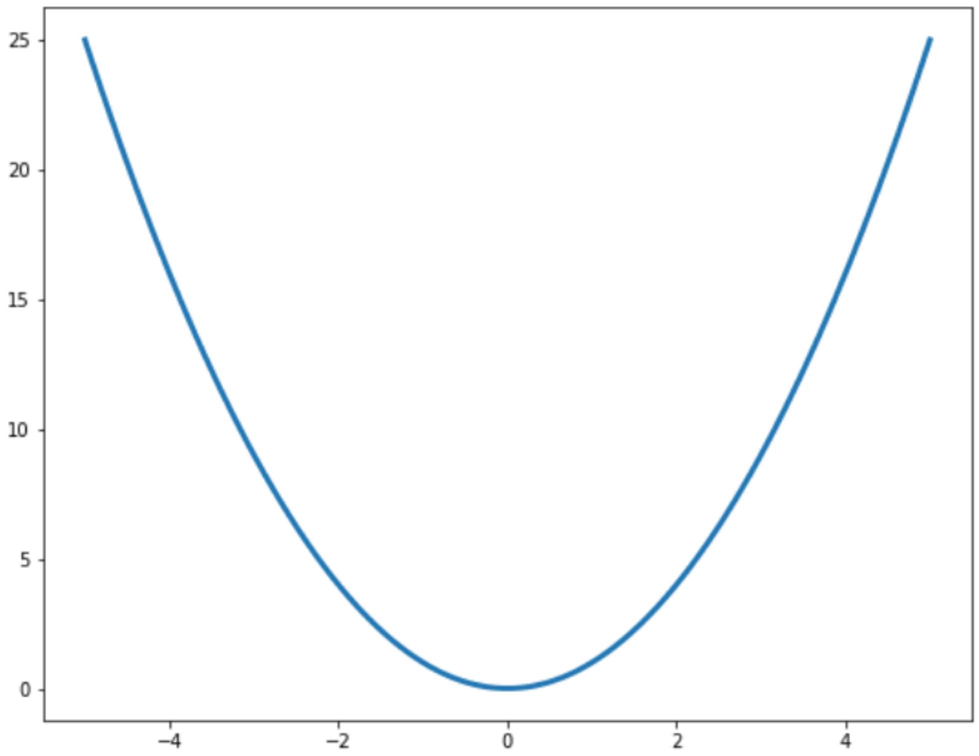
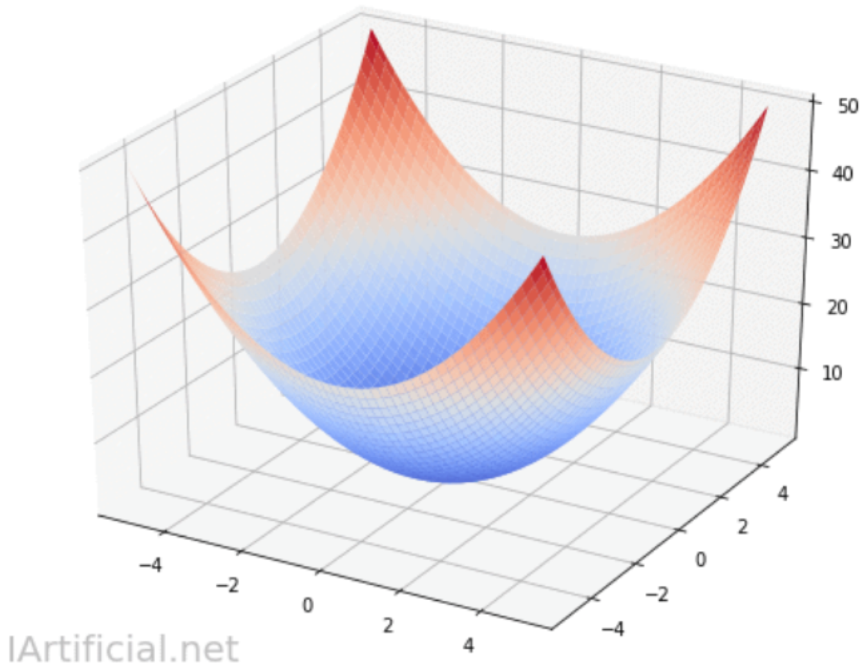
Ejemplo de descenso por gradiente

Formas del Error

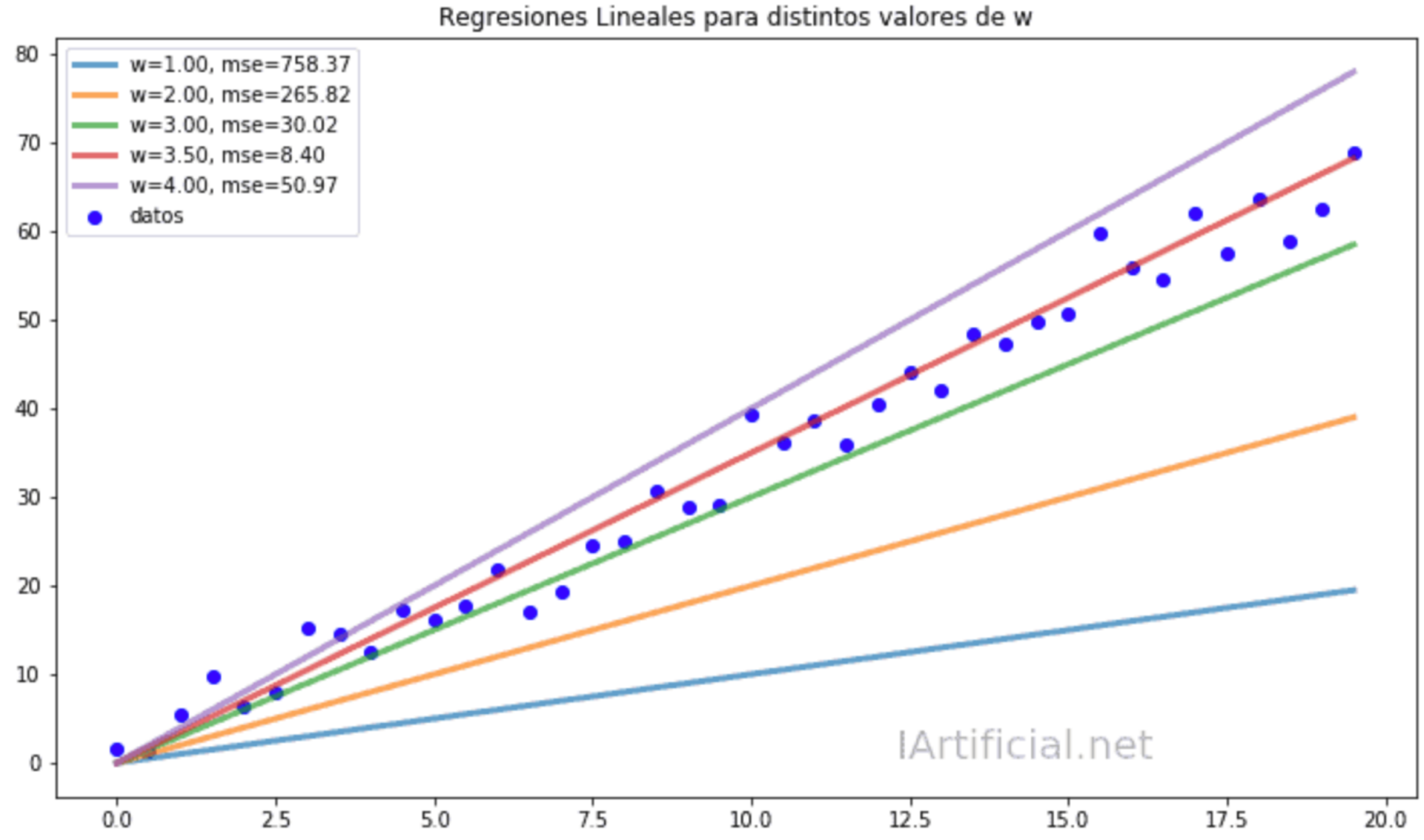




Regresión lineal variable de orden 1

$$y=f(x)=wx+b$$

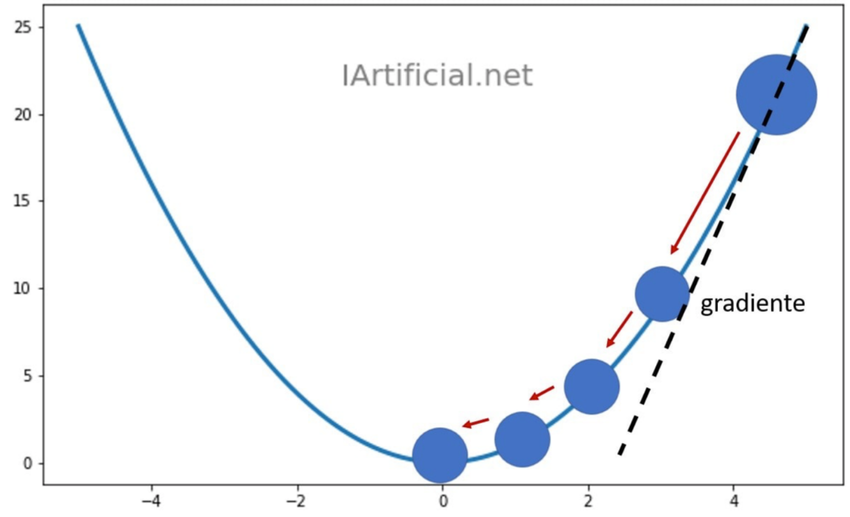
Variaciones sobre w



w = 1.00, mse = 758.37
w = 2.00, mse = 265.82
w = 3.00, mse = 30.02
w = 4.00, mse = 50.97
w = 3.50, mse = 8.40

Navegando por el gradiente

- Lo que vamos buscando es encontrar el conjunto de parámetros que minimizan la función de coste (error cuadrático medio).
- Como vemos en el gráfico, el gradiente representa la pendiente en el punto que nos encontremos de la función de coste. Cuanto más «empinada» sea la pendiente, más queremos ir en sentido contrario (para abajo), para minimizar el error.



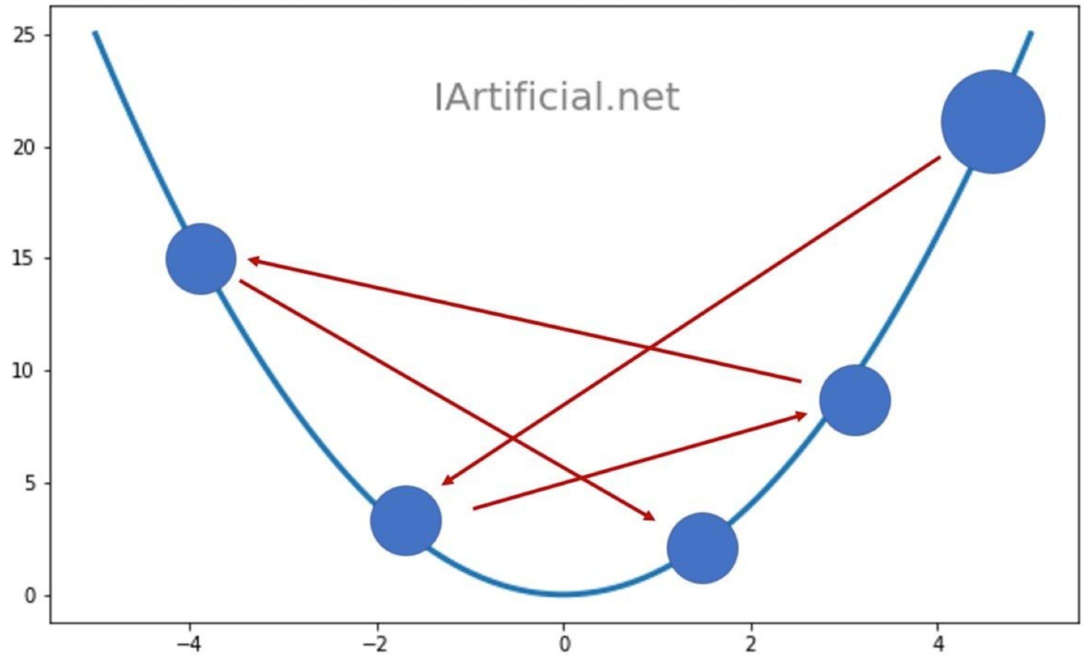
Actualización de los coeficientes

- Usamos '-' (resta), para ir en la dirección opuesta al gradiente
- La actualización de *W* es proporcional al gradiente
  (a mayor pendiente [más empinado], mayor es el cambio)
- α, también llamado 'ratio de aprendizaje', controla el tamaño de la actualización

$$W \leftarrow W - \alpha \frac{\partial MSE}{\partial W}$$

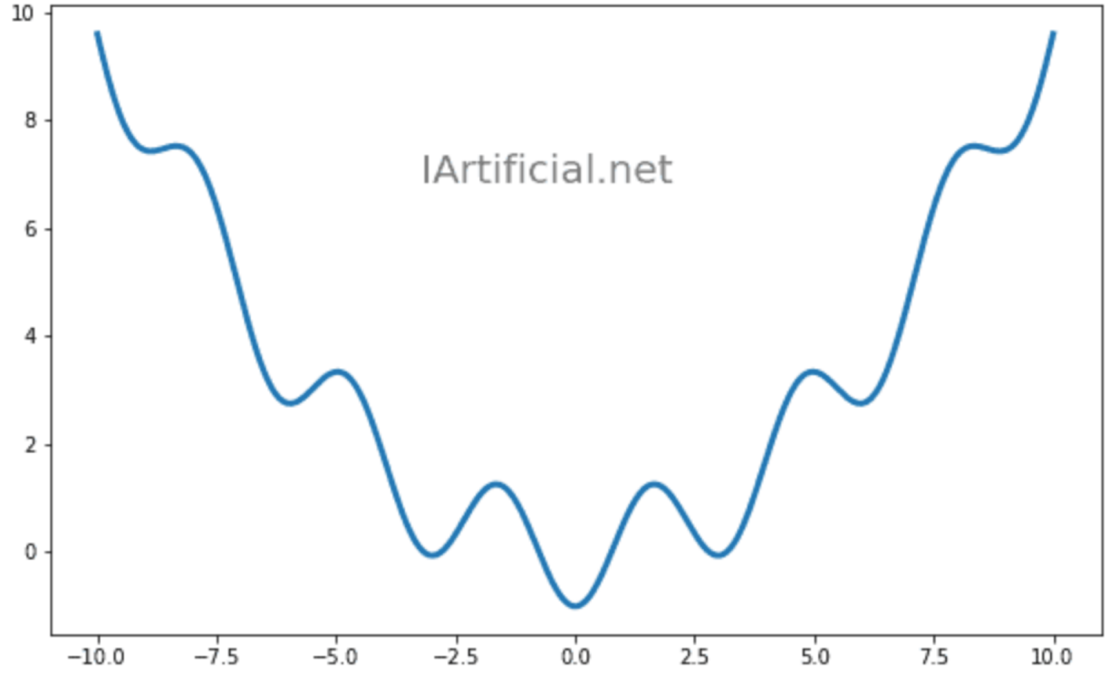
La importancia de elegir un buen ratio de aprendizaje

- si el ratio de aprendizaje es demasiado grande, los cambios en *W* serán también muy grandes y será difícil encontrar los coeficientes que minimicen la función de coste.



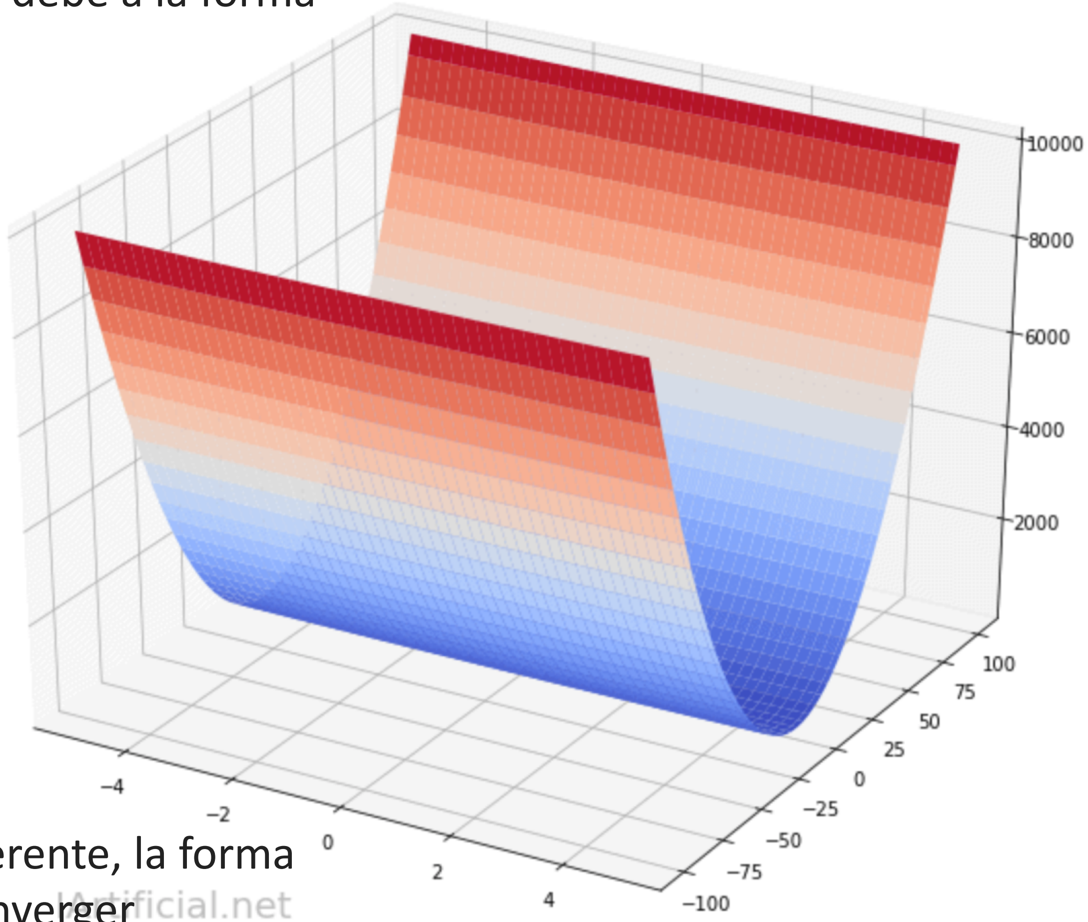
Demasiado pequeño

- Por otra parte, si el ratio de aprendizaje es demasiado pequeño, el gradiente descendiente tardará mucho en encontrar la solución adecuada
- ¿Y si el gradiente descendiente se queda atrapado en un mínimo local?



La importancia de escalar los datos al usar el Gradiente Descendiente

El gradiente descendiente funciona mucho mejor cuando los datos tienen una escala similar. Esto se debe a la forma geométrica de la función de coste



si distintas dimensiones tienen una escala diferente, la forma geométrica será muy diferente, tardará en converger

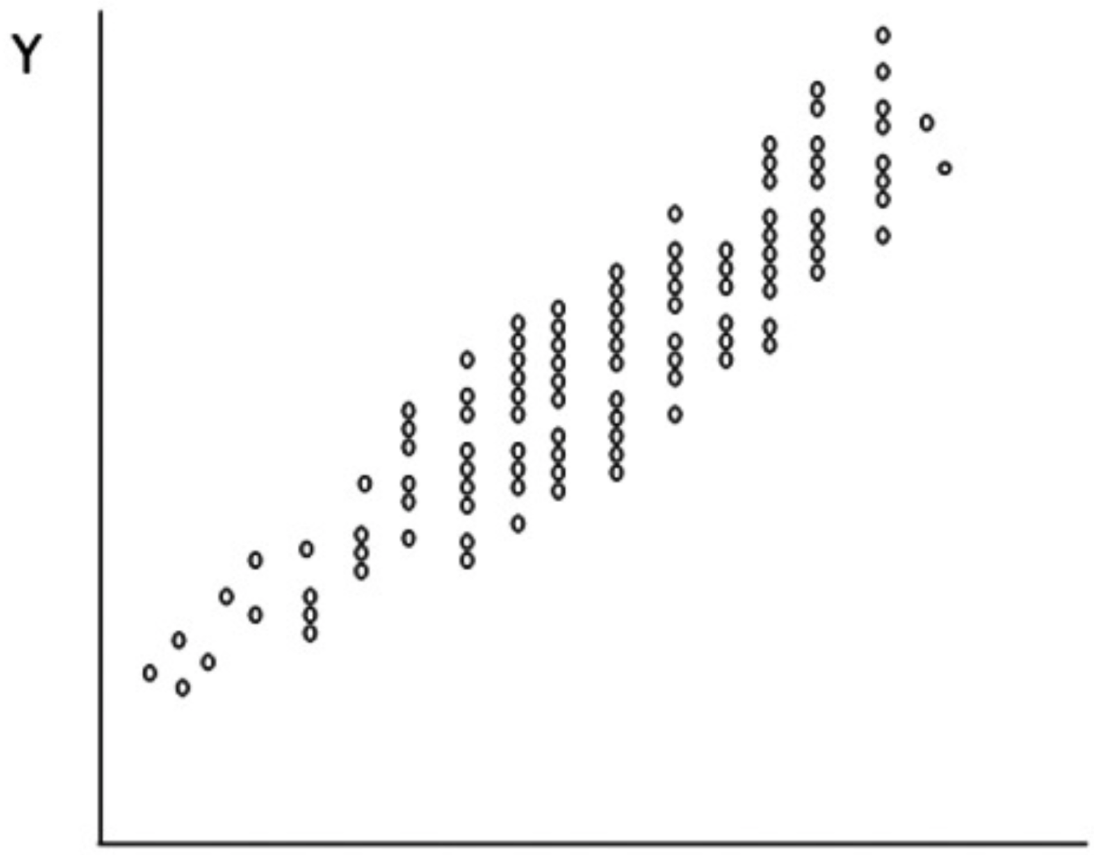
Consideraciones de aplicación

- La regresión lineal permite predecir el comportamiento de una variable (**dependiente** o predicha) a partir de otra (**independiente** o predictora).
- **Tiene presunciones como la linearidad de la relación, la normalidad, la aleatoridad de la muestra y homogeneidad de las varianzas.**
- La regresión **no prueba causalidad**.
- Un artículo que usa regresión debe mencionar o mostrar que se analizó la "**nube de pu**ntos" y que se hizo un **análisis de los residuales**.
- La línea de regresión **no debe extenderse más allá de los datos** obtenidos.

Procedimiento de aplicación

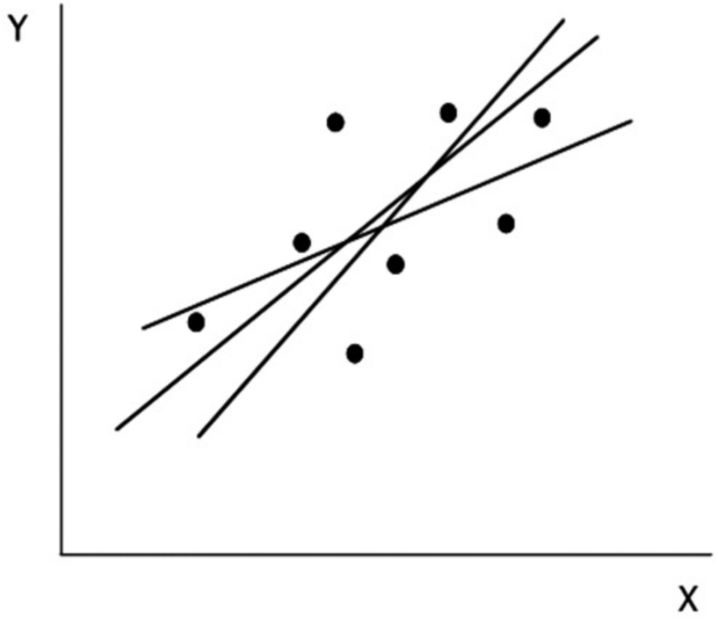
- El procedimiento a seguir puede dividirse en cuatro etapas:
- La primera aproximación es a través de dibujar los puntos en un gráfico cartesiano que muestre la relación entre las dos variables.
- Luego se determina la ecuación de la línea que mejor describa dichos puntos.
- A continuación se calcula la variabilidad de la muestra en torno a la línea de regresión calculada.
- Finalmente se pueden hacer inferencias.

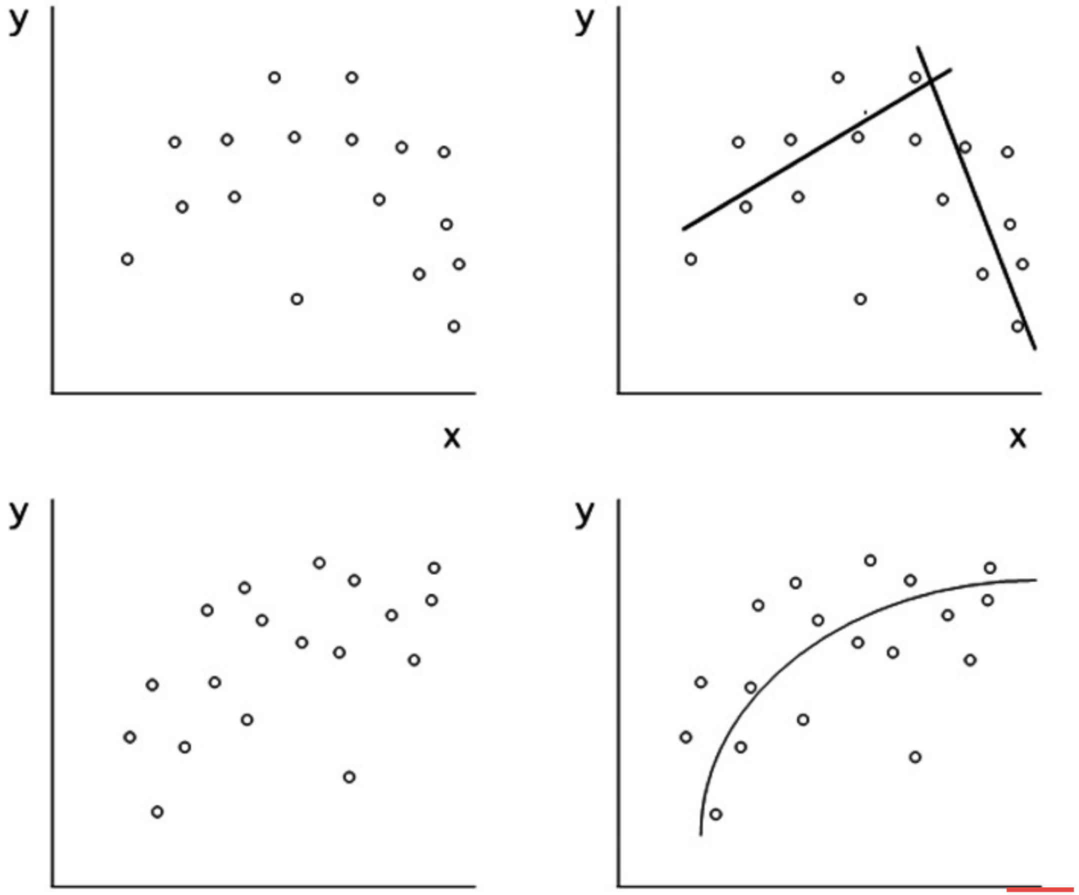
Gráfica de puntos



Aplicabilidad gráfica

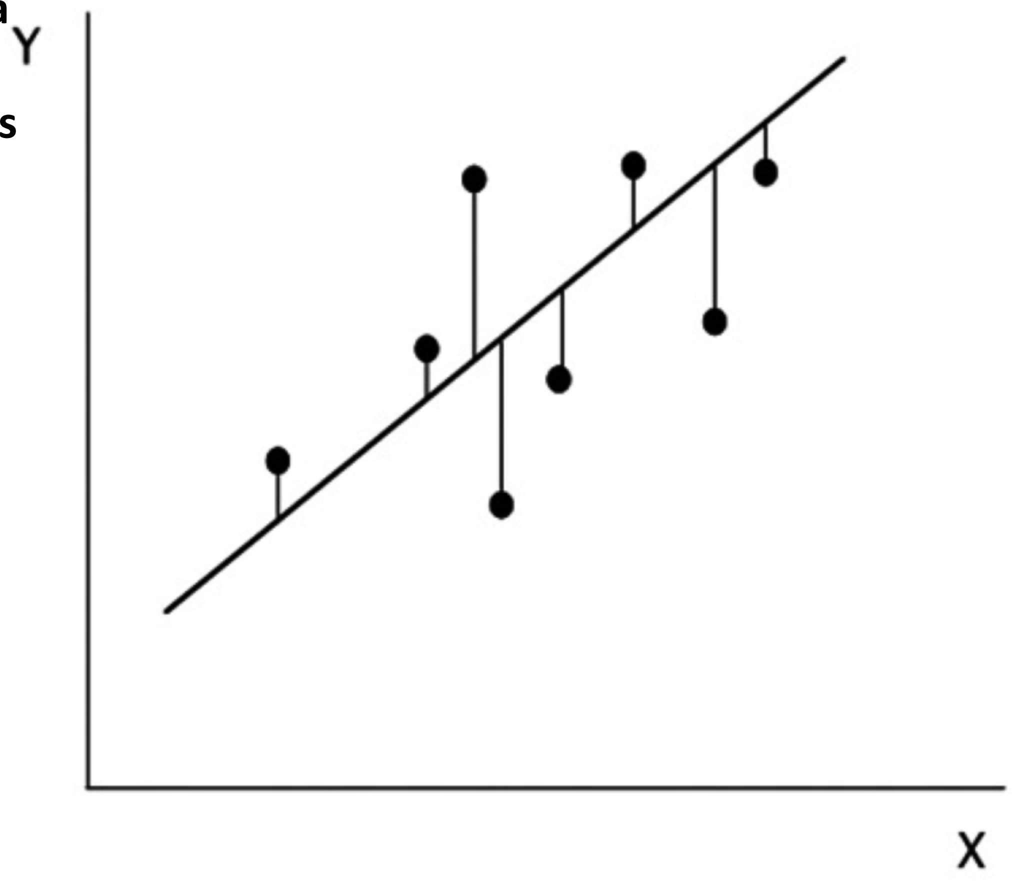
El sentido común nos informa
De la variabilidad de la muestra
Y la aplicabilidad de una regresión lineal





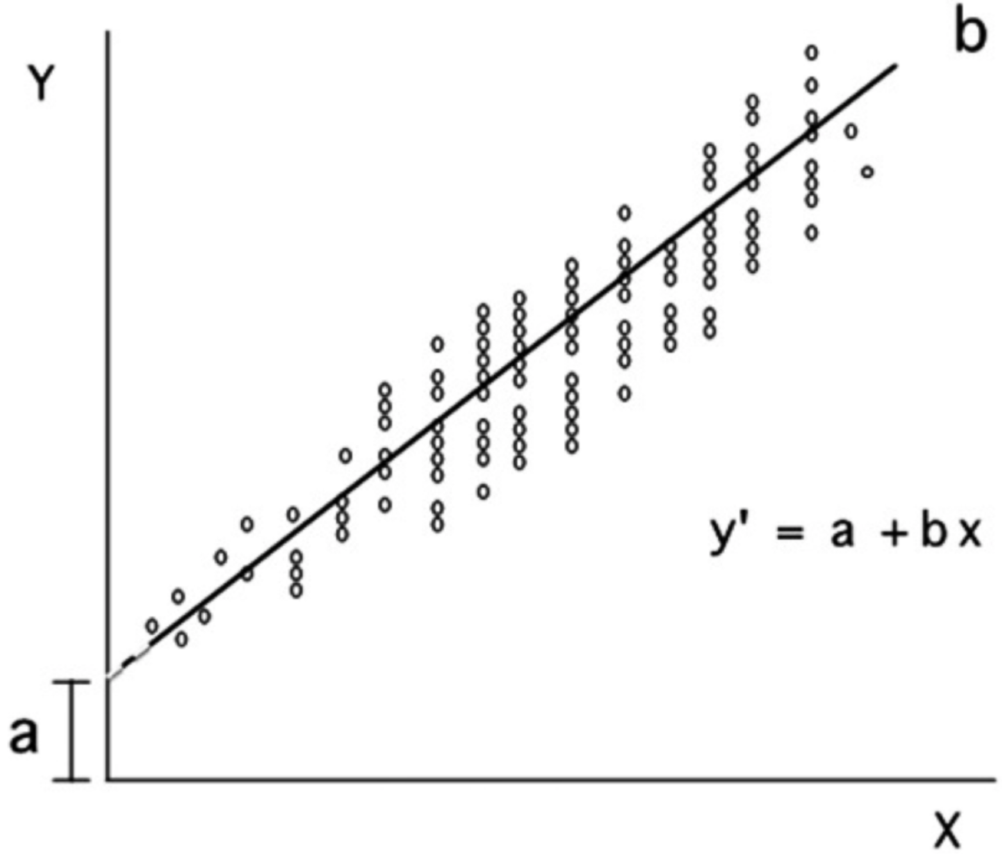
Residuos respecto a nuestra hipótesis

La distancia vertical medida desde cada observación hasta la línea estimada se denomina "residuo". Al igual que la varianza, sus cuadrados estiman la variabilidad de los puntos en relación a la línea.



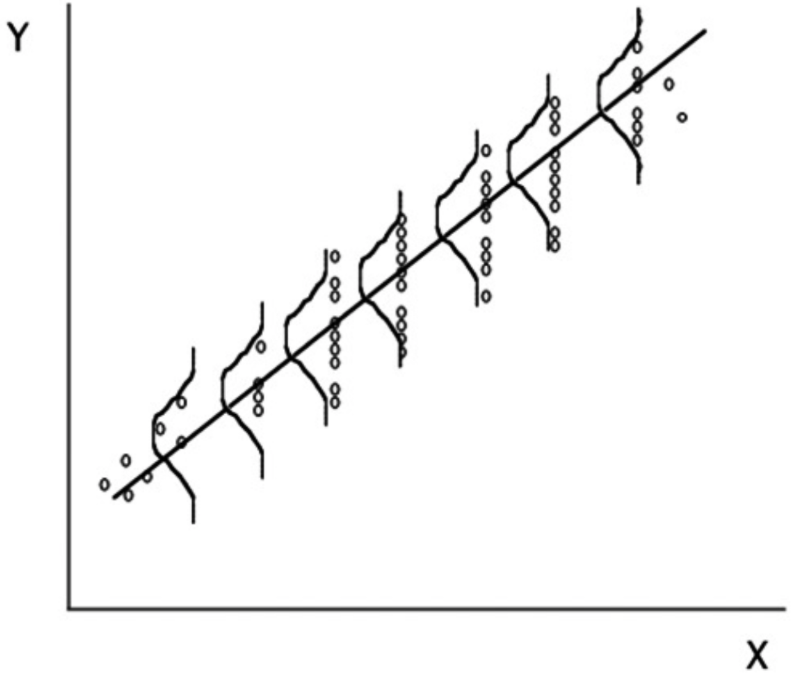
Respecto a una regresión líneal con una dimensión

Resultado predicho = intersección + (pendiente · valor del predictor).



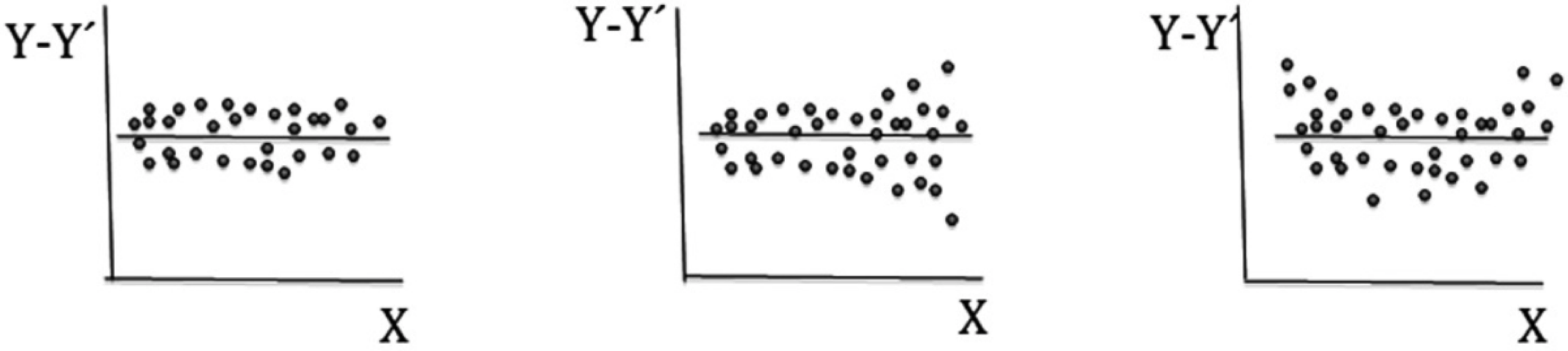
Estudio de los datos

- El modelo de regresión parte de algunas presunciones acerca de las variables y de las muestras experimentales
- La media de la población de Y a un valor dado de X crece (o decrece) linealmente a medida que X aumenta.
- Para cualquier valor de X, los posibles valores de Y se distribuyen normalmente.
- La desviación estándar de la población de Y alrededor de su media a un valor dado de X es la misma que para todos los valores de X.



Estudio de la varianza de los residuos

El **análisis de los residuales** permite verificar dos de las presunciones del modelo de regresión: la linearidad y la igualdad de las varianzas.
En A, los residuos se distribuyen normalmente sobre y bajo la línea de identidad en todo el rango de X.
En B, los residuos van aumentando en la medida que X aumenta.
En C, los residuos no se distribuyen linealmente siendo mayores cuando X es menor o mayor



A B C

Error cuadrático medio o R2

- El error cuadrático medio es una medida de comparación entre diferentes modelos y diferentes datasets.
- Al estar partido por la cardinalidad de la media permite comparar muestras
- R2 es la varianza de los residuos, es decir, el cuadrado de la desviación estándar residual
- esta cuantifica la variación inexplicada por nea de regresión y por lo tanto es un reflejo de cuan bueno es el ajuste de la ecuación obtenida.
- Error cuadrático y Varianza de residuos desde un punto de vista estaditico no es lo mismo. Aunque tengan la misma fórmula. El residuo en sí no es un error, pero si mide la diferencia con nuestra hipótesis y puede asimilarse.