown as red curves, fitted to the data set shown in Figure 1.2.

4 parámetros

9 parámetros

Flexibilidad de un Modelo

¿Cómo medimos la flexibilidad de un modelo?

- Número de parámetros

¿Cómo *limitamos* la flexibilidad de un modelo?

- Menor M
  –> Menor número de parámetros

1 parámetro

2 parámetros

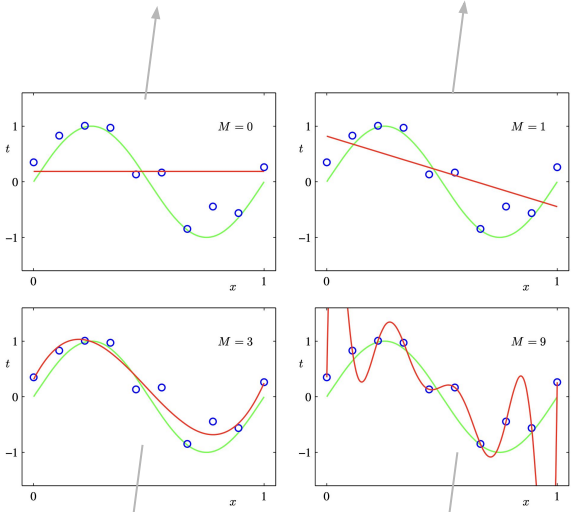


Figure 1.4 Plots of polynomials having various orders M, shown as red curves, fitted to the data set shown in Figure 1.2.

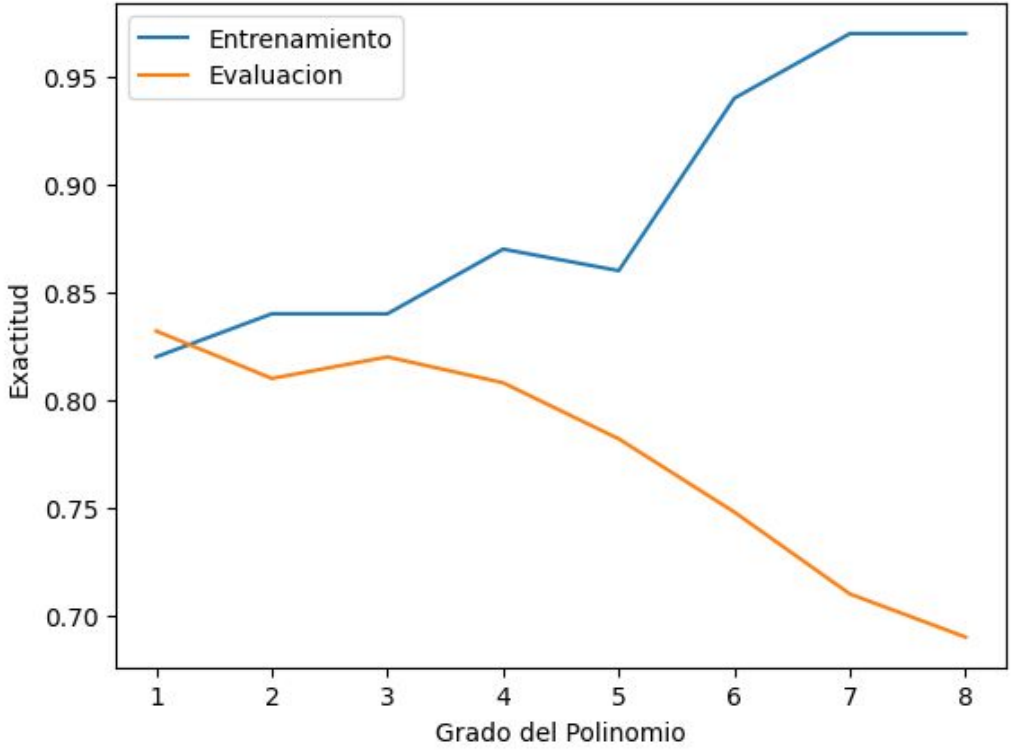
4 parámetros

9 parámetros

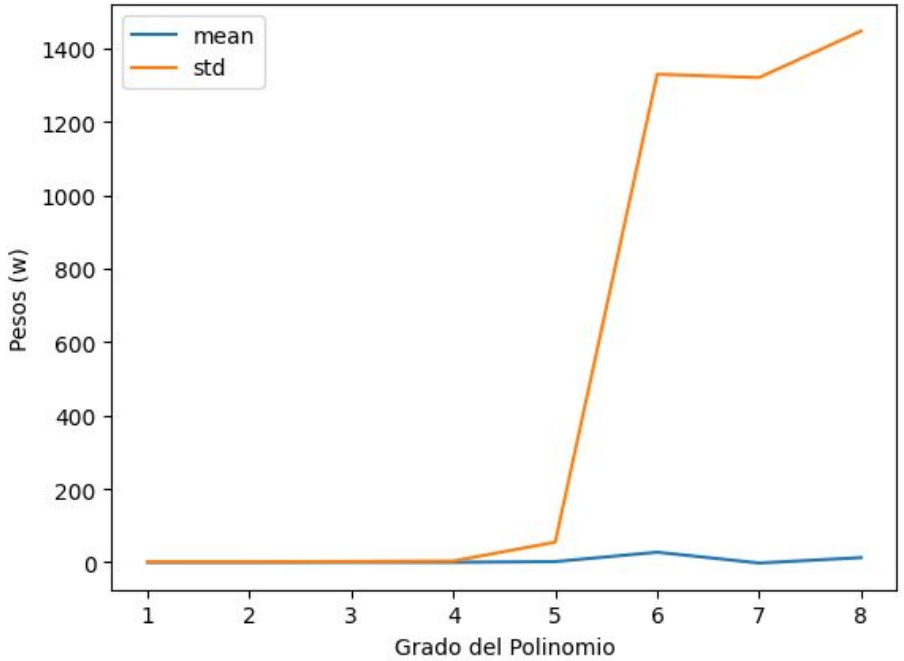
Flexibilidad de un Modelo

```
[13] MAX_DEGREE = 9

     weights_mean = []
     weights_std = []
     accuracies_train = []
     accuracies_test = []
     for d in range(1, MAX_DEGREE):
       print(f"Grado: {d}")
       pipe = make_pipeline(
         PolynomialFeatures(degree=d),
         StandardScaler(),
         LogisticRegression(penalty=None, fit_intercept=False, max_iter=10000)
         )
       pipe.fit(X_train,y_train)
       ws = pipe.steps[-1][1].coef_
       weights_mean.append(ws.mean())
       weights_std.append(ws.std())

       preds = pipe.predict(X_train)
       accuracies_train.append(accuracy_score(preds, y_train))

       preds = pipe.predict(X_test)
       accuracies_test.append(accuracy_score(preds, y_test))
```

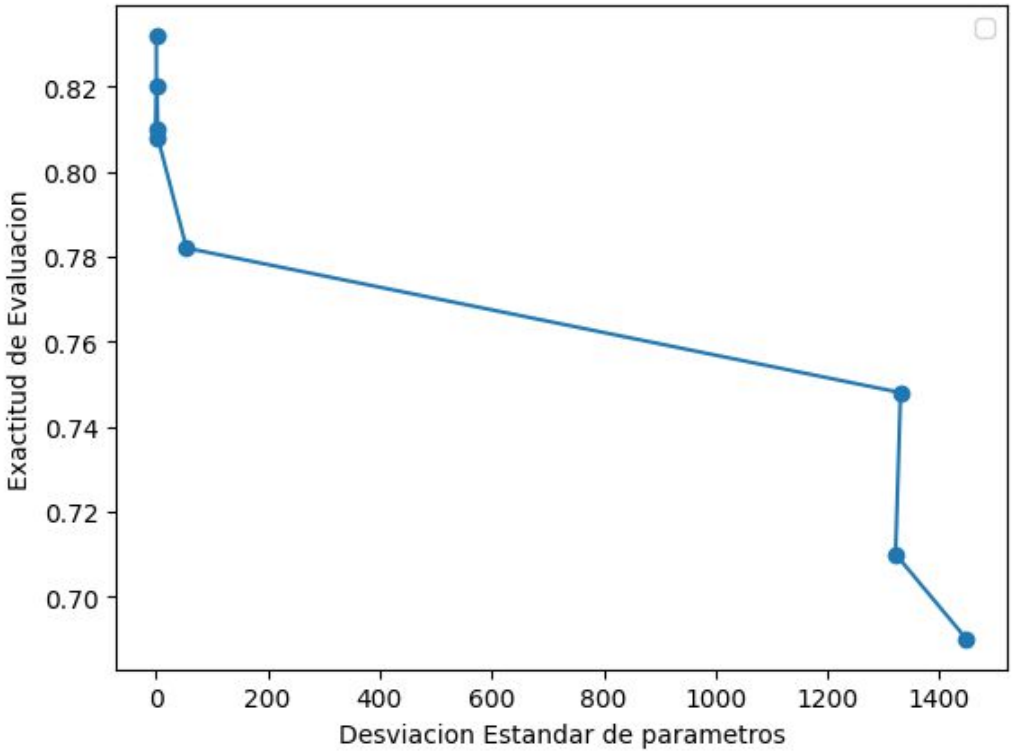
Flexibilidad de un Modelo



Flexibilidad de un Modelo



Flexibilidad de un Modelo



Flexibilidad de un Modelo

¿Cómo medimos la flexibilidad de un modelo?

- Número de parámetros

¿Cómo *limitamos* la flexibilidad de un modelo?

- Menor M
  –> Menor número de parámetros
- ¿Tamaño de los parámetros?

1 parámetro chico

2 parámetros chicos

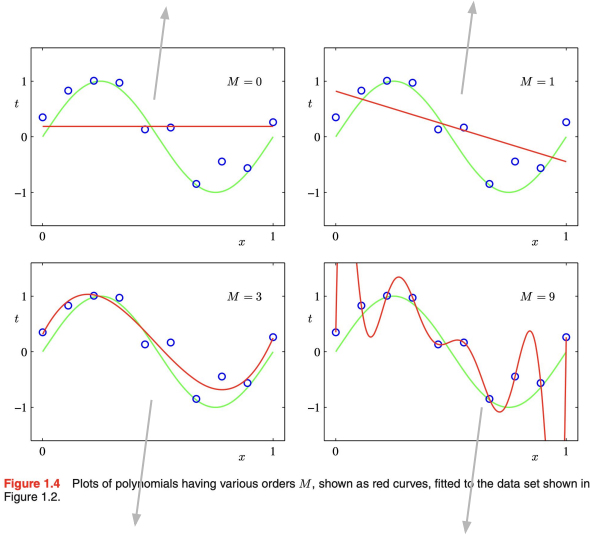


Figure 1.4 Plots of polynomials having various orders M, shown as red curves, fitted to the data set shown in Figure 1.2.

4 parámetros chicos

9 parámetros GRANDES (+ y -)

Regularización

"Mantengamos los pesos pequeños"

¿Cómo se eligen los pesos?

- Función de pérdida

$$L(y,t) = (y-t)^2$$
$$L(y;t) = -\left[t\log(y)(1-t)\log(1-y)\right]$$

¿Cómo puedo hacer para *forzarlos* a ser pequeños?

- Añadir un término que ***penalice*** su tamaño.

$$L(\vec{w};\ \vec{x},\vec{t}) \rightarrow L(\vec{w};\ \vec{x},\vec{t}) + \lambda\, L_{reg}(\vec{w})$$

Término de **regularización**
ó *penalización*

Coeficiente de **regularización**

Regularización

Ridge o L2

- Módulo cuadrado de los coeficientes

$$L_{reg}(\vec{w}) = \|\vec{w}\|_2^2 = \sum_{i=1}^{M} |w_i|^2$$

Regularización

Ridge o L2

- Módulo cuadrado de los coeficientes

$$L_{reg}(\vec{w}) = \|\vec{w}\|_2^2 = \sum_{i=1}^{M} |w_i|^2$$

Lasso o L1

- Módulo de los coeficientes (no continuo)

$$L_{reg}(\vec{w}) = \|\vec{w}\|_1 = \sum_{i=1}^{M} |w_i|$$

Regularización

Comparación Gráfica

$$L_{reg}(\vec{w}) = \|\vec{w}\|_q^q = \sum_{i=1}^{M} w_i^q$$

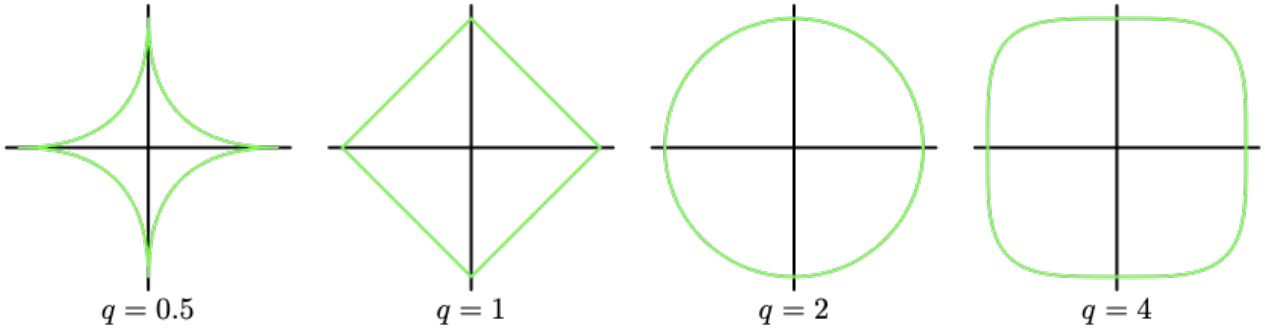
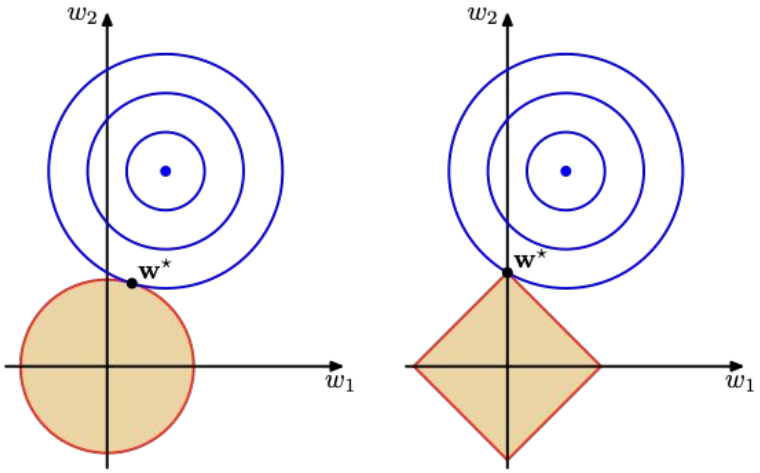


Figure 3.3 Contours of the regularization term in (3.29) for various values of the parameter q.

Regularización

Figure 3.4 Plot of the contours of the unregularized error function (blue) along with the constraint region (3.30) for the quadratic regularizer $q = 2$ on the left and the lasso regularizer $q = 1$ on the right, in which the optimum value for the parameter vector $\mathbf{w}$ is denoted by $\mathbf{w}^\star$. The lasso gives a sparse solution in which $w_1^\star = 0$.



- **L2 (Ridge):**
  Tiende a hacer todos los coeficientes pequeños
- **L1 (Lasso):**
  Tiende a hacer coeficientes 0. Funciona como selector de features

Regularización

Ridge o L2

- Módulo cuadrado de los coeficientes

$$L_{reg}(\vec{w}) = \|\vec{w}\|_2^2 = \sum_{i=1}^{M} |w_i|^2$$

Lasso o L1

- Módulo de los coeficientes (no continuo)

$$L_{reg}(\vec{w}) = \|\vec{w}\|_1 = \sum_{i=1}^{M} |w_i|$$

ElasticNet

- Combinación de L1 y L2

$$L_{reg}(\vec{w}) = \ell \|\vec{w}\|_1 + \frac{1-\ell}{2} \|\vec{w}\|_2^2$$

Regularización: Hiper-parámetros

Una vez escogido el modelo, tengo que escoger:

- El grado del polinomio **M** (parámetro discreto)
- El **coeficiente de regularización** (parámetro continuo)
  (o *los coeficientes* en caso de elasticnet)

La elección de estos se hace midiendo la performance en un conjunto *separado* del de entrenamiento.

Ambos limitan la flexibilidad del modelo, dándonos más libertad de elegir el que generalice mejor

Sobreajuste de Hiper-parámetros

Optimizar algorítmicamente parámetros sobre el conjunto de entrenamiento

“Sobre-Ajusta” el conjunto de entrenamiento

Sobreajuste de Hiper-parámetros

Optimizar algorítmicamente parámetros sobre el conjunto de entrenamiento
“Sobre-Ajusta” el conjunto de entrenamiento

Optimizar algorítmicamente hiper-parámetros sobre el conjunto de evaluación
¿“Sobre-Ajusta” el conjunto de evaluación?

Sobreajuste de Hiper-parámetros

Optimizar algorítmicamente parámetros sobre el conjunto de entrenamiento
“Sobre-Ajusta” el conjunto de entrenamiento

Optimizar algorítmicamente hiper-parámetros sobre el conjunto de evaluación
¿“Sobre-Ajusta” el conjunto de evaluación?