

Ajuste de Curvas e Interpolación

Aplicación de Métodos Numéricos al Ambiente Construido (CV1012)

M.C. Xavier Sánchez Díaz
sax@tec.mx



Outline

- 1 Continuo y diferenciable
- 2 Discreto y aproximable
- 3 Estadística descriptiva básica
- 4 Ajuste de curvas

Recuento

Continuo y diferenciable

Hasta el momento hemos revisado los siguientes temas:

- Métodos numéricos para encontrar raíces de ecuaciones no lineales:
 - ▶ Métodos de intervalos: *bisección, falsa posición*
 - ▶ Métodos abiertos: *punto fijo, Newton-Raphson, secante*
- Matrices:
 - ▶ Operaciones algebraicas con matrices y vectores
 - ▶ Solución de sistemas de ecuaciones lineales usando eliminación

Recuento

Continuo y diferenciable

Hasta el momento hemos revisado los siguientes temas:

- Métodos numéricos para encontrar raíces de ecuaciones no lineales:
 - ▶ Métodos de intervalos: *bisección*, *falsa posición*
 - ▶ Métodos abiertos: *punto fijo*, *Newton-Raphson*, *secante*
- Matrices:
 - ▶ Operaciones algebraicas con matrices y vectores
 - ▶ Solución de sistemas de ecuaciones lineales usando eliminación

Recuento

Continuo y diferenciable

Hasta el momento hemos revisado los siguientes temas:

- Métodos numéricos para encontrar raíces de ecuaciones no lineales:
 - ▶ Métodos de intervalos: *bisección*, *falsa posición*
 - ▶ Métodos abiertos: *punto fijo*, *Newton-Raphson*, *secante*
- Matrices:
 - ▶ Operaciones algebraicas con matrices y vectores
 - ▶ Solución de sistemas de ecuaciones lineales usando eliminación

Recuento

Continuo y diferenciable

Hasta el momento hemos revisado los siguientes temas:

- Métodos numéricos para encontrar raíces de ecuaciones no lineales:
 - ▶ Métodos de intervalos: *bisección*, *falsa posición*
 - ▶ Métodos abiertos: *punto fijo*, *Newton-Raphson*, *secante*
- Matrices:
 - ▶ Operaciones algebraicas con matrices y vectores
 - ▶ Solución de sistemas de ecuaciones lineales usando eliminación

Recuento

Continuo y diferenciable

Hasta el momento hemos revisado los siguientes temas:

- Métodos numéricos para encontrar raíces de ecuaciones no lineales:
 - ▶ Métodos de intervalos: *bisección*, *falsa posición*
 - ▶ Métodos abiertos: *punto fijo*, *Newton-Raphson*, *secante*
- Matrices:
 - ▶ Operaciones algebraicas con matrices y vectores
 - ▶ Solución de sistemas de ecuaciones lineales usando eliminación

Recuento

Continuo y diferenciable

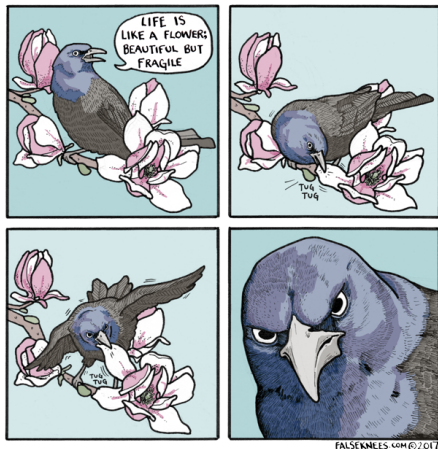
Hasta el momento hemos revisado los siguientes temas:

- Métodos numéricos para encontrar raíces de ecuaciones no lineales:
 - ▶ Métodos de intervalos: *bisección*, *falsa posición*
 - ▶ Métodos abiertos: *punto fijo*, *Newton-Raphson*, *secante*
- Matrices:
 - ▶ Operaciones algebraicas con matrices y vectores
 - ▶ Solución de sistemas de ecuaciones lineales usando eliminación

¿Cómo ha sido el proceso?

Continuo y diferenciable

Hasta ahora, nos dan una ecuación *bonita* y nos dicen qué hacer o qué debemos encontrar en ella. Sin embargo, la vida no es así de fácil. . .



<https://falseknees.com/249.html>

La realidad es distinta

Discreto y aproximable

En ingeniería usualmente tomamos mediciones, y a partir de ello tratamos de hacer generalizaciones.

Para ello, tenemos herramientas como el **ajuste de curvas**, en donde tratamos de encontrar una **función** que *describa* el comportamiento de nuestras **observaciones**.

La realidad es distinta

Discreto y aproximable

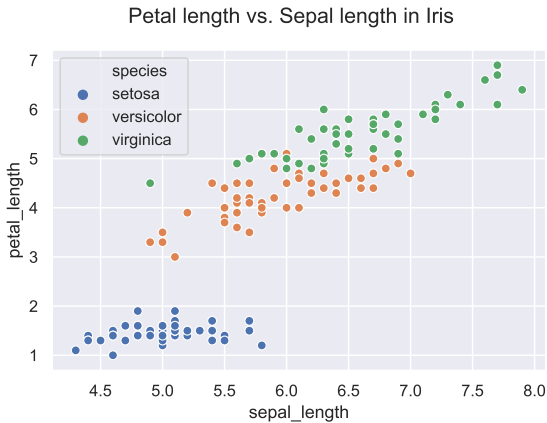
En ingeniería usualmente tomamos mediciones, y a partir de ello tratamos de hacer generalizaciones.

Para ello, tenemos herramientas como el **ajuste de curvas**, en donde tratamos de encontrar una **función** que *describa* el comportamiento de nuestras **observaciones**.

Generalizando

Discreto y aproximable

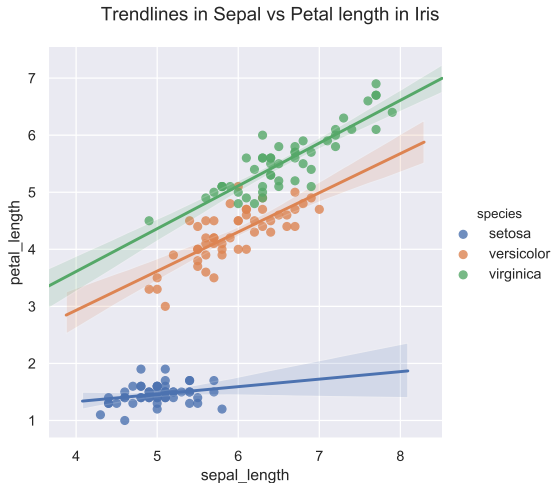
A partir de datos. . .



Generalizando

Discreto y aproximable

...generalizamos.



Nuestras herramientas

Estadística descriptiva básica

Para poder generalizar a partir de una **muestra** de datos, necesitamos saber *más o menos* cómo se comportan.

En este caso, podemos calcular a partir de nuestra **muestra** algunas **estadísticas** que nos ayuden a comprender este comportamiento.

- ¿Cuántos datos tenemos?
- ¿Cuál es el máximo valor? ¿Cuál es el mínimo?
- ¿Qué tan separados están los datos?
- ¿Alrededor de qué valor se concentra la mayoría de los datos?
- ¿Cuál es el dato más común?

Nuestras herramientas

Estadística descriptiva básica

Para poder generalizar a partir de una **muestra** de datos, necesitamos saber *más o menos* cómo se comportan.

En este caso, podemos calcular a partir de nuestra **muestra** algunas **estadísticas** que nos ayuden a comprender este comportamiento.

- ¿Cuántos datos tenemos?
- ¿Cuál es el máximo valor? ¿Cuál es el mínimo?
- ¿Qué tan separados están los datos?
- ¿Alrededor de qué valor se concentra la mayoría de los datos?
- ¿Cuál es el dato más común?

Nuestras herramientas

Estadística descriptiva básica

Para poder generalizar a partir de una **muestra** de datos, necesitamos saber *más o menos* cómo se comportan.

En este caso, podemos calcular a partir de nuestra **muestra** algunas **estadísticas** que nos ayuden a comprender este comportamiento.

- ¿Cuántos datos tenemos?
- ¿Cuál es el máximo valor? ¿Cuál es el mínimo?
- ¿Qué tan separados están los datos?
- ¿Alrededor de qué valor se concentra la mayoría de los datos?
- ¿Cuál es el dato más común?

Nuestras herramientas

Estadística descriptiva básica

Para poder generalizar a partir de una **muestra** de datos, necesitamos saber *más o menos* cómo se comportan.

En este caso, podemos calcular a partir de nuestra **muestra** algunas **estadísticas** que nos ayuden a comprender este comportamiento.

- ¿Cuántos datos tenemos?
- ¿Cuál es el máximo valor? ¿Cuál es el mínimo?
- ¿Qué tan separados están los datos?
- ¿Alrededor de qué valor se concentra la mayoría de los datos?
- ¿Cuál es el dato más común?

Nuestras herramientas

Estadística descriptiva básica

Para poder generalizar a partir de una **muestra** de datos, necesitamos saber *más o menos* cómo se comportan.

En este caso, podemos calcular a partir de nuestra **muestra** algunas **estadísticas** que nos ayuden a comprender este comportamiento.

- ¿Cuántos datos tenemos?
- ¿Cuál es el máximo valor? ¿Cuál es el mínimo?
- ¿Qué tan separados están los datos?
- ¿Alrededor de qué valor se concentra la mayoría de los datos?
- ¿Cuál es el dato más común?

Nuestras herramientas

Estadística descriptiva básica

Para poder generalizar a partir de una **muestra** de datos, necesitamos saber *más o menos* cómo se comportan.

En este caso, podemos calcular a partir de nuestra **muestra** algunas **estadísticas** que nos ayuden a comprender este comportamiento.

- ¿Cuántos datos tenemos?
- ¿Cuál es el máximo valor? ¿Cuál es el mínimo?
- ¿Qué tan separados están los datos?
- ¿Alrededor de qué valor se concentra la mayoría de los datos?
- ¿Cuál es el dato más común?

Nuestras herramientas

Estadística descriptiva básica

Para poder generalizar a partir de una **muestra** de datos, necesitamos saber *más o menos* cómo se comportan.

En este caso, podemos calcular a partir de nuestra **muestra** algunas **estadísticas** que nos ayuden a comprender este comportamiento.

- ¿Cuántos datos tenemos?
- ¿Cuál es el máximo valor? ¿Cuál es el mínimo?
- ¿Qué tan separados están los datos?
- ¿Alrededor de qué valor se concentra la mayoría de los datos?
- ¿Cuál es el dato más común?

Media aritmética

Estadística descriptiva básica

La **media aritmética** (mejor conocida como **promedio**) es una excelente manera de obtener información inmediata sobre el comportamiento *promedio* (duh) de nuestra muestra:

Media

$$\bar{y} = \frac{\sum y_i}{n}$$

donde n es el **tamaño de la muestra** (o sea, cuántos datos tenemos), y y_i es el i -ésimo elemento en nuestra muestra.

Mediana

Estadística descriptiva básica

Otra medida de centralidad importante es la **mediana**, que es el valor de **en medio** de los datos **cuando están ordenados**.

Es importante recalcar que la **media** y la **mediana** **no son lo mismo**, aunque a veces pueden ser idénticas.

- *El sueldo promedio del mexicano hace referencia a la **media** de los sueldos: sumas todos los sueldos y los divides entre los entrevistados para saber lo que esperas que gane un mexicano comúnmente.*
- *El sueldo del mexicano promedio hace referencia a la **mediana** de los sueldos: ordenas todos los sueldos, y tomas el de en medio para saber lo que esperas que gane un mexicano de clase media.*

Mediana

Estadística descriptiva básica

Otra medida de centralidad importante es la **mediana**, que es el valor de **en medio** de los datos **cuando están ordenados**.

Es importante recalcar que **la media y la mediana no son lo mismo**, aunque **a veces pueden ser idénticas**.

- *El sueldo promedio del mexicano hace referencia a la **media** de los sueldos: sumas todos los sueldos y los divides entre los entrevistados para saber lo que esperas que gane un mexicano comúnmente.*
- *El sueldo del mexicano promedio hace referencia a la **mediana** de los sueldos: ordenas todos los sueldos, y tomas el de en medio para saber lo que esperas que gane un mexicano de clase media.*

Mediana

Estadística descriptiva básica

Otra medida de centralidad importante es la **mediana**, que es el valor de **en medio** de los datos **cuando están ordenados**.

Es importante recalcar que **la media y la mediana no son lo mismo**, aunque **a veces pueden ser idénticas**.

- *El sueldo promedio del mexicano* hace referencia a la **media** de los sueldos: sumas todos los sueldos y los divides entre los entrevistados para saber *lo que esperas que gane un mexicano comúnmente*.
- *El sueldo del mexicano promedio* hace referencia a la **mediana** de los sueldos: ordenas todos los sueldos, y tomas el de en medio para saber *lo que esperas que gane un mexicano de clase media*.

Mediana

Estadística descriptiva básica

Otra medida de centralidad importante es la **mediana**, que es el valor de **en medio** de los datos **cuando están ordenados**.

Es importante recalcar que **la media y la mediana no son lo mismo**, aunque **a veces pueden ser idénticas**.

- *El sueldo promedio del mexicano* hace referencia a la **media** de los sueldos: sumas todos los sueldos y los divides entre los entrevistados para saber *lo que esperas que gane un mexicano comúnmente*.
- *El sueldo del mexicano promedio* hace referencia a la **mediana** de los sueldos: ordenas todos los sueldos, y tomas el de en medio para saber *lo que esperas que gane un mexicano de clase media*.

Moda

Estadística descriptiva básica

Otra medida comúnmente empleada es la **moda**, que viene a ser el dato que **más se repite** en la **muestra**.

Siguiendo con nuestro ejemplo anterior, la **moda** vendría a representar *lo que gana la mayoría de los mexicanos*.

Moda

Estadística descriptiva básica

Otra medida comúnmente empleada es la **moda**, que viene a ser el dato que **más se repite** en la **muestra**.

Siguiendo con nuestro ejemplo anterior, la **moda** vendría a representar *lo que gana la mayoría de los mexicanos*.

Midiendo la dispersión

Estadística descriptiva básica

La **desviación estándar** es la medida de *dispersión* más común:

Desviación estándar

$$S_y = \sqrt{\frac{S_t}{n - 1}}$$

donde S_t es la **suma total de los cuadrados de los residuales** entre cada dato y la media:

$$S_t = \sum (y_i - \bar{y})^2$$

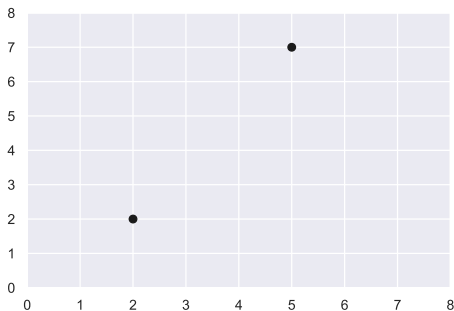
Otra medida de dispersión muy utilizada es la **varianza**, que es igual al **cuadrado de la desviación estándar**.

Regresión lineal

Ajuste de Curvas

Como vimos anteriormente, queremos obtener información a partir de nuestras mediciones, para poder **generar** un **modelo matemático** (o sea una función).

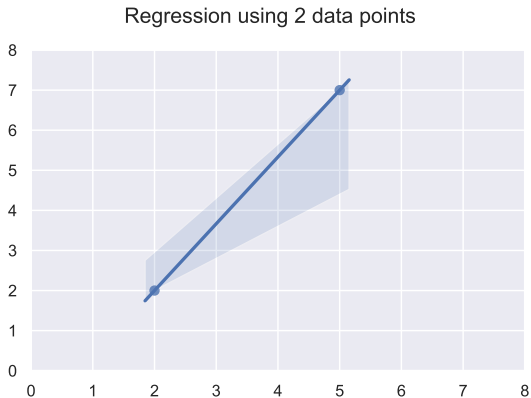
¿Cuál es la manera más sencilla de unir dos puntos?



Regresión lineal

Ajuste de Curvas

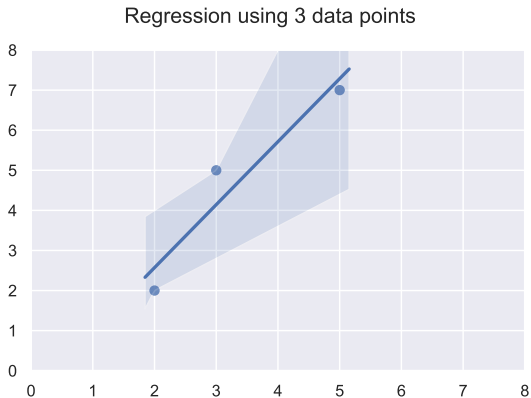
¿Cómo se ve una línea recta si usamos dos puntos?



Regresión lineal

Ajuste de Curvas

¿Y si agregamos otra medición?



Regresión lineal

Ajuste de Curvas

¿Y si agregamos otra más?



Mínimos cuadrados

Ajuste de curvas

¿Cómo sabemos si la línea que hicimos es la *mejor* aproximación que tenemos para la tendencia?

- Podemos contar a cuántos puntos le atinamos y cuántos no
- Podemos revisar por cuánto fallamos en cada punto y sumarlo
- Podemos revisar por cuánto fallamos en cada punto y sacarle valor absoluto
- Podemos revisar por cuánto fallamos en cada punto, elevarlo al cuadrado, y sacar un promedio

Mínimos cuadrados

Ajuste de curvas

¿Cómo sabemos si la línea que hicimos es la *mejor* aproximación que tenemos para la tendencia?

- Podemos contar a cuántos puntos le atinamos y cuántos no
- Podemos revisar por cuánto fallamos en cada punto y sumarlo
- Podemos revisar por cuánto fallamos en cada punto y sacarle valor absoluto
- Podemos revisar por cuánto fallamos en cada punto, elevarlo al cuadrado, y sacar un promedio

Mínimos cuadrados

Ajuste de curvas

¿Cómo sabemos si la línea que hicimos es la *mejor* aproximación que tenemos para la tendencia?

- Podemos contar a cuántos puntos le atinamos y cuántos no
- Podemos revisar por cuánto fallamos en cada punto y sumarlo
- Podemos revisar por cuánto fallamos en cada punto y sacarle valor absoluto
- Podemos revisar por cuánto fallamos en cada punto, elevarlo al cuadrado, y sacar un promedio

Mínimos cuadrados

Ajuste de curvas

¿Cómo sabemos si la línea que hicimos es la *mejor* aproximación que tenemos para la tendencia?

- Podemos contar a cuántos puntos le atinamos y cuántos no
- Podemos revisar por cuánto fallamos en cada punto y sumarlo
- Podemos revisar por cuánto fallamos en cada punto y sacarle valor absoluto
- Podemos revisar por cuánto fallamos en cada punto, elevarlo al cuadrado, y sacar un promedio

Mínimos cuadrados

Ajuste de curvas

¿Cómo sabemos si la línea que hicimos es la *mejor* aproximación que tenemos para la tendencia?

- Podemos contar a cuántos puntos le atinamos y cuántos no
- Podemos revisar por cuánto fallamos en cada punto y sumarlo
- Podemos revisar por cuánto fallamos en cada punto y sacarle valor absoluto
- Podemos revisar por cuánto fallamos en cada punto, elevarlo al cuadrado, y sacar un promedio

Mínimos cuadrados

Ajuste de curvas

¿Cómo sabemos si la línea que hicimos es la *mejor* aproximación que tenemos para la tendencia?

- Podemos contar a cuántos puntos le atinamos y cuántos no
- Podemos revisar por cuánto fallamos en cada punto y sumarlo
- Podemos revisar por cuánto fallamos en cada punto y sacarle valor absoluto
- Podemos revisar por cuánto fallamos en cada punto, elevarlo al cuadrado, y sacar un promedio

Mínimos cuadrados

Ajuste de curvas

Para obtener la ecuación de una recta necesitamos dos elementos:

$$y = a_1x + a_0$$

Pendiente

$$a_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2}$$

Ordenada al origen

$$a_0 = \bar{y} - a_1 \bar{x}$$

Calidad de la regresión

Ajuste de Curvas

Para revisar la **calidad** de nuestra regresión, podemos usar una estadística llamada r^2 .

Calculando r^2

$$r^2 = \frac{S_t - S_r}{S_t}$$

Donde S_t es la suma de todos los cuadrados de los residuales con la media:

$$S_t = \sum_{i=1}^n (y_i - \bar{y})^2$$

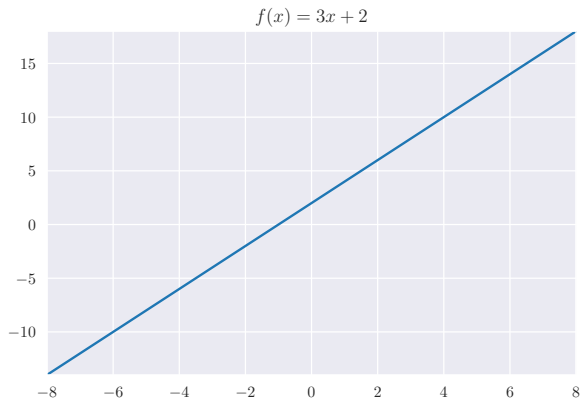
y S_r es la suma de todos los cuadrados de los residuales con nuestra regresión:

$$S_r = \sum_{i=1}^n (y_i - a_0 - a_1 x_i)^2$$

Pendientes

Interpolación

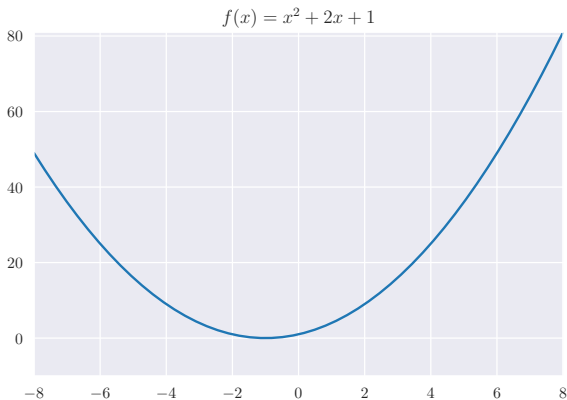
¿Cuántas pendientes tiene la siguiente *función*?



Pendientes

Interpolación

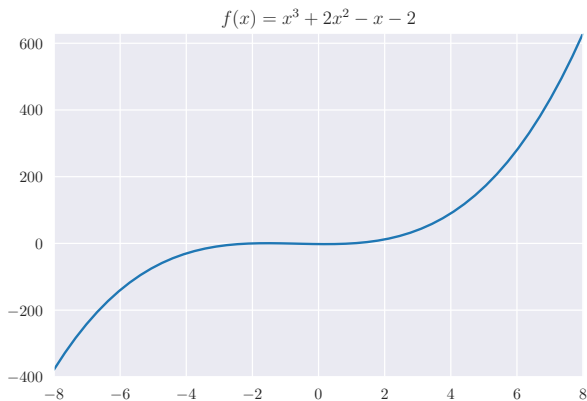
¿Cuántas pendientes tiene la siguiente *función*?



Pendientes

Interpolación

¿Cuántas pendientes tiene la siguiente *función*?



Ajuste polinomial

Interpolación

Como ya vimos, se puede ajustar una función que *aproxime* los puntos. Sin embargo, también se pueden usar métodos para encontrar una función que pase **exactamente** por cada uno de ellos.

- Para unir n puntos, siempre se existirá un polinomio de grado $n - 1$ que describa su comportamiento exacto
- Calcular este polinomio para grados muy altos es computacionalmente costoso
- Para grados muy altos, el error en los extremos tiende a ser mucho más grande

*Story time:
Overfitting*

Ajuste polinomial

Interpolación

Como ya vimos, se puede ajustar una función que *aproxime* los puntos. Sin embargo, también se pueden usar métodos para encontrar una función que pase **exactamente** por cada uno de ellos.

- Para unir n puntos, siempre se existirá un **polinomio** de **grado** $n - 1$ que describa su comportamiento exacto
- Calcular este polinomio para grados muy altos es computacionalmente **costoso**
- Para grados muy altos, el **error** en los extremos tiende a ser mucho más grande

*Story time:
Overfitting*

Ajuste polinomial

Interpolación

Como ya vimos, se puede ajustar una función que *aproxime* los puntos. Sin embargo, también se pueden usar métodos para encontrar una función que pase **exactamente** por cada uno de ellos.

- Para unir n puntos, siempre se existirá un **polinomio** de **grado** $n - 1$ que describa su comportamiento exacto
- Calcular este polinomio para grados muy altos es computacionalmente **costoso**
- Para grados muy altos, el **error** en los extremos tiende a ser mucho más grande

*Story time:
Overfitting*

Ajuste polinomial

Interpolación

Como ya vimos, se puede ajustar una función que *aproxime* los puntos. Sin embargo, también se pueden usar métodos para encontrar una función que pase **exactamente** por cada uno de ellos.

- Para unir n puntos, siempre se existirá un **polinomio** de **grado** $n - 1$ que describa su comportamiento exacto
- Calcular este polinomio para grados muy altos es computacionalmente **costoso**
- Para grados muy altos, el **error** en los extremos tiende a ser mucho más grande

*Story time:
Overfitting*

Ajuste polinomial

Interpolación

Como ya vimos, se puede ajustar una función que *aproxime* los puntos. Sin embargo, también se pueden usar métodos para encontrar una función que pase **exactamente** por cada uno de ellos.

- Para unir n puntos, siempre se existirá un **polinomio** de **grado** $n - 1$ que describa su comportamiento exacto
- Calcular este polinomio para grados muy altos es computacionalmente **costoso**
- Para grados muy altos, el **error** en los extremos tiende a ser mucho más grande

*Story time:
Overfitting*

Ajuste polinomial

Interpolación

Como ya vimos, se puede ajustar una función que *aproxime* los puntos. Sin embargo, también se pueden usar métodos para encontrar una función que pase **exactamente** por cada uno de ellos.

- Para unir n puntos, siempre se existirá un **polinomio** de **grado** $n - 1$ que describa su comportamiento exacto
- Calcular este polinomio para grados muy altos es computacionalmente **costoso**
- Para grados muy altos, el **error** en los extremos tiende a ser mucho más grande

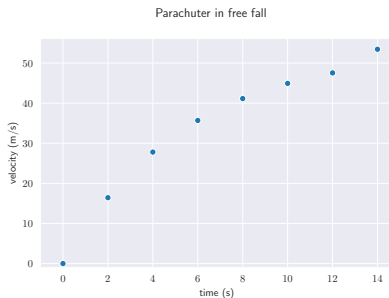
Story time:
Overfitting

El concepto de interpolación

Interpolación

Interpolar hace referencia a obtener puntos **intermedios** a partir de mediciones dadas:

t (s)	v (m/s)
0	0.00
2	16.42
4	27.8
6	35.68
8	41.14
10	44.92
12	47.54
∞	53.44

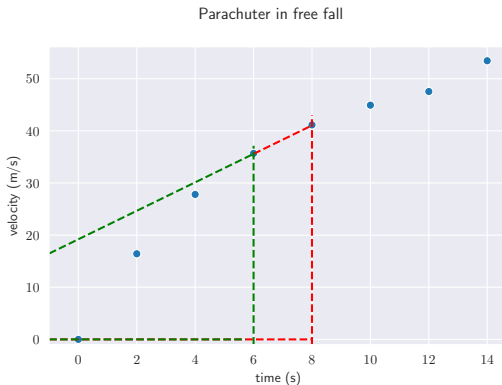


¿Cuál será la velocidad a los 5 segundos?

Interpolación lineal

Interpolación

La **interpolación lineal** (o clásica) consiste en unir entre puntos con líneas rectas.



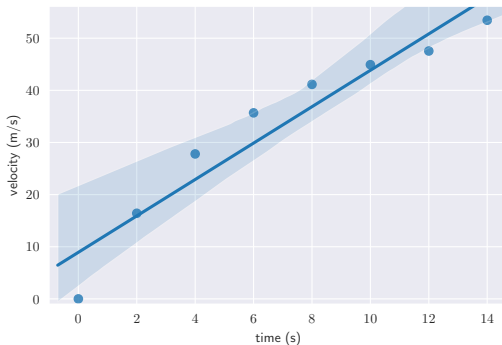
Interpolación con mínimos cuadrados

Interpolación

También podemos **interpolarse por mínimos cuadrados** linealmente, evaluando $f(5)$:

$$f(x) = 3.49107143x + 8.93$$

Least squares linear regression for the Parachuter problem

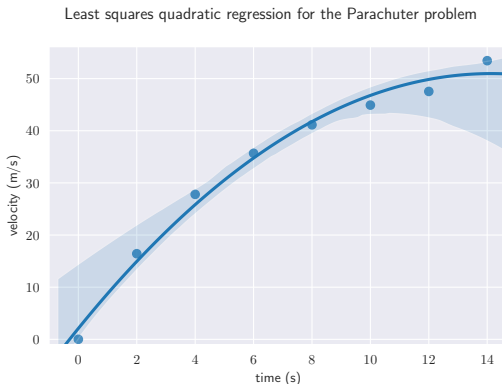


Interpolación con mínimos cuadrados

Interpolación

O con un polinomio de segundo grado:

$$f(x) = -0.244375x^2 + 6.91232143x + 2.0875$$



Interpolación con mínimos cuadrados

Interpolación

O con un polinomio de grado 7 (que es terrible):

$$\begin{aligned}f(x) = & 0.000006.23139881x^7 - 0.000261718750x^6 + 0.00439322917x^5 \\ & - 0.00384635417x^4 + 0.0212750000x^3 - 1.23945833x^2 \\ & + 10.0833095x - 0.00000000000003.01457752\end{aligned}$$

Least squares 7th order regression for the Parachuter problem

