#CrecimientoProfesional
#AprendeConLosPioneros

Online

CAPACITACIÓN
PROFESIONAL

Especialización en Machine Learningcon Python

Sesión 8

Docente: Luis Felipe Garayar Burneo

Reglas

Se requiere **puntualidad** para un mejor desarrollo del curso.

Para una mayor concentración **mantener silenciado el micrófono** durante la sesión.

Las preguntas se realizarán **a través del chat** y en caso de que lo requieran **podrán activar el micrófono**.

Realizar las actividades y/o tareas encomendadas en **los plazos determinados**.

Identificarse en la sala Zoom con el primer nombre y primer apellido.

MALLA CURRICULAR

- Introducción al Machine Learning (4 hrs. académicas)
- Estructura y Operaciones de Datasets (4 hrs. académicas)
- Análisis Exploratorio (4 hrs. académicas)
- Métodos Multivariados y Balanceo de Datos (4 hrs. académicas)
- Modelos Supervisados para Regresión (8 hrs. académicas)
- Modelos Supervisados para Clasificación (12 hrs. académicas)
- Modelos No Supervisados (8 hrs. académicas)
- Introducción al Deep Learning (4 hrs. académicas)
- Introducción a MLOps (4 hrs. académicas)

CERTIFICACIÓN FINAL

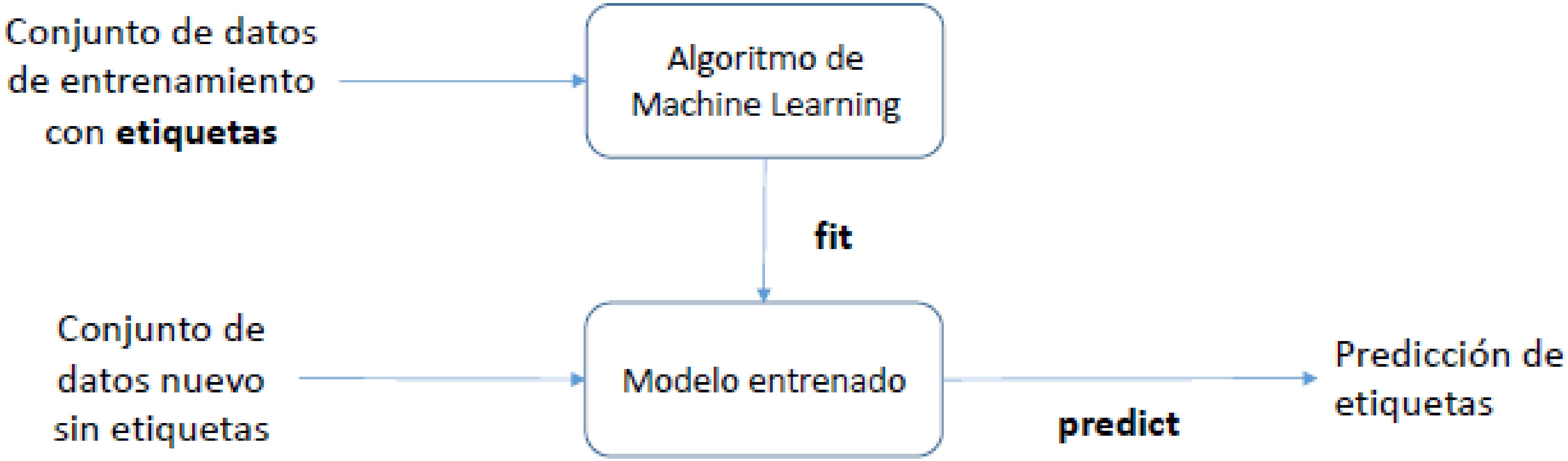
Especialista en Machine Learning

(56 horas académicas)

Contenido

- Entendiendo los métodos supervisados
- Regresión lineal simple. Suposiciones del modelo. Multicolinealidad.
- Regresión lineal múltiple. Multicolinealidad.
- Introducción a las regresiones penalizadas (Ridge, Lasso, elastic net)
- Comparación entre modelos. Evaluación de Desempeño.

Entendiendo los métodos supervisados

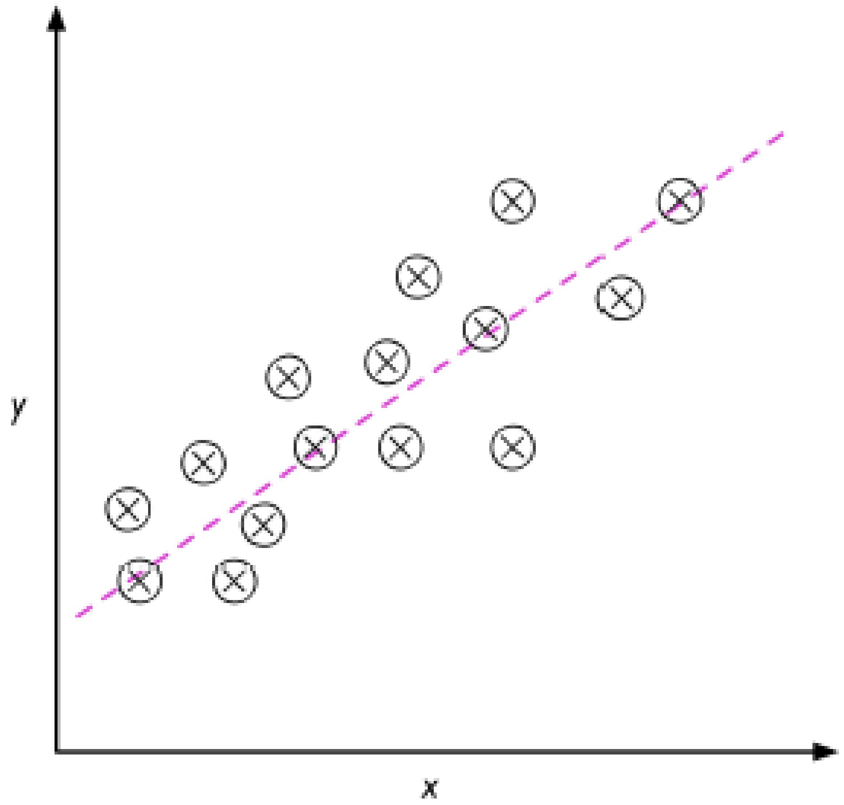


Aprendizaje supervisado: Regresión

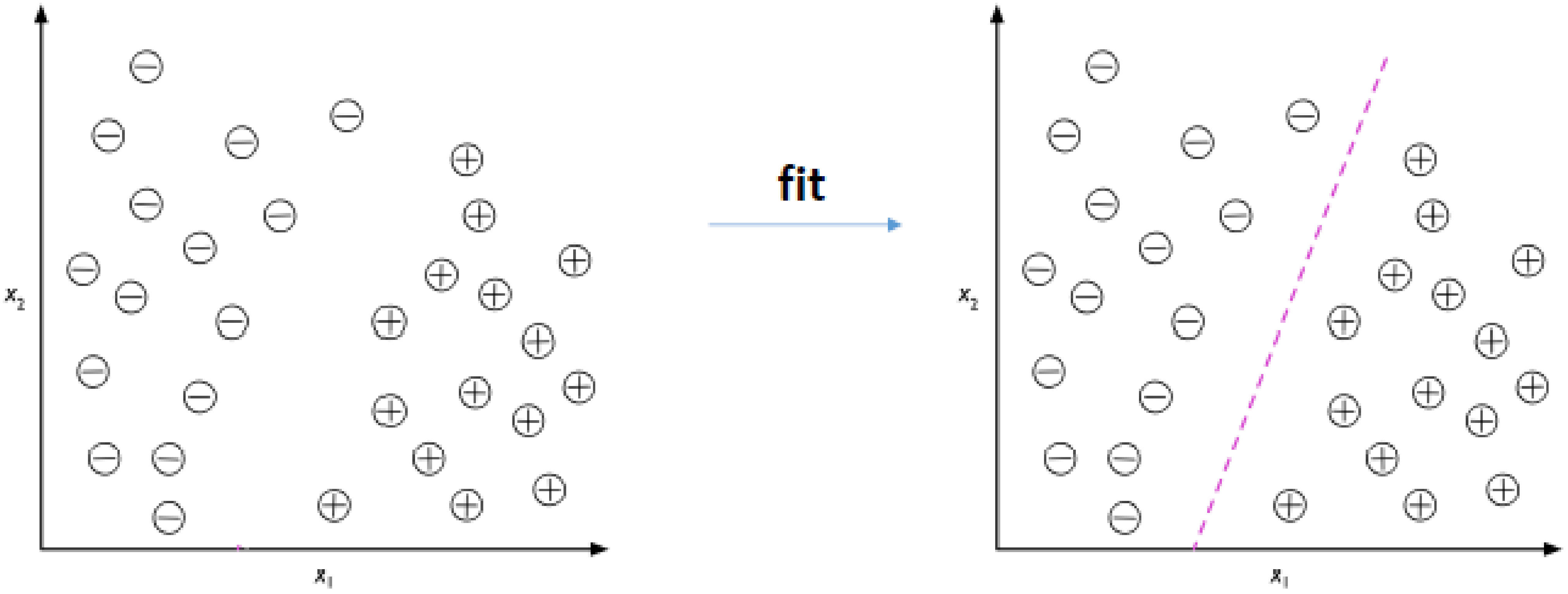
En un modelo de regresión queremos predecir un valor continuo de target.

En la imagen, dada una variable x intentamos estimar el valor de la variable y

La línea elegida (el modelo) es resultado de ajustar un modelo en el que se intenta minimizar la distancia entre los valores reales y la predicción del modelo.



Aprendizaje supervisado: Clasificación



Aprendizaje supervisado: Términos

Sueldo	Antiguedad en sistema financiero	Edad	Capacidad de endeudamiento	Pagó el préstamo
				Sí
				Sí
				Sí
				No
				No

Cantidad de muestras

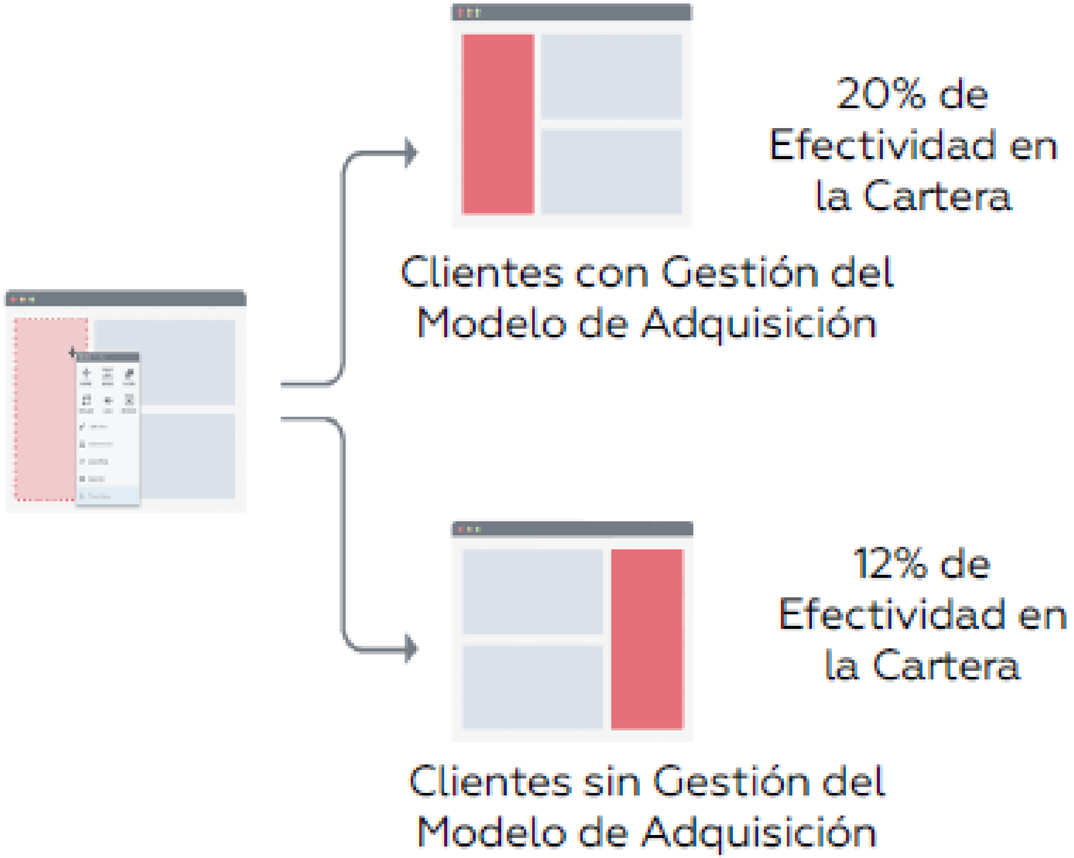
Características
(**Features**, atributos, dimensiones)

Etiqueta
(Label / **Target**)

Metodología de desarrollo de algoritmos supervisados

- ❑ Entendimiento del problema de la naturaleza a resolver con datos. Búsqueda de stakeholders o sponsors.
- ❑ Análisis exploratorio de datos:
    - ✓ Completitud de los datos / Imputación de valores perdidos.
    - ✓ Detección de outliers.
    - ✓ Transformaciones.
    - ✓ Recodificaciones.
- ❑ Balanceo de datos.
- ❑ Selección de variables. (Met. Estadísticas vs ML).
- ❑ Modelamiento y entendimiento de drivers o factores que influyen en la solución.
- ❑ Validación técnica. Validación negocio.
- ❑ Implementación.

Validación técnica y de negocio mediante un piloto



CAPACITACIÓN PROFESIONAL

Implementación de la solución analítica supervisada

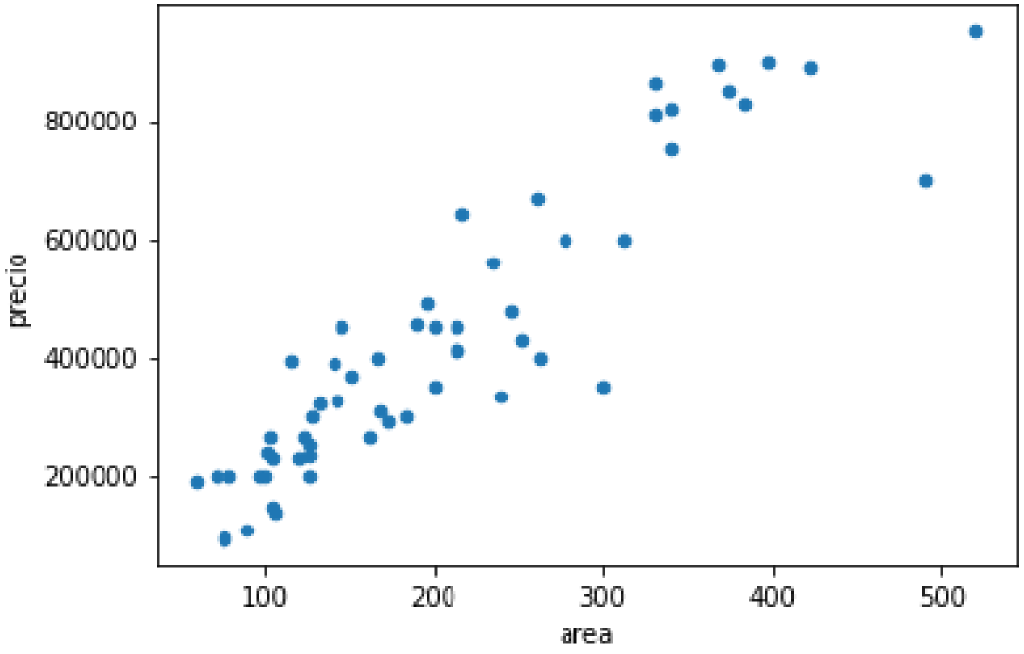
Probabilidad	N° Clientes	N° Sucesos VD	% Sucesos VD/ N° Clientes
0,9	10 000	100	36%
0,8	10 000	60	22%
0,7	10 000	40	14%
0,6	10 000	33	12%
0,5	10 000	20	7%
0,4	10 000	10	4%
0,3	10 000	5	2%
0,2	10 000	5	2%
0,1	10 000	3	1%
0	10 000	3	1%
Total	100 000	279	100%

GRUPOS DE EJECUCIÓN	N° CLIENTES	% SUCESOS ACUMULADOS	EFECTIVIDAD	LIFT
RECOMENDADO	30 000	200	0,67%	2,39
MEDIO	30 000	63	0,21%	0,75
BAJO	40 000	16	0,04%	0,14
TOTAL	100 000	279	0,28%	

Regresión Lineal: Empecemos con un ejemplo

Imaginemos que tenemos una base de datos con precios de casas.

Área	Precio
132.3	319000
141.5	389000
71.2	199000
277.4	599000
338.7	750000
330.5	810000
422.9	890000
367.6	895000
251.3	429000

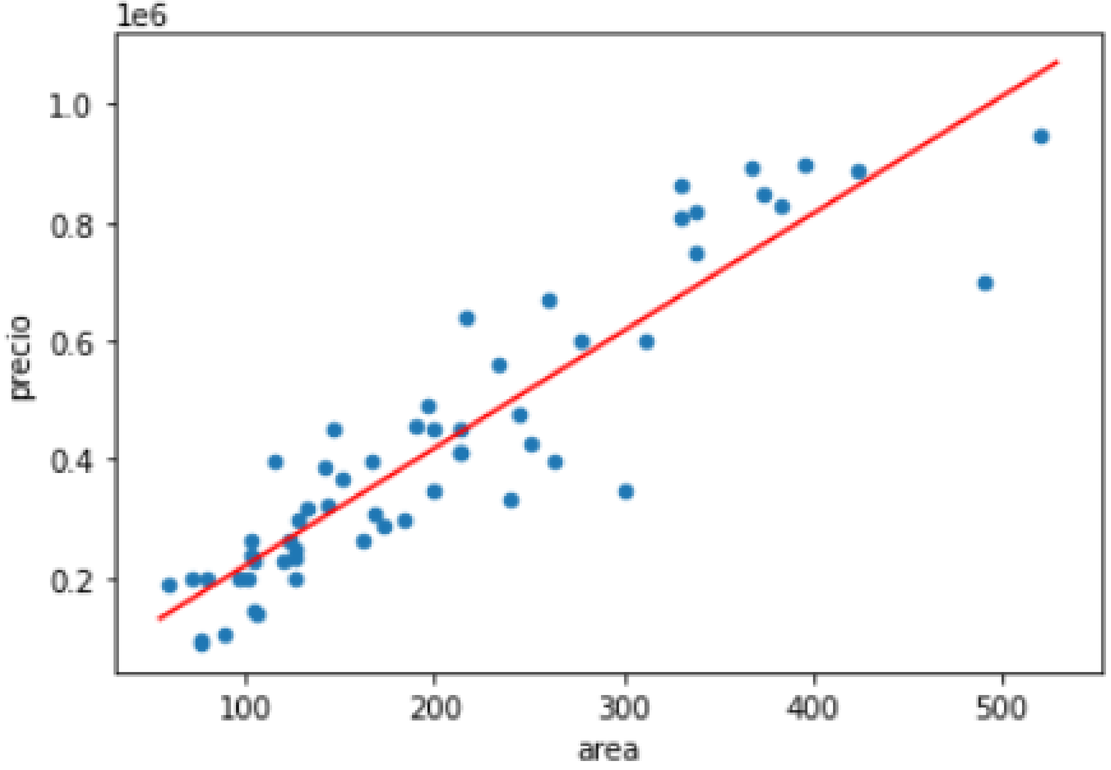


¿cuál sería el precio de una casa de 150 m2?

CAPACITACIÓN
PROFESIONAL

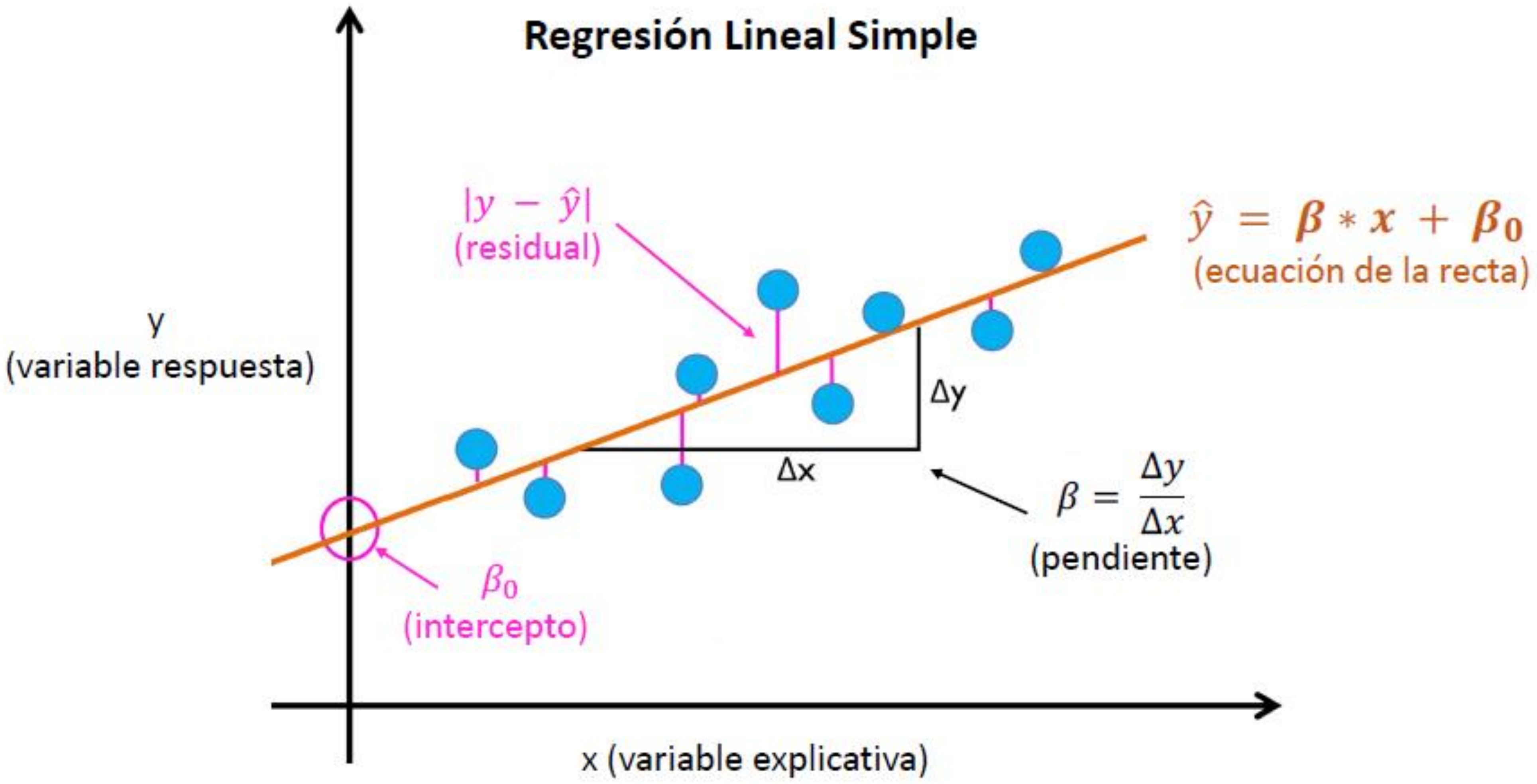
Regresión Lineal Simple

- Es un método de aprendizaje supervisado
- Es un método de regresión porque estimamos un valor numérico real
- Se predice una variable de respuesta con una variable explicativa

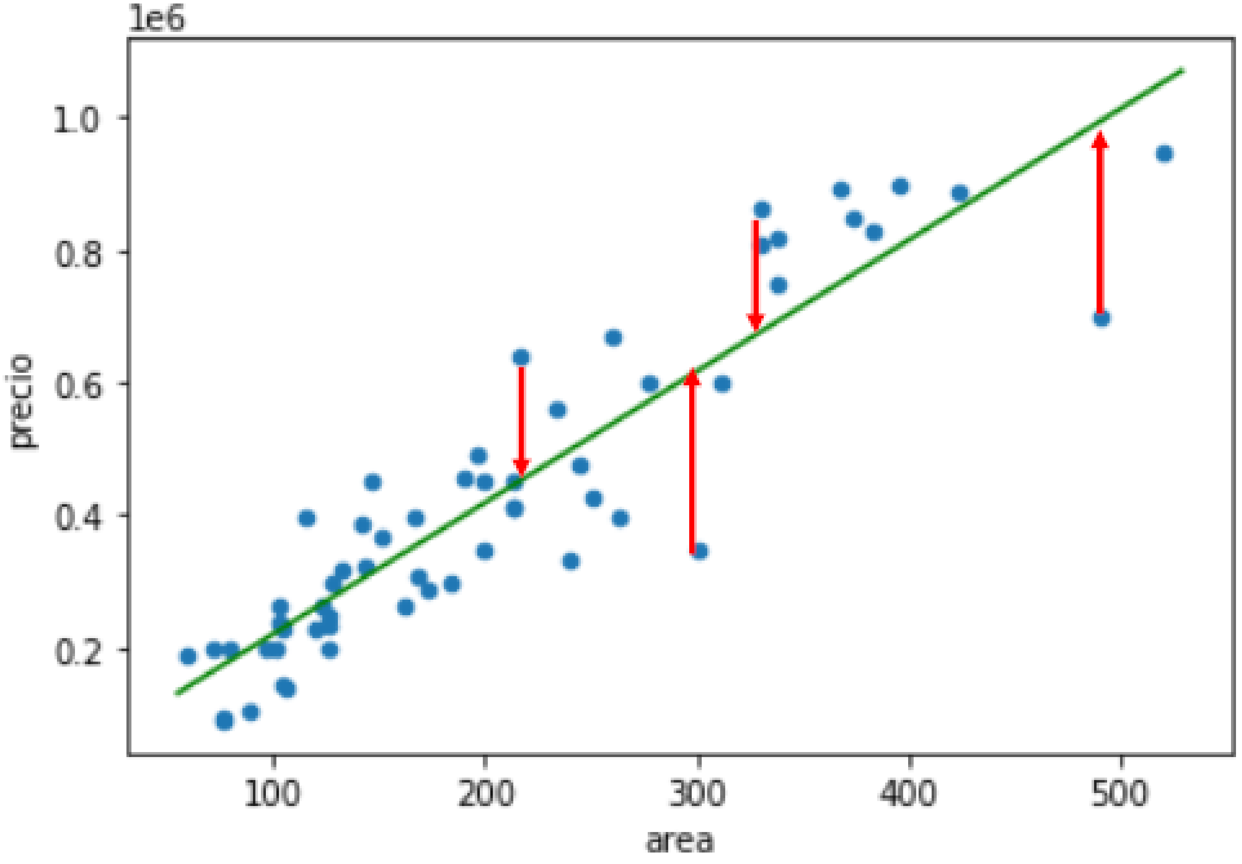


CAPACITACIÓN PROFESIONAL

Regresión Lineal Simple



Regresión Lineal Simple: eligiendo la mejor recta



$$J(\beta) = \frac{1}{2}\sum_{1}^{n} (\hat{y} - y)^2$$

$$\hat{y} = \beta * x + \beta_0$$

$$y: valor\ real$$

Utilizando un algoritmo de optimización, se eligen los **mejores betas** para minimizar la función de costo

Regresión Lineal Simple: interpretación del modelo

$$\hat{Y} = 600 + 300X$$

- b=300 → Cambio en Y por cada unidad de cambio en X. Por cada año de experiencia laboral, el sueldo mensual aumenta 300 €.
- a=600 → Valor medio de Y cuando X=0. Sueldo medio de aquellas personas sin experiencia laboral.

Una persona con 3 años de experiencia laboral, ¿qué sueldo mensual tendrá? Interpreta el resultado.

$$X = 3 \Rightarrow \hat{Y} = 600 + 300 * 3 = 1500$$

- $\hat{Y} = 1500$ → Valor promedio previsto para todos los sujetos que han obtenido en la variable X un valor de Xi. Las personas con 3 años de experiencia tienen un sueldo promedio de 1500 €.

Regresión Lineal Simple: interpretación del modelo

Si una persona con 3 años de experiencia laboral tiene un sueldo mensual de 1700 €, ¿cuál será su error asociado? Interpreta el resultado.

$$e = Y - \hat{Y} = 1700 - 1500 = 200$$

- El modelo estimó un sueldo de 1500 € para una persona con 3 años de experiencia laboral. Si esta persona concreta tiene un sueldo de 1700 €, esta diferencia de 200 € es el error; aquello que el modelo no explica.

Componentes de la variación en la regresión lineal

$$\sum_{i=1}^{N}\left(Y-\overline{Y}\right)^2=\sum_{i=1}^{N}(\hat{Y}-\overline{Y})^2+\sum_{i=1}^{N}(Y-\hat{Y})^2$$

- Suma de cuadrados total = Suma de Cuadrados **Explicada** + Suma de Cuadrados **No Explicada**

- Variación Total = **Variación Explicada** + Variación **No Explicada**

Explicada por el modelo Y = f(x).	No explicada por el modelo Y = f(x).

CAPACITACIÓN PROFESIONAL

Contraste de significación global (ANOVA)

Prueba F para la significancia del modelo (general)

Muestra si hay una relación lineal entre todas las variables x (consideradas en forma conjunta) e y

Usa el estadístico de prueba F

Hipótesis:

H_0: $\beta_1 = \beta_2 = \ldots = \beta_k = 0$ (No hay relación lineal)

H_A: Al menos un $\beta_i \neq 0$ (Existe relación lineal entre (y) y al menos un x_i)

Contraste de significación global (ANOVA)

A partir de esta información muestral, podemos calcular el numerador y denominador del estadístico F_{exp}.

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	Estadistico F
Regresión	SCE	$k-1$	$\frac{SCE}{k-1}$	$\frac{\frac{SCE}{k-1}}{\frac{SCR}{n-k}}$
Residual	SCR	$n-k$	$\frac{SCR}{n-k}$	
Total	SCT	$n-1$		

Contraste de significación global (ANOVA)

La tabla Anova es el estándar estadístico para resumir las principales características de un modelo de regresión

Fuente	SS	DF	MS
Modelo	1591.9902	1	1591.9902
Residuales	851.469256	72	11.8259619
Total	2443.45946	73	33.4720474

Suma de residuales al cuadrado

Número de Obs	=	74
F(1,72)	=	134.62
Prob > F	=	0.0000
R-2	=	0.6515
R-2 Ajustado	=	0.6467
Root MSE	=	3.4389

Significancia global

R^2 y R^2 – ajustado

VAR	Coef	Err. Sta.	t	P>|t|	[95% Inter. de conf.]	
X_	-.0060087	.0005179	-11.60	0.000	-.0070411	-.0049763
_cons	39.44028	1.614003	24.44	0.000	36.22283	42.65774

Intervalo de confianza

Coeficiente beta

Significancia individual (estadístico t)

P-value

Métricas de evaluación para la regresión lineal

Coeficiente de determinación

Expresa la proporción de la variación de los resultados que puede explicarse por el modelo

Error cuadrático medio
(mean squared error)

$$MSE = \frac{1}{n}\sum_{1}^{n} (\hat{y} - y)^2$$

$$\hat{y} = \beta * x + \beta_0$$
$$y: valor\ real$$

Raíz del error cuadrático medio
(root mean squared error)

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

$$\hat{y} = \beta * x + \beta_0$$
$$y: valor\ real$$

$$R^2 = 1 - \frac{SSE}{SST}$$

$$SSE = \sum_{1}^{n} (\hat{y} - y)^2$$

sum of squared errors

$$SST = \sum_{1}^{n} (y - \mu_y)^2$$

sum of total squares

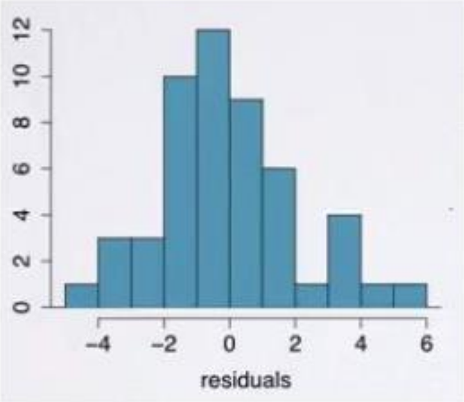
Supuestos paramétricos para la regresión lineal

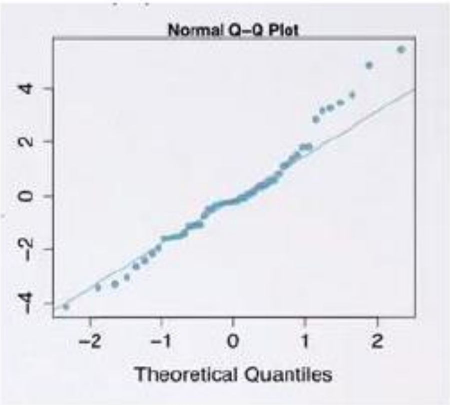
- **Independencia**
  Entre los residuales (ei), mediante Durbin Watson: Si DW = 2 los ei son completamente independientes. Entre 1.5 y 2.5 se considera que existe independencia, DW<2 autocorrelación negativa.

- **Homocedasticidad**
  Igualdad de varianzas de ei y los pronósticos. El supuesto implica que la variación de los ei sea uniforme en todo el rango de valores de los pronósticos (gráfico sin asociación).

- **Normalidad**
  Los residuales tiene distribución normal?, por la prueba de Shapiro, si el p-value > 0.05 entonces se afirma que poseen esta distribución.

- **No Colinealidad (Evitar la Multicolinealidad)**
  No correlación entre variables regresoras.

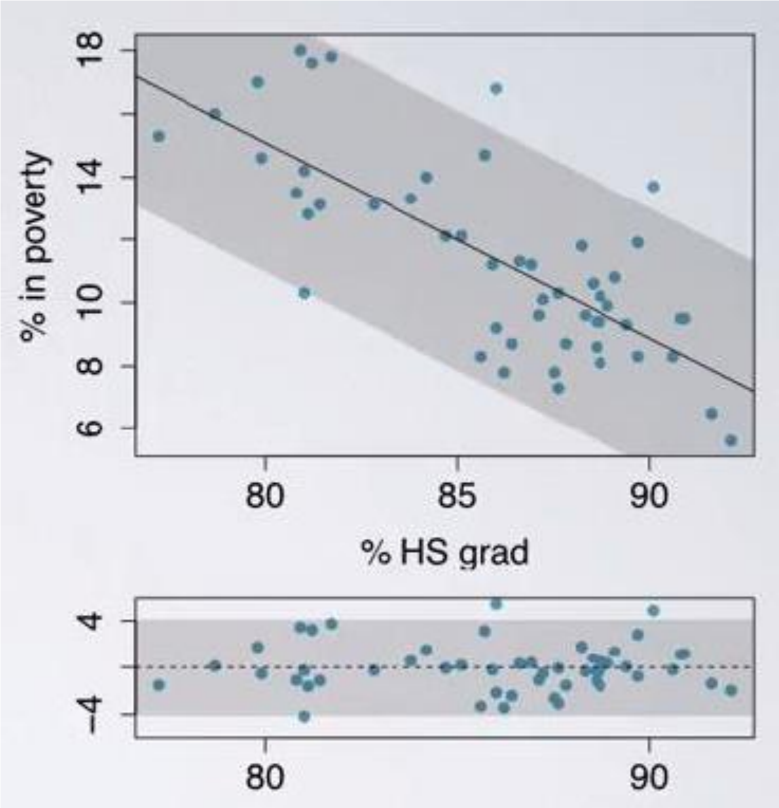




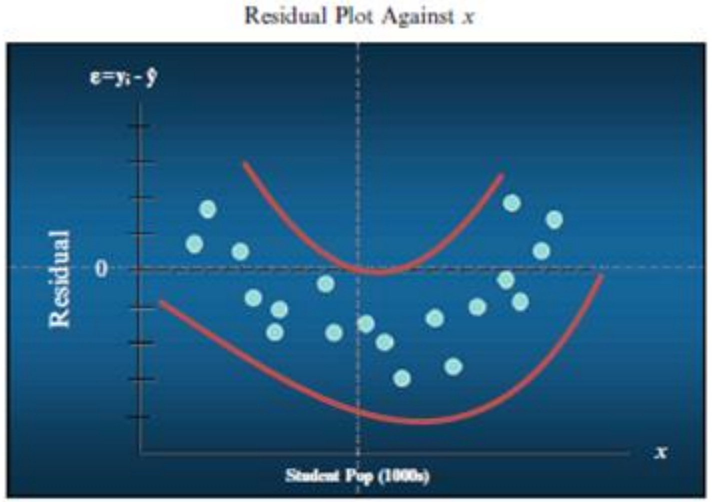
CAPACITACIÓN PROFESIONAL

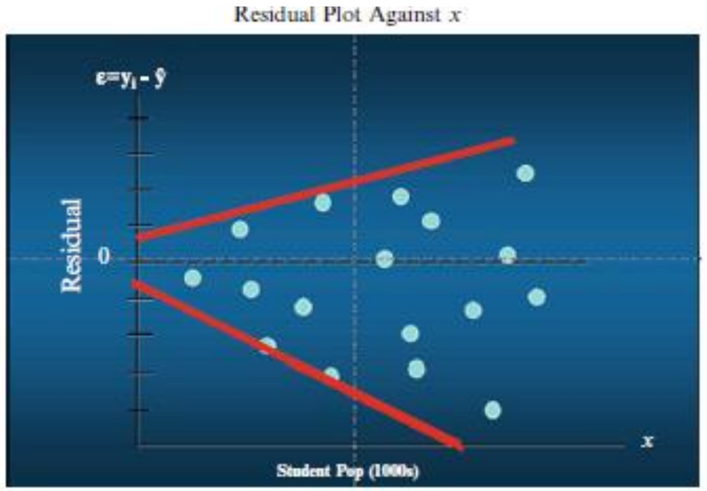
Supuestos paramétricos para la regresión lineal

Homocedasticidad

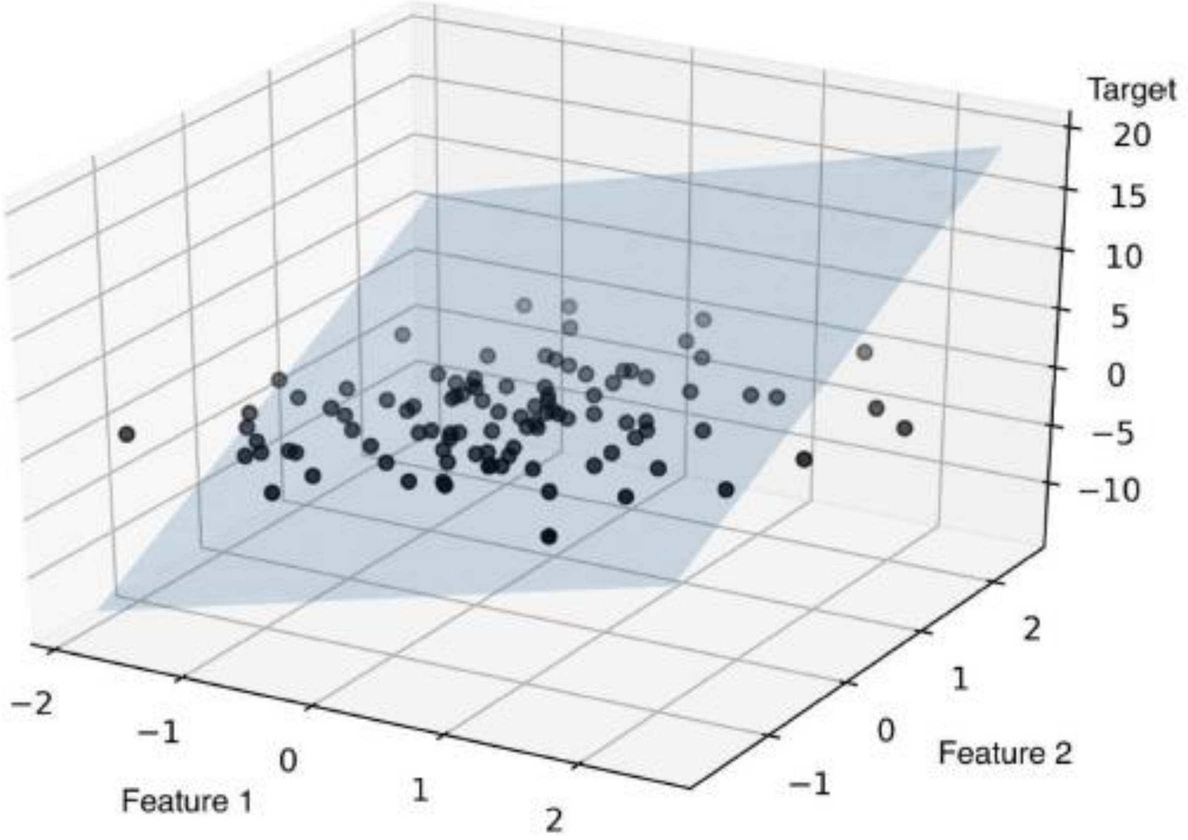


NO DEBE SER





Regresión Lineal Múltiple



Podemos generalizar la idea de la regresión simple utilizando múltiples variables explicativas

$$\hat{y} = \beta_0 + \beta_1 * x_1 + \beta_2 * x_2 \; ... + \beta_j * x_j$$

La función de costo sigue siendo la misma, lo que cambia es la estimación $\hat{y}$

$$J(\beta) = \frac{1}{2}\sum_{1}^{n} (\hat{y} - y)^2$$

Regresión Lineal Múltiple: multicolinealidad

Multicolinealidad

La multicolinealidad ocurre cuando un predictor está linealmente relacionado con uno o varios de los otros predictores del modelo. La consecuencia es no poder identificar de forma precisa el efecto individual que tiene cada predictor sobre la variable respuesta, lo que se traduce en un incremento de la varianza de los coeficientes de regresión estimados hasta el punto de que resulta imposible establecer su significancia estadística.

Como reconocerlo:

- Si el R2 es alto pero los coeficientes son no significativos, es un indicio.
- Si existe correlación entre las variables, si en su mayoría son mayores a 0.5, es un indicio.

Factor de Inflación de la Varianza (**VIF**),

$$VIF_{\hat{\beta}_j} = \frac{1}{1 - R^2}$$

VIF = 1, ausencia total
1<VIF<5, moderado
5 <VIF< 10, muy afectado

Posibles soluciones para la multicolinealidad

Algunas de las posibles soluciones al problema de multicolinealidad son las siguientes:

- mejora del diseño muestral estrayendo la información máxima de la variables observadas.
- eliminación de las variables que se sospechan son causantes de la multicolinealidad.
- en caso de disponer de pocas observaciones, aumentar el tamaño de la muestra.
- utilizar la relación extramuestral que permita realizar relaciones entre los parámetros (información a priori) que permita estimar el modelo por mínimos cuadrados restringidos.

Regresión Lineal Múltiple: variables cualitativas

Variables cualitativas

• Muchas veces en el modelo de regresión aparecen factores cualitativos (sexo, raza, estado civil,….). En estos casos la información relevante se puede representar con la ayuda de variables ficticias.

• Las variables ficticias son variables binarias que toman valor 0,1.

• Al definir una variable ficticia debemos decidir a qué acontecimiento se le asigna el valor 1, y a cuál el 0.

Ejemplo:
Supóngase que se pretende transformar la variable "**medios de transporte más comunes**" de tres categorías: 1=autobús, 2=tren y 3=avión.
La conversión podría efectuarse por medio de dos variables dicotómicas, F1 , F2 y F3. Los valores que éstas tomarían para representar cada categoría serían los siguientes:

Categoría	F1	F2	F3
Autobús	1	0	0
Tren	0	1	0
Avión	0	0	1

Regresión Espúrea

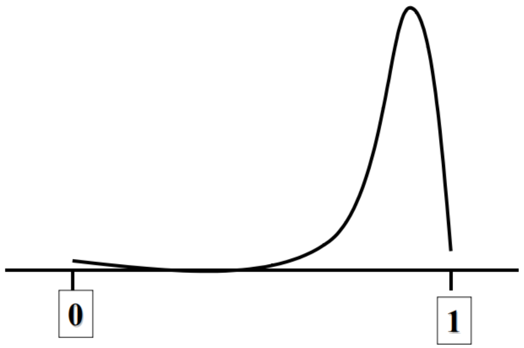
La relación espúrea da la impresión de la existencia de un vínculo apreciable entre dos grupos que es inválido cuando se examina objetivamente. Es una relación en la cual dos acontecimientos no tienen conexión lógica, aunque se puede implicar que la tienen debido a un tercer factor no considerado aún (llamado "factor de confusión" o "variable escondida").

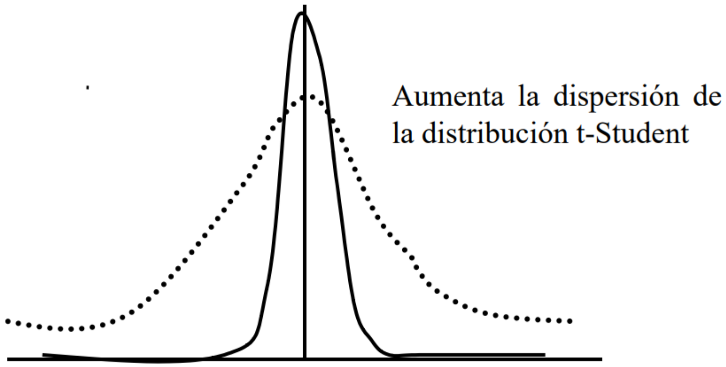
Ejemplo: examinando las ventas de helados de una ciudad. Estas son más altas cuando la tasa de sofocamientos es mayor. Sostener que la venta de helados causa los sofocamientos sería implicar una relación espuria entre las dos. En realidad, una ola de calor puede haber causado ambas. La ola de calor es un ejemplo de variable escondida.

Problemas de la regresión espúrea:

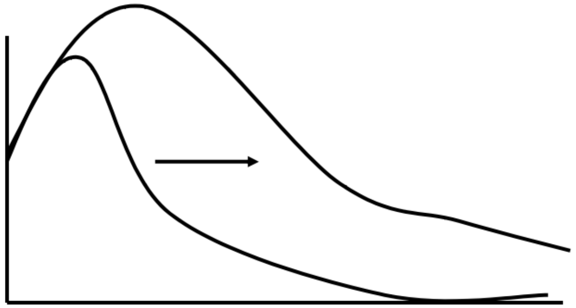
1) Los estimadores son estadísticamente significativos, presentando estadísticos t y F elevados, que rechazan la hipótesis nula.
2) El valor de la R2 es muy cercano al valor de 1, indicando que el modelo es adecuado
3) El estadístico DW tiende a cero. *Una regla para determinar si la regresión es falsa DW << R2*

Los valores de R^2 tienden a agruparse alrededor de 0.95





Se presentan valores muy altos de t-calculado



Se presentan valores muy altos del estadístico F

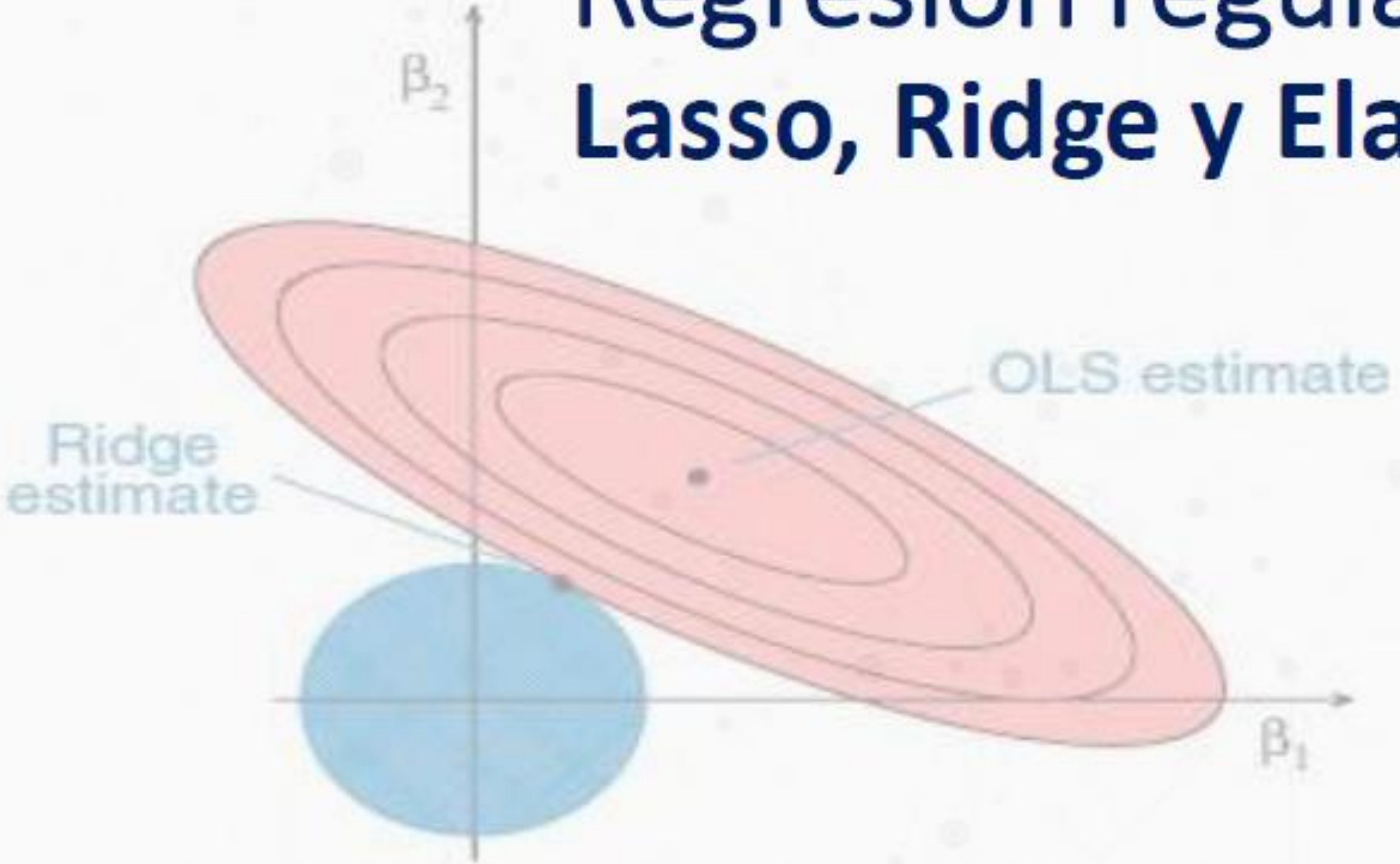
Algunos problemas con los modelos de regresión

La **regresión lineal** trata de modelar la relación entre una variable continua y una o más variables independientes mediante el ajuste de una ecuación lineal. Tres de las limitaciones que aparecen en la práctica al tratar de emplear este tipo de modelos (ajustados por mínimos cuadrados ordinarios) son:

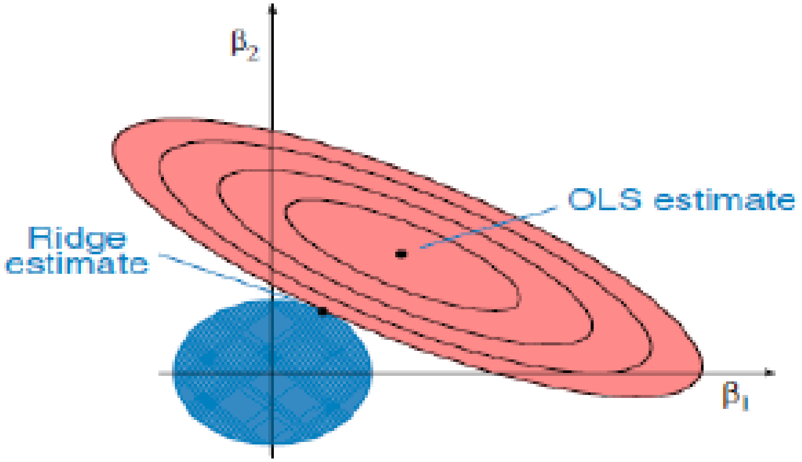
- Se ven perjudicados por la incorporación de predictores correlacionados.
- No realizan selección de predictores, todos los predictores se incorporan en el modelo aunque no aporten información relevante.
- No pueden ajustarse cuando el número de predictores es superior al número de observaciones.

Una forma de atenuar el impacto de estos problemas es utilizar estrategias de regularización como ***ridge*, *Lasso* o *Elastic Net***, que fuerzan a que los coeficientes del modelo tiendan a cero, minimizando así el riesgo de ***overfitting***, reduciendo varianza, atenuado el efecto de la correlación entre predictores y reduciendo la influencia en el modelo de los predictores menos relevantes.

Regresión regularizada:
Lasso, Ridge y Elastic net



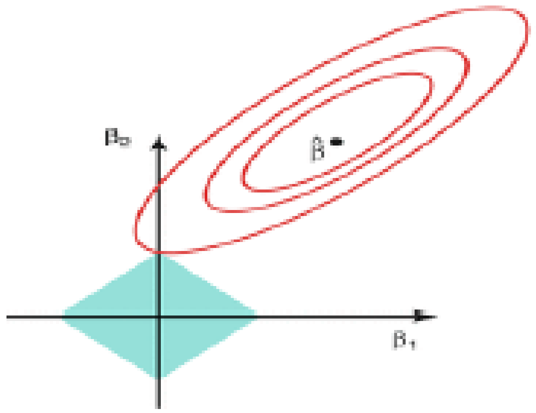
Modelo de regresión penalizada: Ridge



$$\underbrace{\sum_{i=1}^{n}\left(y_i - \beta_o - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2}_{\text{Mínimos cuadrados}} + \underbrace{\lambda\sum_{j=1}^{p}\beta_j^2}_{\text{Penalización ridge}}$$

- Es útil cuando existe colinearidad entre las variables utilizadas para el entrenamiento.
- El factor λ (lambda) sirve para controlar la intensidad de la regularización. Se utiliza la norma de regularización L2.
- La elección de este parámetro involucra un balance entre los componentes de sesgo y varianza del error cuadrático medio al estimar β.

Modelo de regresión penalizada: Lasso



$$\underbrace{\sum_{i=1}^{n}\left(y_i - \beta_o - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2}_{\text{Mínimos cuadrados}} + \underbrace{\lambda\sum_{j=1}^{p}|\beta_j|}_{\text{Penalización lasso}}$$

- Permite reducir a valores cercanos a cero los coeficientes de las variables menos relevantes.
- Se utiliza para la reducción de la dimensionalidad.
- El factor λ (lambda) controla el nivel de regularización.
- Lasso es una técnica de regresión lineal regularizada, como Ridge, con la leve diferencia en la penalización. (Norma L1 en lugar de L2)

Modelo de regresión penalizada: Lasso

- Para valores crecientes de λ , los coeficientes βj se contraen hacia cero como en Ridge (shrinkage), con la diferencia de que algunos de ellos se anulan.

- Esto es, Lasso produce estimación y selección de variables en forma continua y simultánea, siendo especialmente útil en el caso $p>=n$.

- En los últimos años se han presentado algunas generalizaciones y extensiones de las técnicas presentadas anteriormente, especialmente diseñadas para ciertas situaciones particulares.

Modelo de regresión penalizada: elastic net

Es una combinación de Ridge y Lasso. Se decide, que peso se le da a cada método de penalización y se implementa la regresión.

$$\underbrace{\sum_{i=1}^{n}\left(y_i-\beta_o-\sum_{j=1}^{p}\beta_j x_{ij}\right)^2}_{\text{Mínimos cuadrados}} + \lambda_1 \underbrace{\sum_{j=1}^{p}\beta_j^2}_{\text{Ridge}} + \lambda_2 \underbrace{\sum_{j=1}^{p}|\beta_j|}_{\text{Lasso}}$$

$$\hat{\beta}^e = \arg\min_{\beta}\sum_{i=1}^{n}(y_i - x_i^T\beta)^2 + \frac{1-\alpha}{2}\lambda\sum_{i=1}^{p}\beta_i^2 + \alpha\lambda\sum_{i=1}^{p}|\beta_i|$$

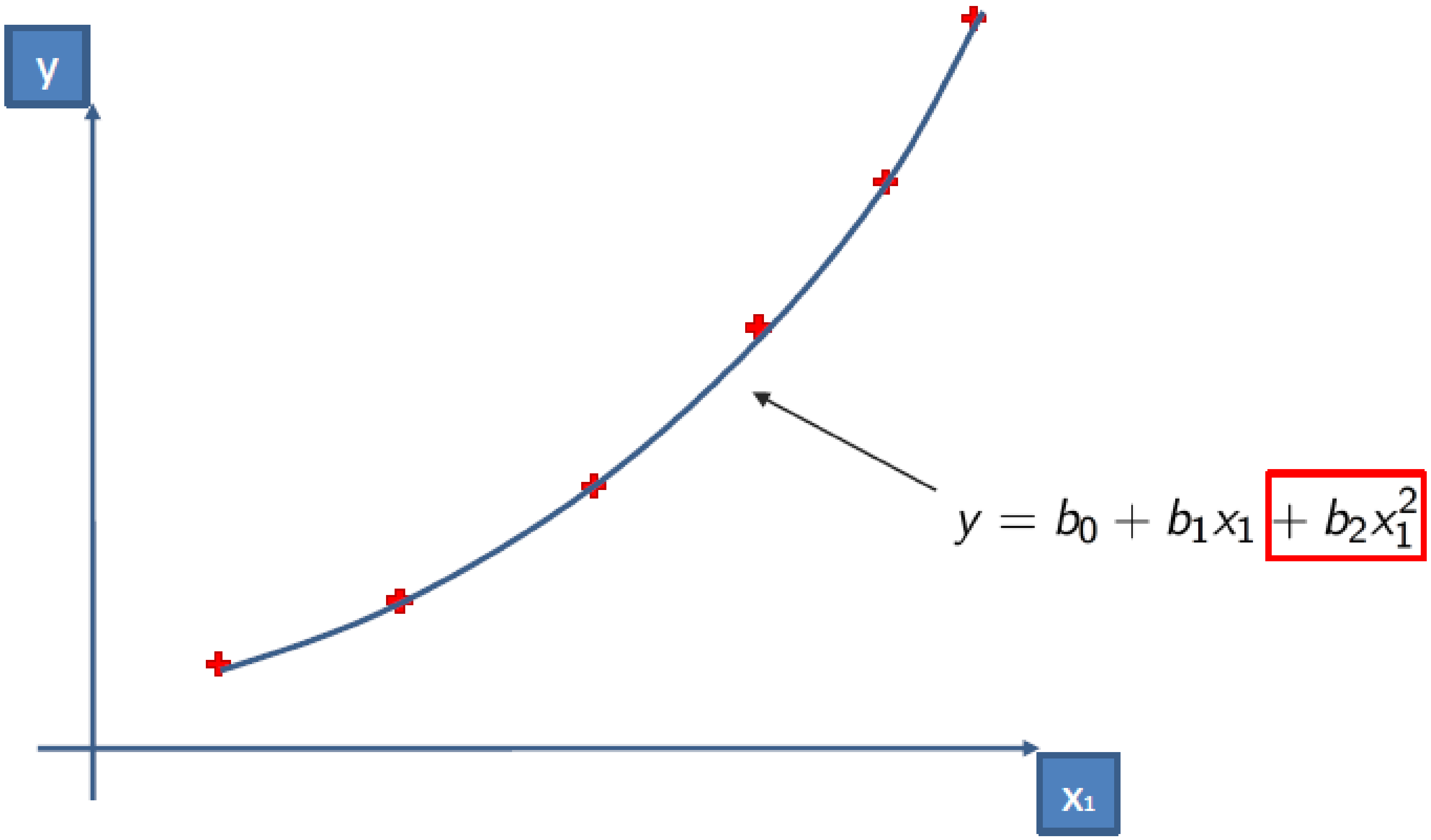
donde λ es un parámetro de precisión

- Si $\lambda = 0$, regresión lineal tradicional ($\hat{\beta}^e = \hat{\beta}$).
- Si $\lambda = \infty$, $\hat{\beta}^e = 0$
- Si $\alpha = 0$, entonces $\hat{\beta}^e = \hat{\beta}^{ridge}$
- Si $\alpha = 1$, entonces $\hat{\beta}^e = \hat{\beta}^{lasso}$

Regresiones

Regresión Lineal Simple	$y = b_0 + b_1 x_1$
Regresión Lineal Múltiple	$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n$
Regresión Lineal Polinómica	$y = b_0 + b_1 x_1 + b_2 x_1^2 + \ldots + b_n x_1^n$

Regresión Polinómica



CAPACITACIÓN
PROFESIONAL

Perú

Online