FLACSO ARGENTINA

Facultad Latinoamericana de Ciencias Sociales. Sede Argentina.

Área Comunicación y Cultura.

CURSO DE POSGRADO

BIG DATA E INTELIGENCIA TERRITORIAL

Clase 1. Modelos Lineales

Regresión lineal

Regresión lineal

Estadística

- Método de cuadrados mínimos (solución exacta) - OLS
- Chequeo de los supuestos (assumptions).
- **Importante seguir los pasos si voy a realizar una inferencia utilizando regresión lineal.**

Recordemos los supuestos

- ***Independencia***
- ***Linealidad***
- ***Homocedasticidad***
- ***Normalidad***

(revisar los apuntes de la materia Estadística Computacional)

Mas información en https://openintrostat.github.io/ims-tutorials/

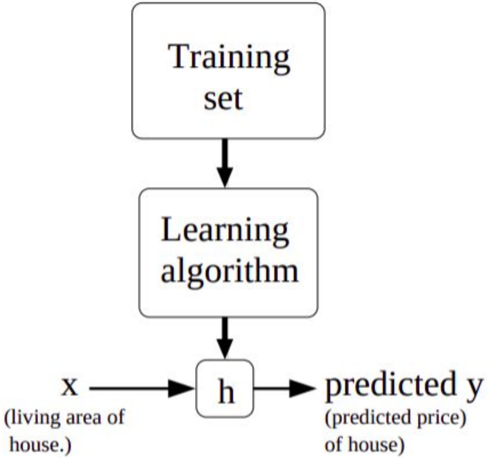
Regresión lineal

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2$$

$$h(x) = \sum_{i=0}^{n} \theta_i x_i = \theta^T x,$$

$$J(\theta) = \frac{1}{2} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2.$$

Función de costo



El parámetro theta se actualiza según

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta).$$

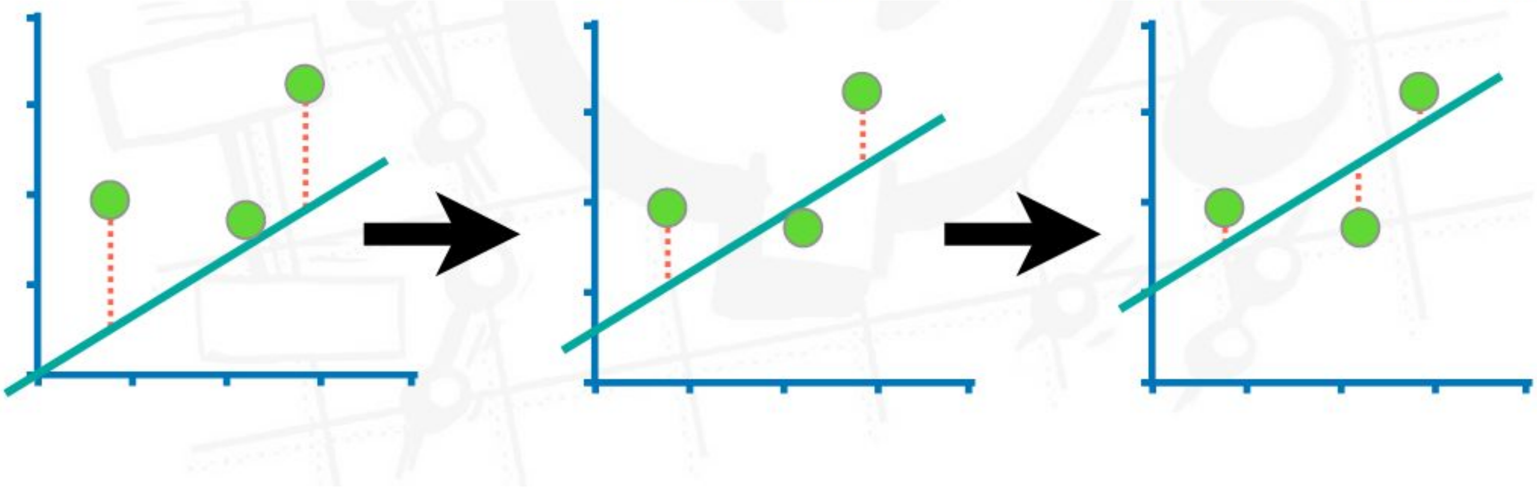
tasa de aprendizaje

Fuente: https://see.stanford.edu/materials/aimlcs229/cs229-notes1.pdf

Regresión lineal

- La gran diferencia está en que **<u>los parámetros los vamos a "aprender"</u>** mediante el descenso por el gradiente.
- Esta solución no es exacta.
- El descenso por el gradiente es la base para otros modelos más complejos.
- El objetivo: minimizar la función de costo J

Descenso por el gradiente



Arrancamos con un valor aleatorio

Luego, mejoramos ese valor, una vez en cada paso

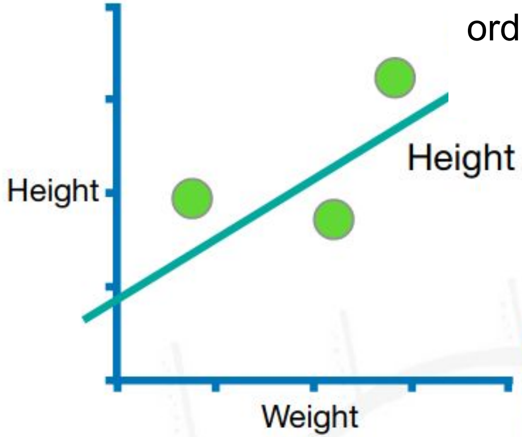
Hasta que alcanzamos una solución óptima

***Fuente**: Stochastic gradient descent - study guide - Josh Hammer*

Descenso por el gradiente

Lo esencial es tener una *idea intuitiva* del descenso por el gradiente

El descenso por el gradiente se usa para optimizar parámetros. En este caso, queremos optimizar el valor de la ordenada y pendiente.



$$\text{Height} = \textbf{intercept} + 0.64 \times \text{Weight}$$

Por ahora vamos a tomar un valor inicial de 0.64 para la pendiente

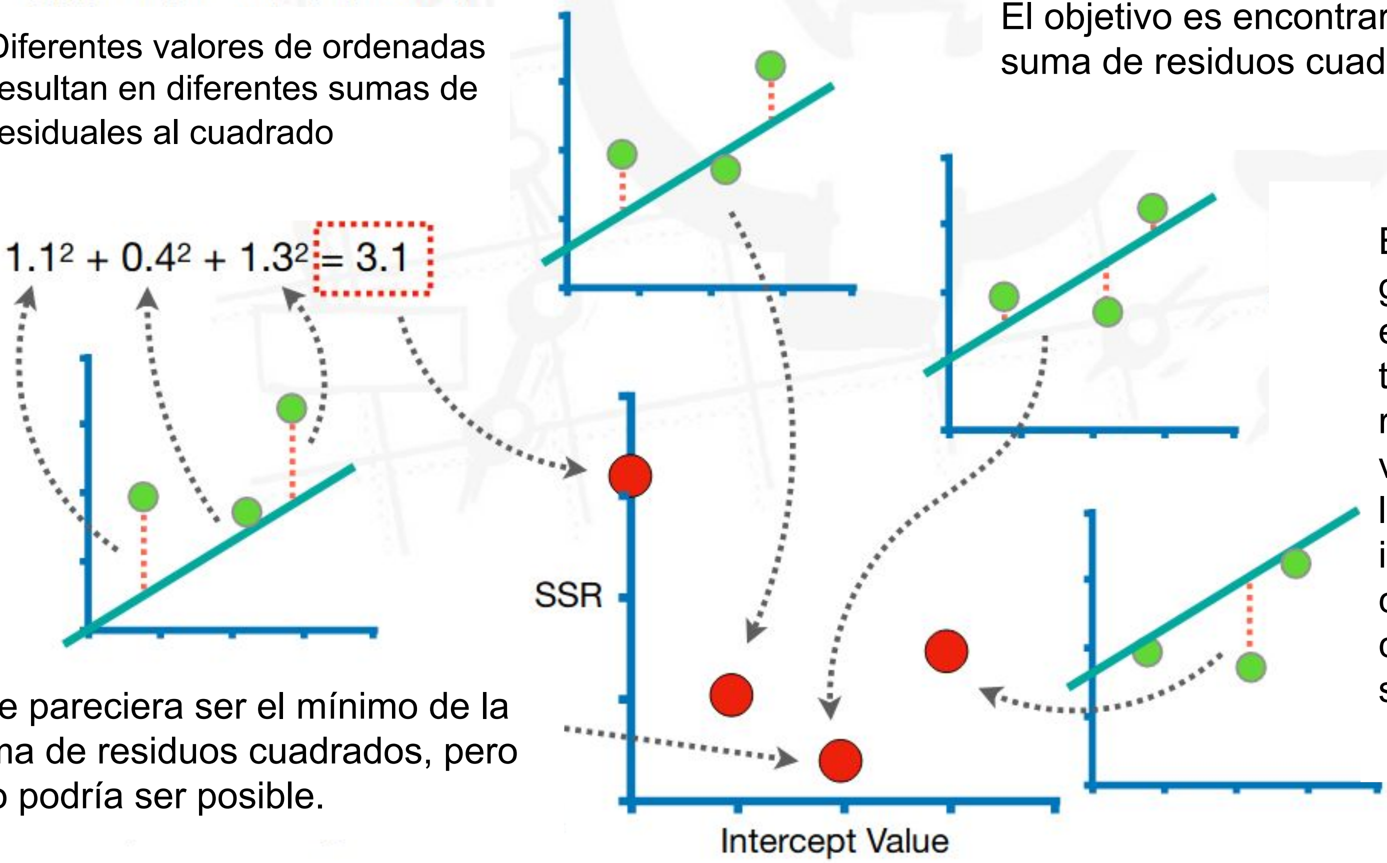
Utilizamos la función de pérdida para evaluar los posibles candidatos a "mejor" parámetro.

Fuente: *Stochastic gradient descent - study guide - Josh Hammer*

Descenso por el gradiente

Diferentes valores de ordenadas resultan en diferentes sumas de residuales al cuadrado

El objetivo es encontrar el **mínimo** de la suma de residuos cuadrados



El descenso por el gradiente resuelve este problema testeando relativamente pocos valores por fuera de la solución óptima e incrementando la cantidad de valores cercanos a la solución óptima.

Este pareciera ser el mínimo de la suma de residuos cuadrados, pero otro podría ser posible.

Descenso por el gradiente

Los <u>Residuales</u> son la diferencia entre los valores observados y los predichos

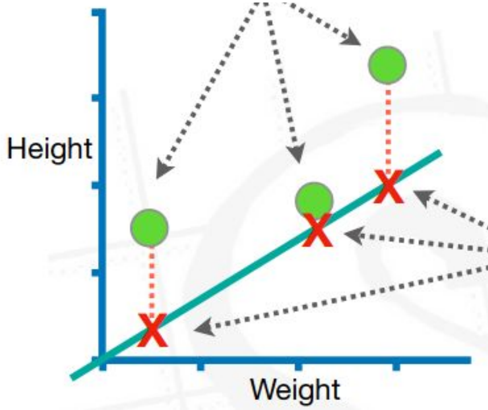
Reemplazo Predicted Height

Residual = (Observed Height - Predicted Height) = (Observed - (**intercept** + 0.64 × Weight)

Las **<u>alturas observadas</u>** son los valores que medimos

Las **<u>alturas predichas</u>** vienen de la ecuación de la recta

Podemos reemplazar en la ecuación



Predicted Height = **intercept** + 0.64 × Weight

La función de pérdida es el SSR (residual sum of squares)

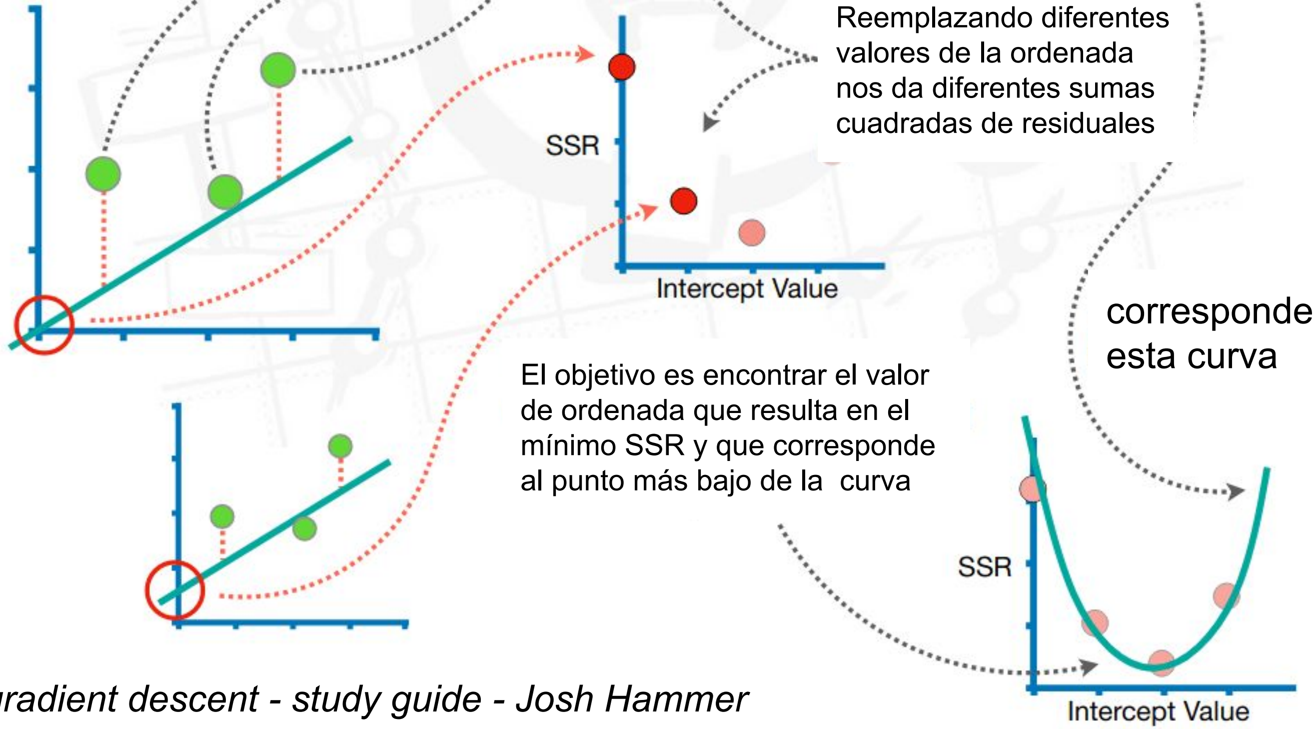
Fuente: *Stochastic gradient descent - study guide - Josh Hammer*

Función de pérdida (loss function): suma de residuales cuadrados (SSR)

Sum of Squared Residuals (SSR) = (Height - (**intercept** + 0.64 x Weight))^2
+ (Height - (**intercept** + 0.64 x Weight))^2
+ (Height - (**intercept** + 0.64 x Weight))^2

Ecuación de SSR

Hay un término en la suma de cada punto observado



Reemplazando diferentes valores de la ordenada nos da diferentes sumas cuadradas de residuales

corresponde a esta curva

El objetivo es encontrar el valor de ordenada que resulta en el mínimo SSR y que corresponde al punto más bajo de la curva

Fuente: *Stochastic gradient descent - study guide - Josh Hammer*

SSR y RMSE

Residual sum of squares (RSS)

$$RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Mean squared error (MSE)

$$MSE = \frac{1}{n} \Sigma \underbrace{\left(y - \widehat{y} \right)^2}_{\text{The square of the difference between actual and predicted}}$$

Root mean square error (RMSE)

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}}$$

Resumiendo

- OLS me da la solución exacta a un problema de regresión. Pero no siempre podemos calcular esa solución exacta debido al costo computacional que demanda.
- En esos casos, tenemos el descenso por el gradiente que es una solución aproximada para encontrar los mejores parámetros.
- El descenso por el gradiente "aprende" los parámetros, en este caso, de la recta de regresión lineal. Nosotros sólo podemos controlar los hiperparámetros, es decir, el valor inicial, la cantidad de pasos, la tasa de aprendizaje, la velocidad al que va la tasa de aprendizaje, etc.
- El descenso por el gradiente es importante porque es la forma de aprender los parámetros en modelos más complejos como son las redes neuronales (deep learning).

Bibliografía

- Apuntes de Estadística Computacional (este curso) para regresión lineal.
- Tutoriales OpenIntro https://openintrostat.github.io/ims-tutorials/
- https://see.stanford.edu/materials/aimlcs229/cs229-notes1.pdf
- (video) Descenso por el gradiente https://www.youtube.com/watch?v=vMh0zPT0tLI&t=488s