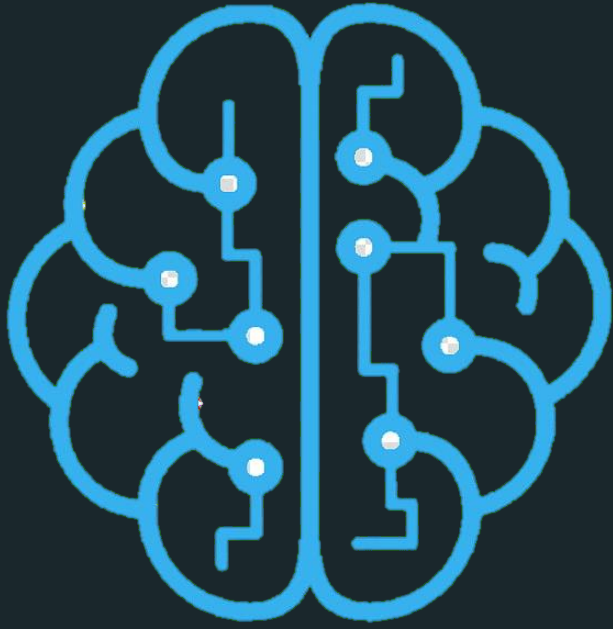


APRENDIZAJE DE MÁQUINA

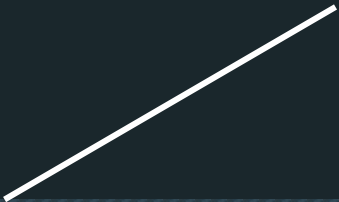
Aprendizaje Supervisado - Regresion



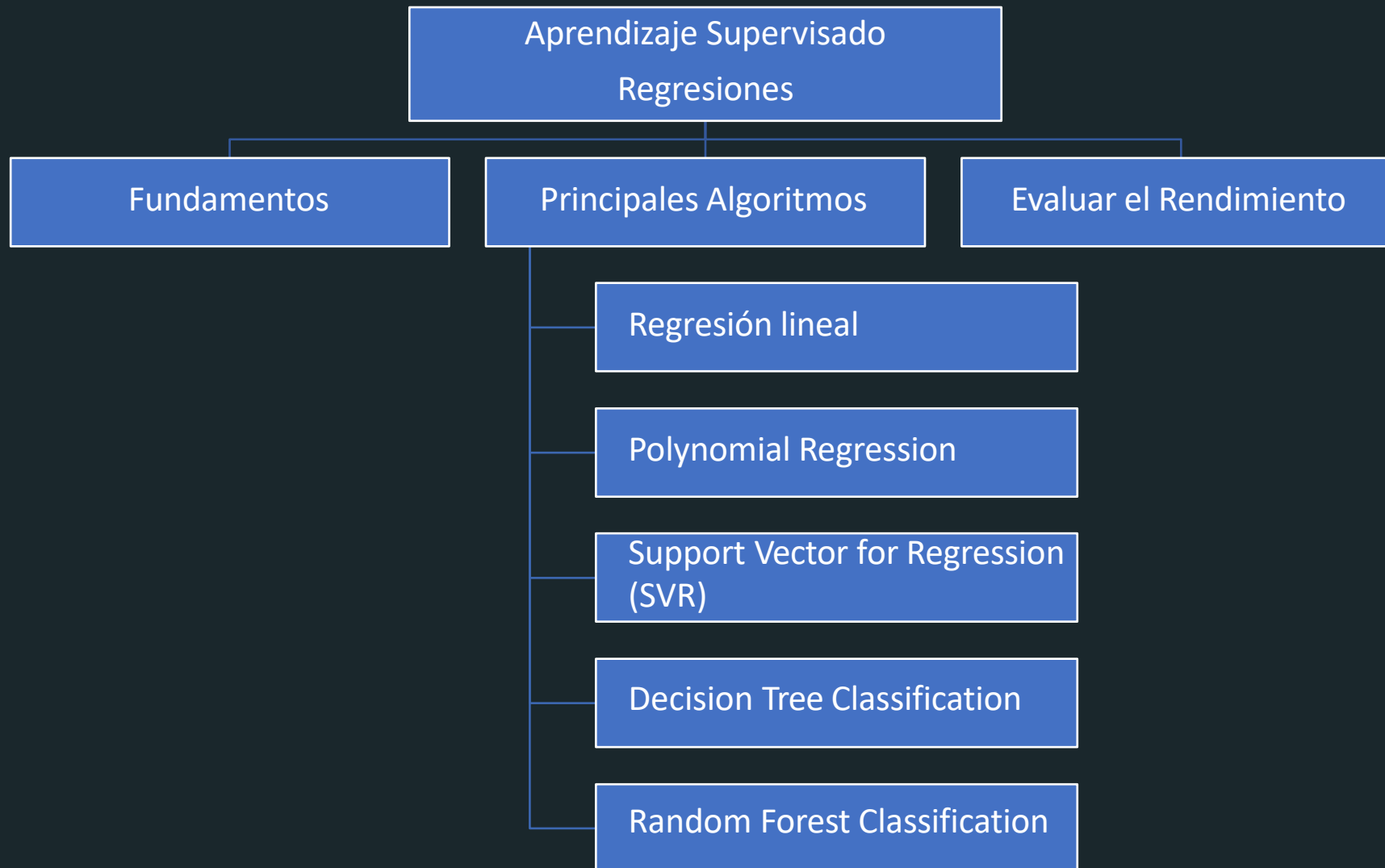


Aprendizaje Supervisado - Regresión

Fundamentos, Principales Algoritmos, Evaluar Rendimiento



Red de Cointenidos:



Fundamentos: Regresiones

- Los modelos de regresión (tanto lineal como no lineal) se utilizan muchísimo para predecir valores numéricos como por ejemplo el sueldo. Si nuestra variable independiente es tiempo entonces podemos hacer predicciones de valores futuros, sin embargo nuestro modelo puede predecir también valores desconocidos del presente. Las técnicas de Regresión son muy variadas, desde la regresión lineal hasta la SVR o la Regresión con Bosques Aleatorios.
- En esta parte, vamos a entender e implementar los siguientes modelos de Regresión dentro del Machine Learning:
 - 1.Simple Linear Regression
 - 2.Multiple Linear Regression
 - 3.Polynomial Regression
 - 4.Support Vector for Regression (SVR)
 - 5.Decision Tree Classification
 - 6.Random Forest Classification

Fundamentos: Que es una regresión?

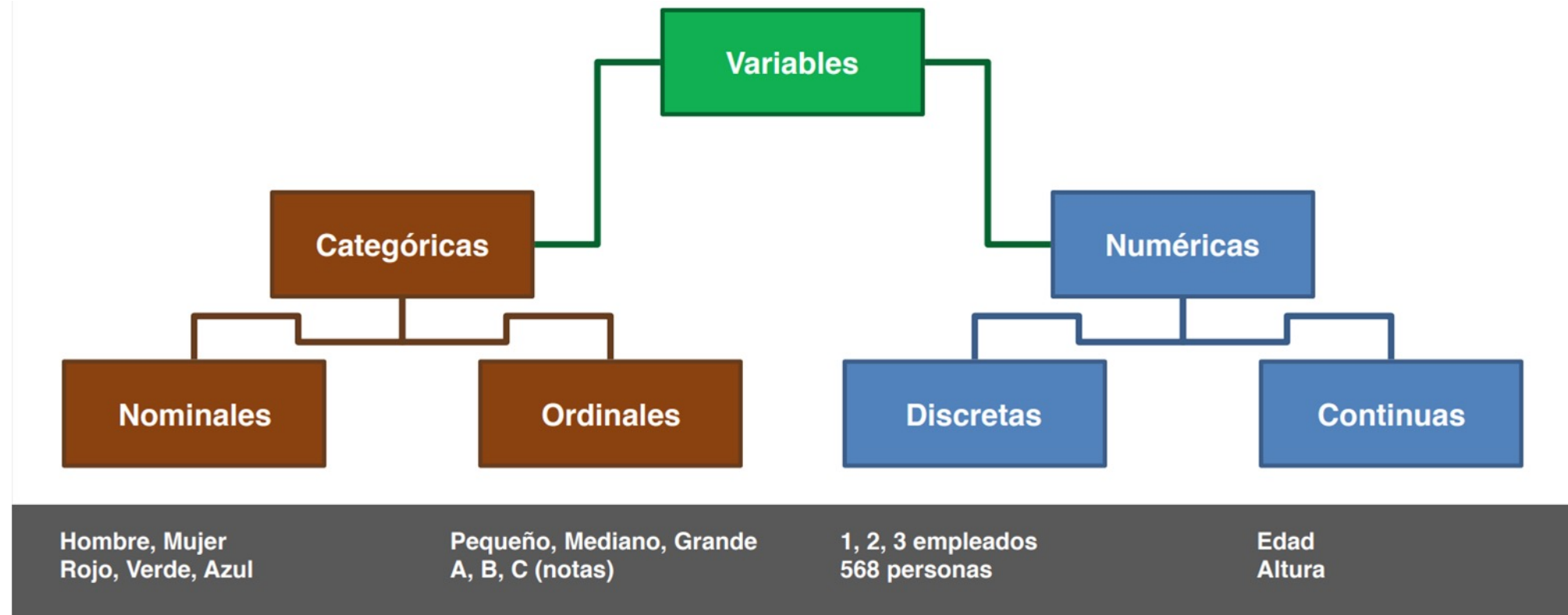
En estadística, se llama análisis de la regresión al proceso estadístico de estimar las relaciones que existen entre variables.

...

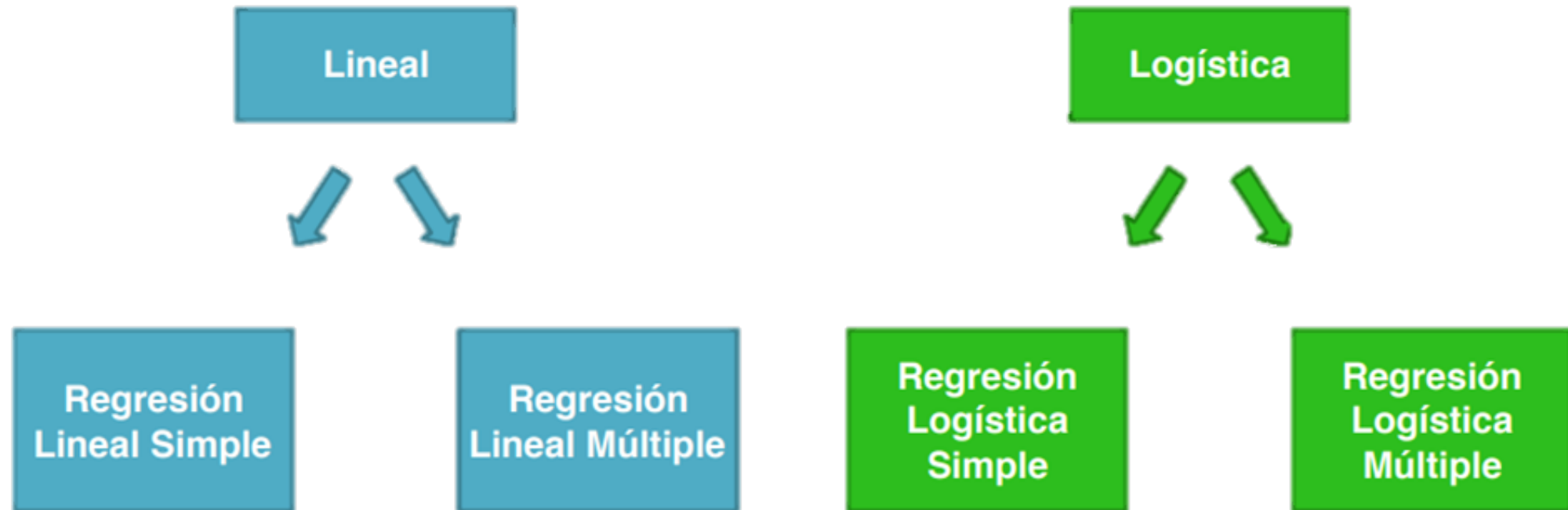
Se centra en estudiar las relaciones entre una variable dependiente de una o más variables independientes.

-Wikipedia

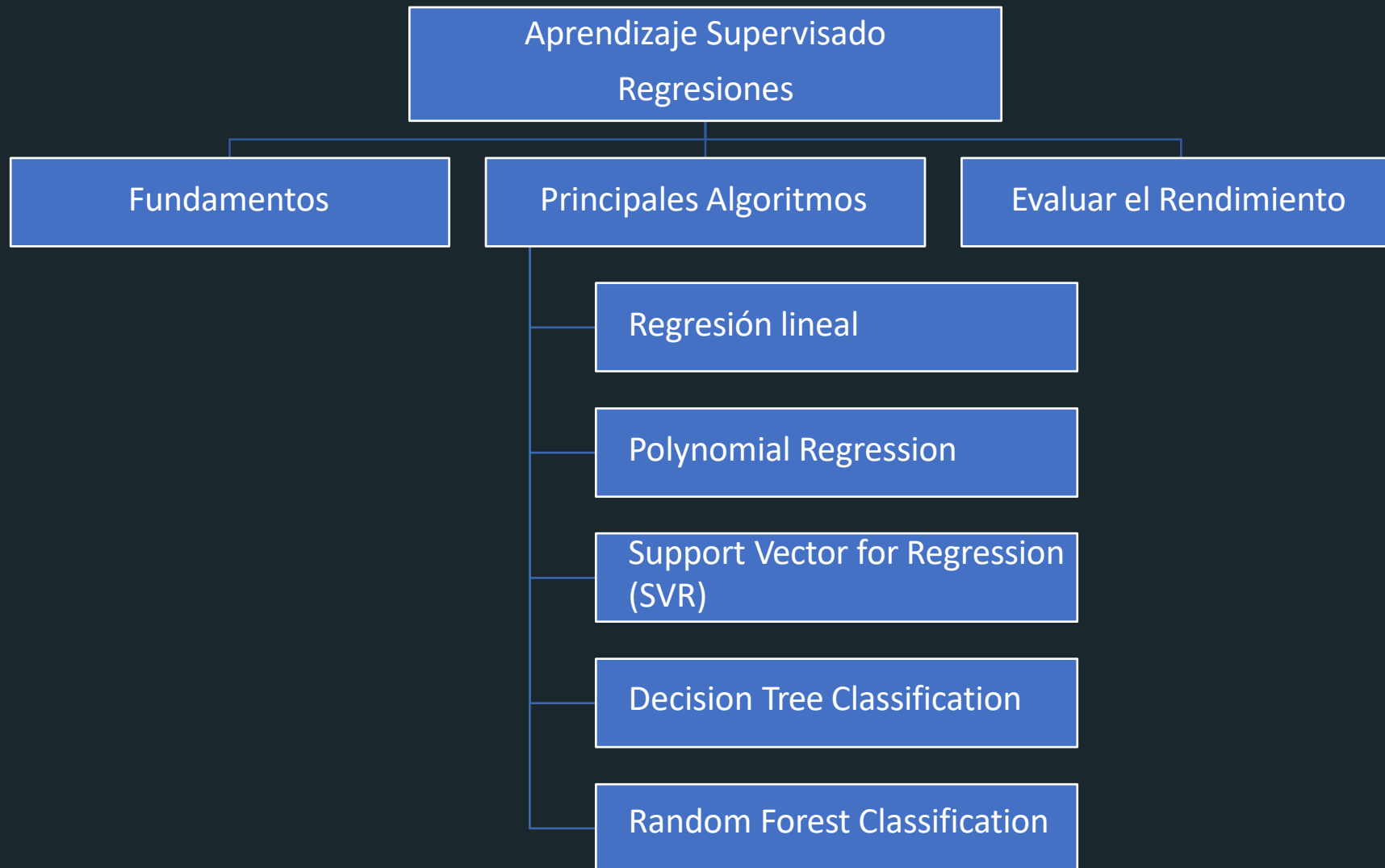
Fundamentos: Tipos de Variables



Fundamentos: Tipos de Regresiones



Red de Contenidos:



Regresión Lineal

Regresiones Lineales

Regresión Lineal Simple

$$y = b_0 + b_1 * x_1$$

Diagram illustrating the components of the Simple Linear Regression equation:

- Variable Dependiente (VD)**: Points to y .
- Constante**: Points to b_0 .
- Coeficiente**: Points to b_1 .
- Variable Independiente (VI)**: Points to x_1 .

Regresión Lineal Múltiple

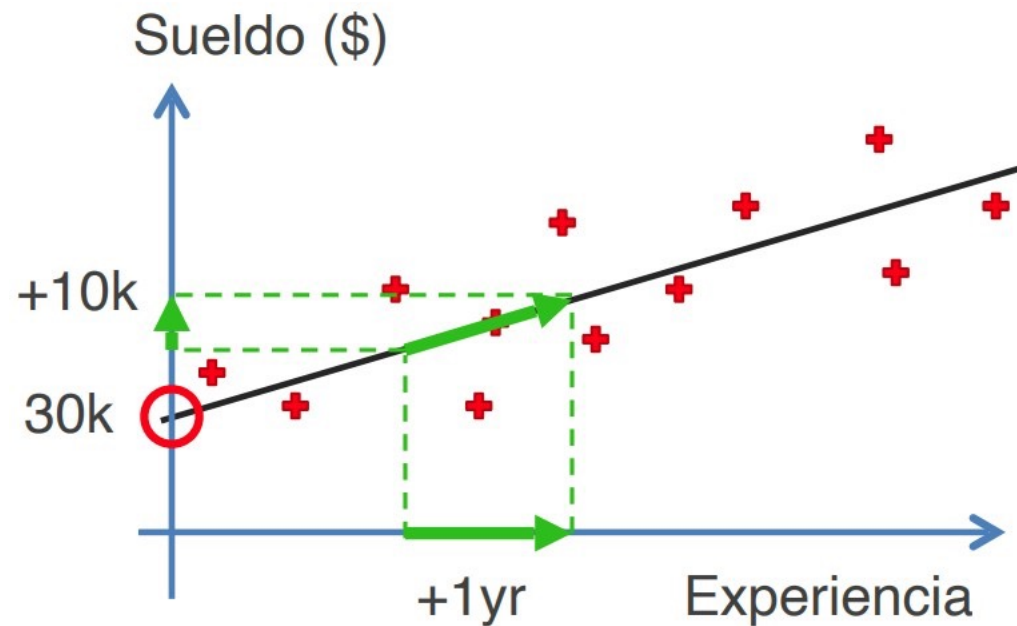
$$y = b_0 + b_1 * x_1 + b_2 * x_2 + \dots + b_n * x_n$$

Diagram illustrating the components of the Multiple Linear Regression equation:

- Variable Dependiente (VD)**: Points to y .
- Constante**: Points to b_0 .
- Coeficientes**: Points to the set of coefficients b_1, b_2, \dots, b_n .
- Variables Independientes (VIs)**: Points to the set of independent variables x_1, x_2, \dots, x_n .

Regresiones Lineales

Regresión Lineal Simple



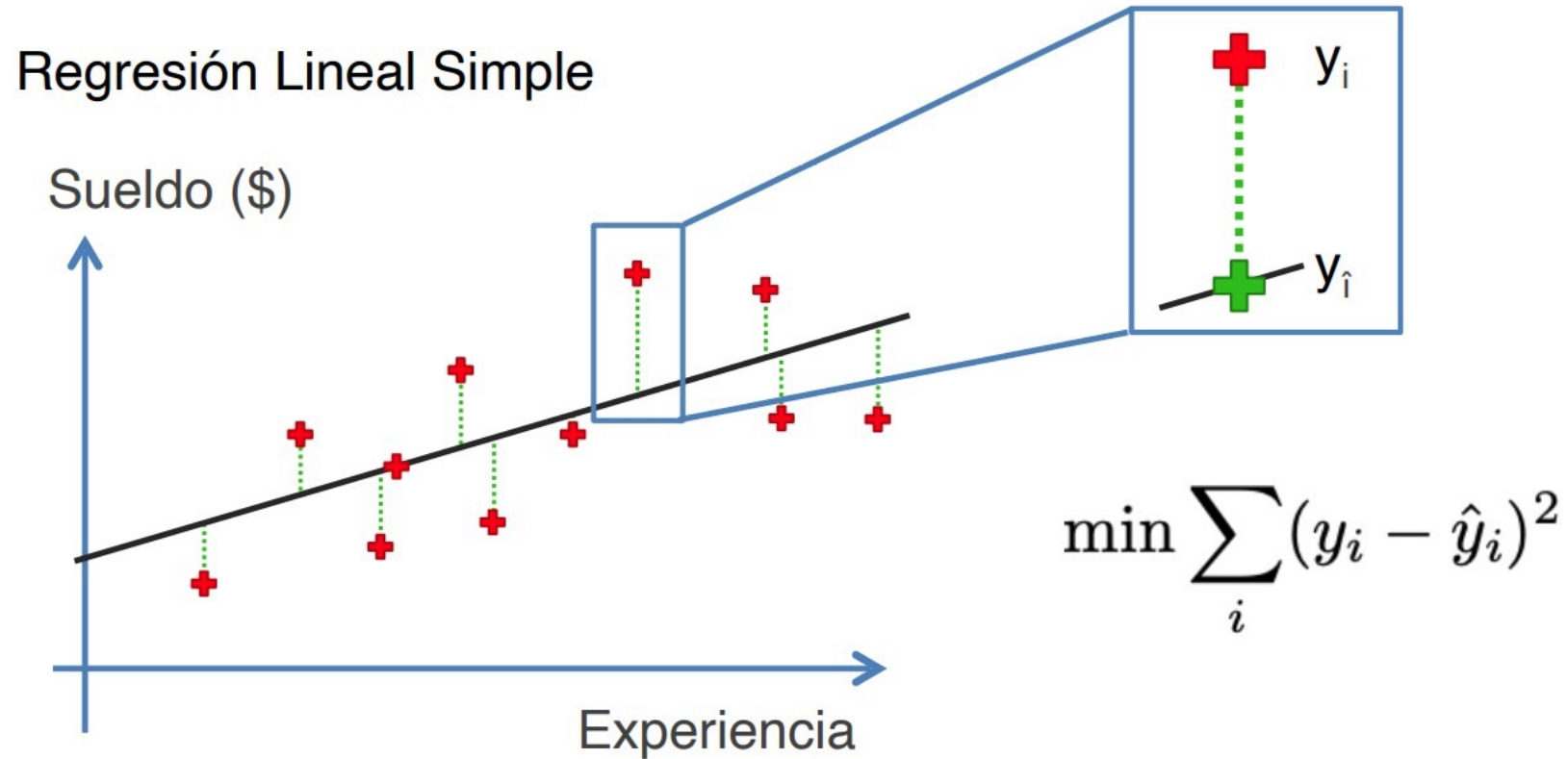
$$y = b_0 + b_1 * x$$



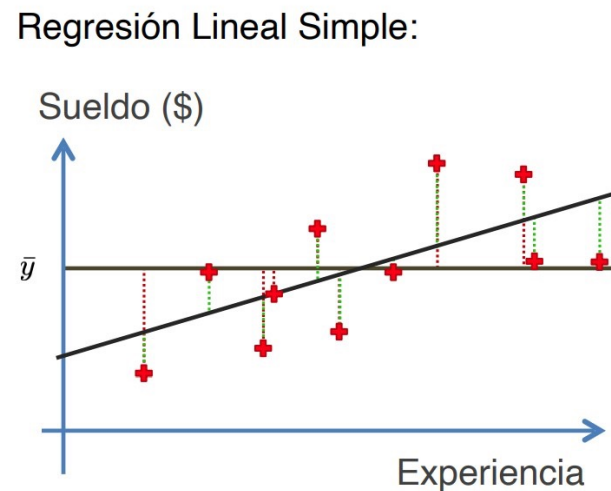
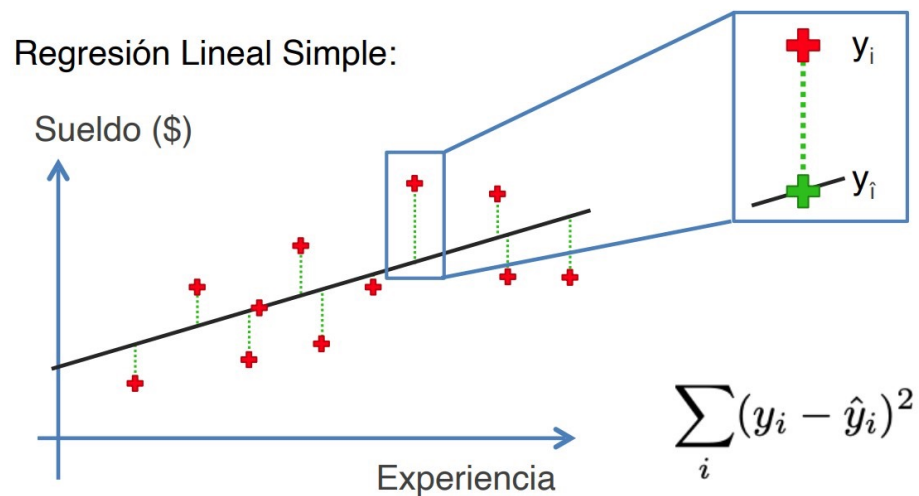
$$\text{Sueldo} = b_0 + b_1 * \text{Experiencia}$$

Métodos de Mínimos Cuadrados

Regresiones Lineales



Regresiones Lineales: R cuadrado



$$SS_{\text{res}} = \sum_i (y_i - \hat{y}_i)^2$$
$$SS_{\text{tot}} = \sum_i (y_i - \bar{y})^2$$

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}}$$

Mientras el R cuadrado tienda a 1 es mejor

Regresiones Lineales: Múltiples

The diagram illustrates the components of the multiple linear regression equation: $y = b_0 + b_1 * x_1 + b_2 * x_2 + \dots + b_n * x_n$. Green arrows point from descriptive labels to specific parts of the equation: 'Variable Dependiente (VD)' points to 'y'; 'Variables Independientes (VIs)' points to the set of x variables; 'Constante' points to b_0 ; and 'Coeficientes' points to the set of b coefficients.

Variable Dependiente (VD)

Variables Independientes (VIs)

$$y = b_0 + b_1 * x_1 + b_2 * x_2 + \dots + b_n * x_n$$

Constante

Coeficientes

Regresión lineal Múltiple

- Restricciones de la regresión lineal:
 1. Linealidad
 2. Homocedasticidad
 3. Normalidad multivariable
 4. Independencia de los errores
 5. Ausencia de multicolinealidad

Regresión Lineal Múltiple

- Variables Dummy

					Variables Dummy	
Profit	R&D Spend	Admin	Marketing	State	New York	California
192,261.83	165,349.20	136,897.80	471,784.10	New York	1	0
191,792.06	162,597.70	151,377.59	443,898.53	California	0	1
191,050.39	153,441.51	101,145.55	407,934.54	California	0	1
182,901.99	144,372.41	118,671.85	383,199.62	New York	1	0
166,187.94	142,107.34	91,391.77	366,168.42	California	0	1

$$y = b_0 + b_1*x_1 + b_2*x_2 + b_3*x_3 + b_4*D_1 + b_5*D_2$$



$$D_1 = 1 - D_2$$

El algoritmo no podrá diferenciar entre estos 3 componentes

Si tenemos n categorías de una variable, entonces debemos quedarnos con $n-1$ variables dummies

Profit	R&D Spend	Admin	Marketing	State	New York	California
192,261.83	165,349.20	136,897.80	471,784.10	New York	1	0
191,792.06	162,597.70	151,377.59	443,898.53	California	0	1
191,050.39	153,441.51	101,145.55	407,934.54	California	0	1
182,901.99	144,372.41	118,671.85	383,199.62	New York	1	0
166,187.94	142,107.34	91,391.77	366,168.42	California	0	1

$$y = b_0 + b_1*x_1 + b_2*x_2 + b_3*x_3 + b_4*D_1$$



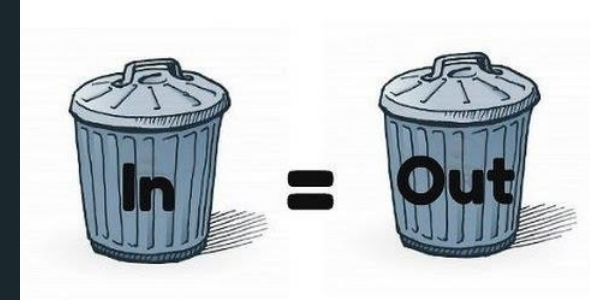
Regresión Lineal Múltiple: Construcción del Modelo

- En la regresión múltiple, intervienen muchas variables queriendo aportar al modelo, sin embargo no todas son útiles. Algunas no aportarán y otras harán la misma tarea.
- El trabajo del modelador es decidir cuales son mejor conservar.
- Es mejor un modelo con pocas variables que prediga mejor o igual que otro con mas variables.



Regresión Lineal Múltiple: Construcción del Modelo

1. No por añadir mas variables vamos a tener mas información.
2. Si crece el numero de variables será mas difícil comprenderlo y explicarlo
3. Existen 5 métodos para quedarnos con las mejores variables



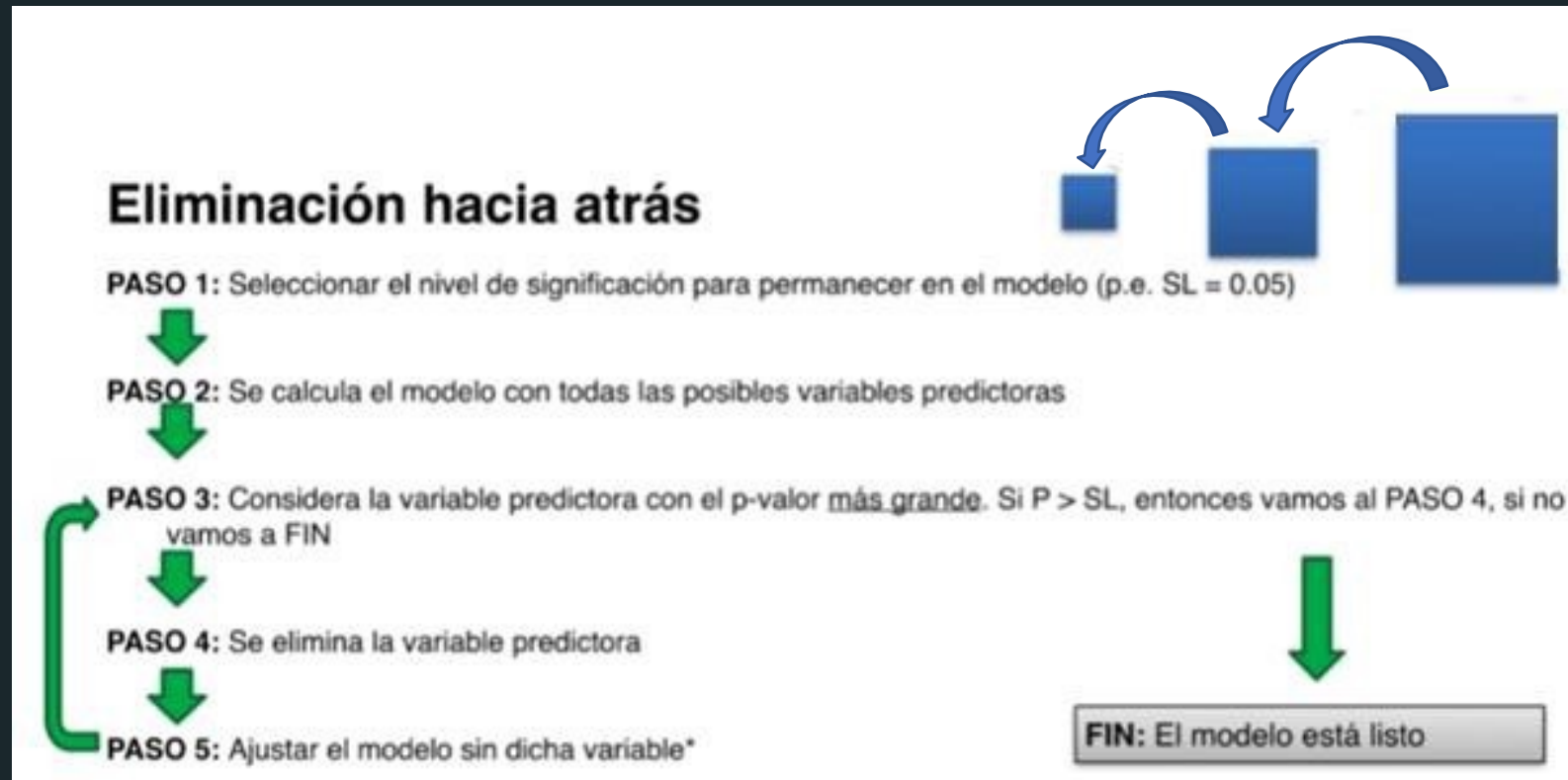
Regresión Lineal Múltiple: Construcción del Modelo

5 métodos de construir modelos:

- Exhaustivo (All-in)
 - Eliminación hacia atrás
 - Selección hacia adelante
 - Eliminación Bidireccional
 - Comparación de scores
- Regresión paso a paso

Regresión Lineal Múltiple: Construcción del Modelo

- Backward Elimination



Regresión Polinómica

Regresión Polinómica

Regresión
Lineal Simple

$$y = b_0 + b_1x_1$$

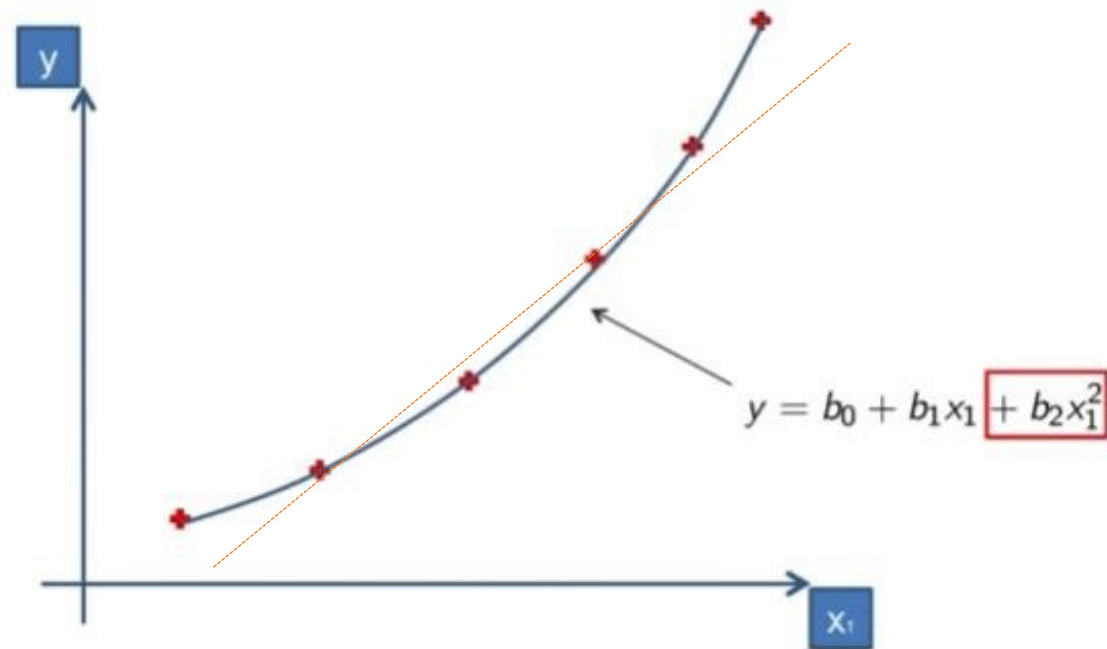
Regresión
Lineal Múltiple

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n$$

Regresión
Lineal
Polinómica

$$y = b_0 + b_1x_1 + b_2x_1^2 + \dots + b_nx_1^n$$

Regresión Polinómica

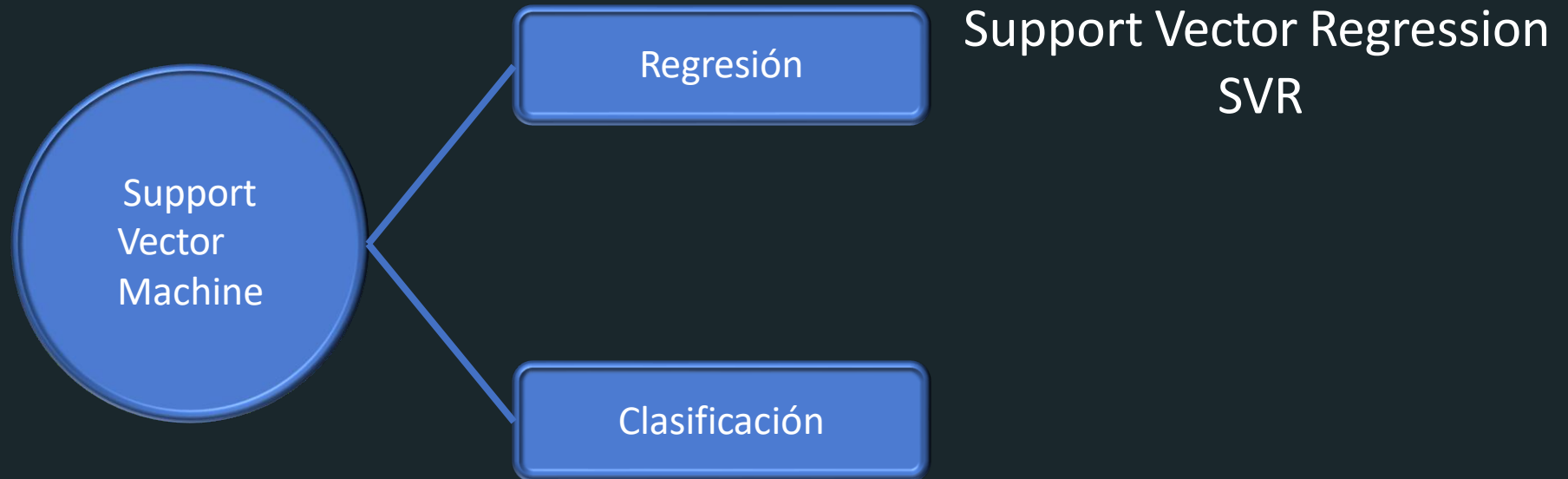


Support Vector Machine

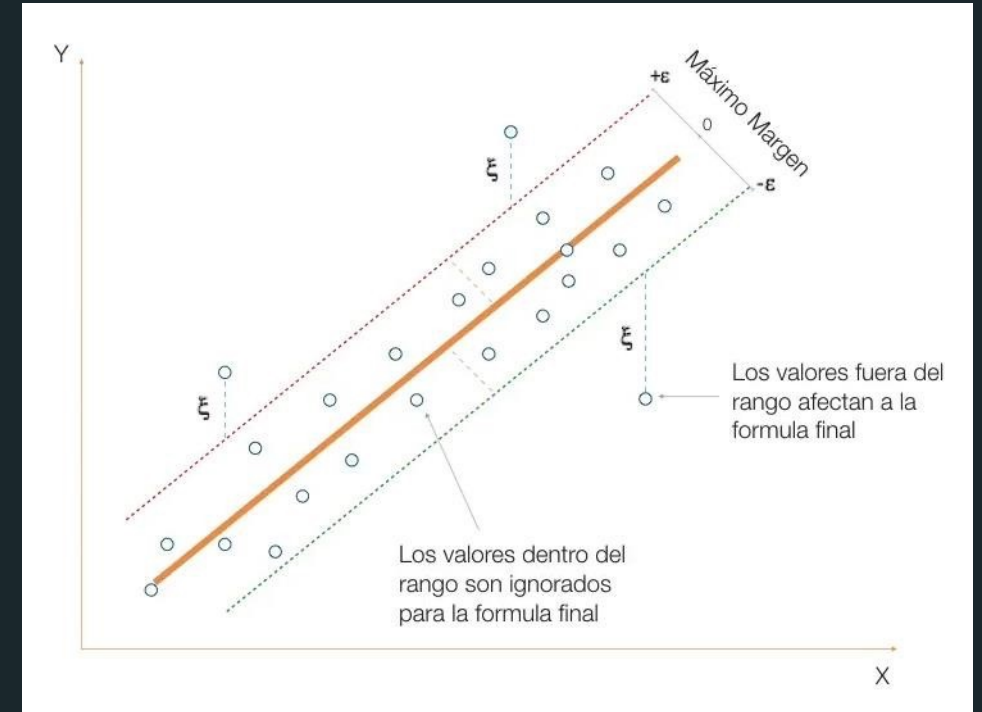
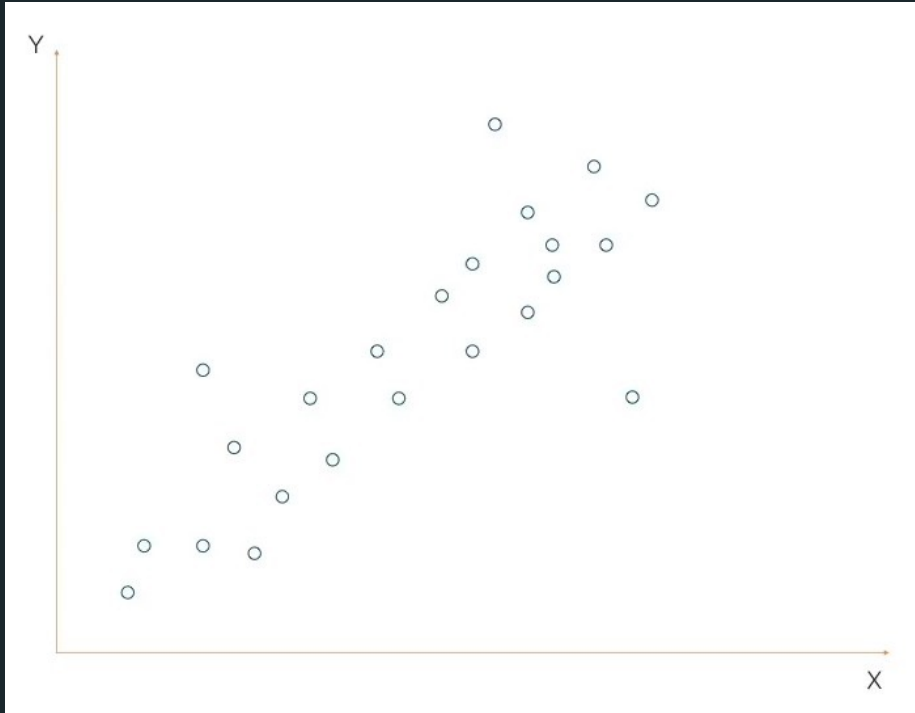
Regression (SRV)

Maquinas de Soporte Vectorial

Las máquinas de vectores de soporte o máquinas de vector soporte son un conjunto de algoritmos de aprendizaje supervisado desarrollados por Vladimir Vapnik y su equipo en los laboratorios AT&T



Maquinas de Soporte Vectorial

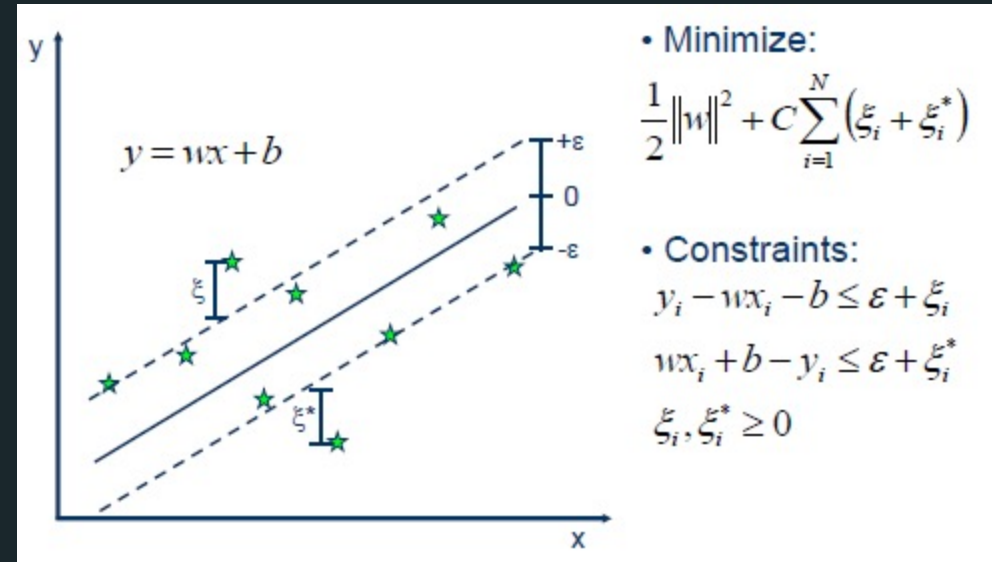


Maquinas de Soporte Vectorial

Se basa en buscar la curva con su respectivo margen máximo, que modele la tendencia de los datos y, según ella predecir cualquier otro dato en el futuro.

Todo lo que se considera fuera de ese rango es considerado error

Esta curva siempre viene acompañada con un rango (máximo margen), tanto del lado positivo como en el negativo, el cual tiene el mismo comportamiento o forma de la curva.



Maquinas de Soporte Vectorial

Que pasa si no es lineal?

