



# REGRESIÓN

APRENDIZAJE AUTOMÁTICO - CEIOT - FIUBA

Dr. Ing. Facundo Adrián Lucianna

# REPASO CLASE ANTERIOR

- Definición de Machine Learning y Big Data
- Tipos de aprendizaje:
  - Aprendizaje supervisado: Regresión y clasificación
  - Aprendizaje no supervisado: Agrupamiento y reducción dimensional
  - Aprendizaje profundo: Redes neuronales

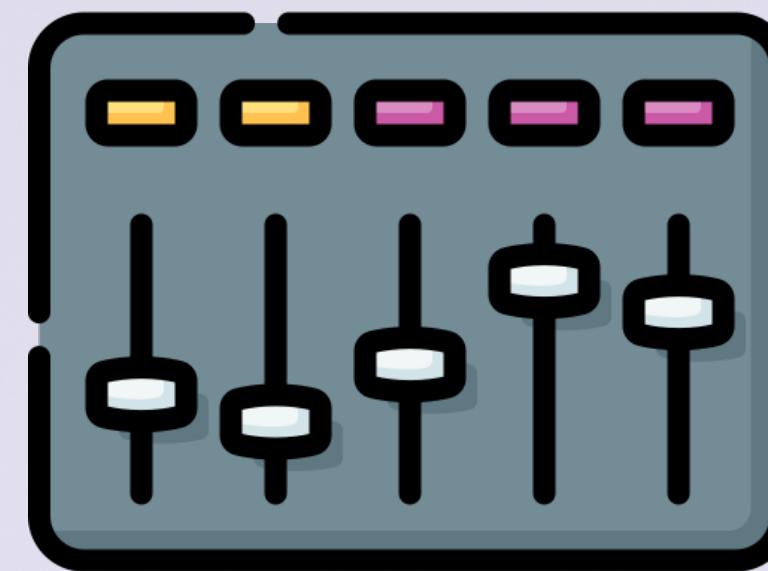
# REPASO CLASE ANTERIOR

Observation →

Features					Label
Position	Experience	Skill	Country	City	Salary (\$)
Developer	0	1	USA	New York	103100
Developer	1	1	USA	New York	104900
Developer	2	1	USA	New York	106800
Developer	3	1	USA	New York	108700
Developer	4	1	USA	New York	110400
Developer	5	1	USA	New York	112300
Developer	6	1	USA	New York	116100
Developer	7	1	USA	New York	117800

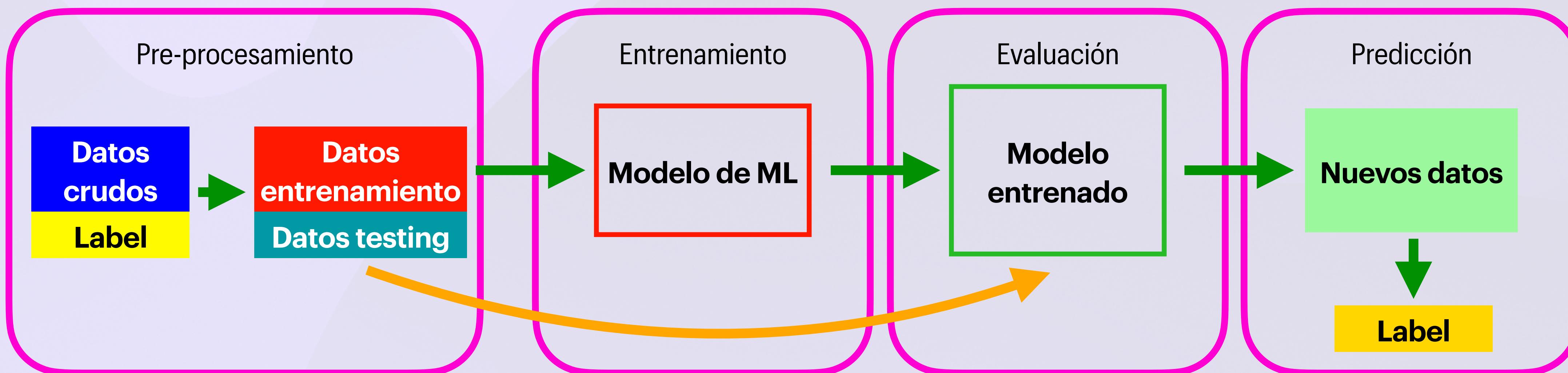
# REPASO CLASE ANTERIOR

- Los algoritmos de Machine Learning tienen parámetros “internos” que no dependen de los datos. Estos parámetros se llaman **hiperparámetros**. Por ejemplo, una red neuronal tiene como hiperparámetros la función de activación o la constante de entrenamiento.



- Llamamos **generalización** a la capacidad del modelo de hacer predicciones nuevas utilizando datos nuevos.

# REPASO CLASE ANTERIOR





# TIPO DE VARIABLES

# TIPOS DE VARIABLES

Recordemos que lo más importante en Data Science, y en consecuencia en Machine Learning, lo más importante son los

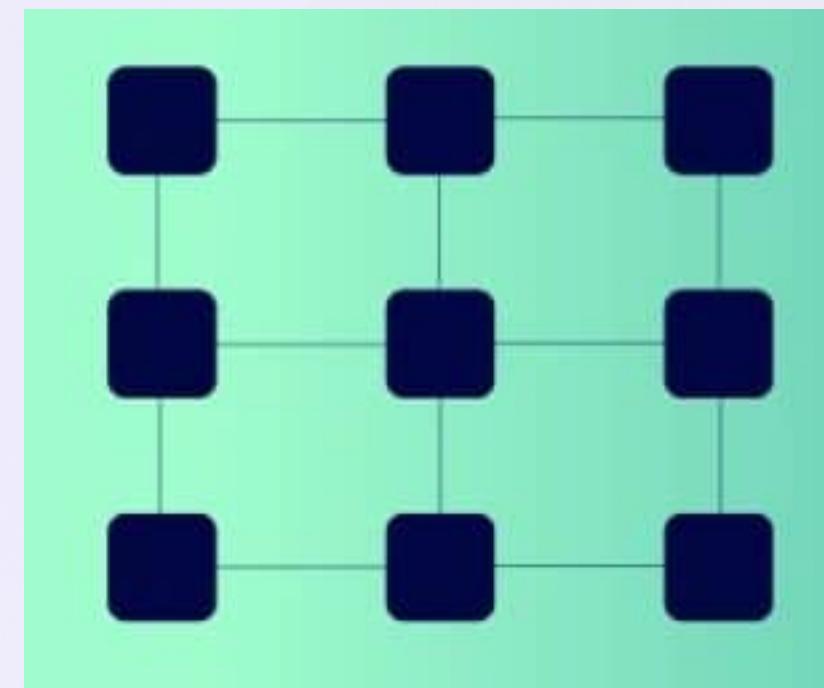
**datos**

Nos permite describir un objeto al que podemos llamar entidad

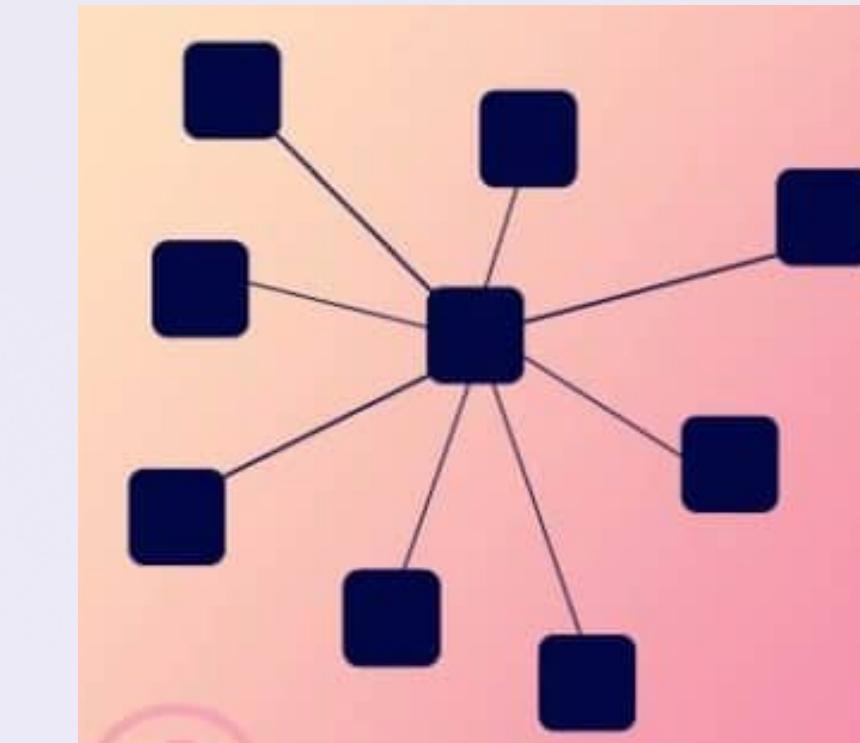
# TIPOS DE VARIABLES

datos

Estructurados



No estructurados



# TIPOS DE VARIABLES

Un **dato** nos permite describir una entidad mediante:

- **Atributos**  
≡
- **Variables**  
≡
- **Features**

El desafío en Machine Learning es elegir las variables correctas. Hay muchas técnicas con la finalidad de permitirnos esto, pero la mayoría de las veces resulta en un proceso iterativo de **prueba y error**.

# TIPOS DE VARIABLES

Una parte importante, es identificar el tipo de variable



# VARIABLES NUMÉRICAS

Son aquellas que representan números y con ellas se pueden realizar operaciones aritméticas.

- **Discretas:** Son números enteros, cosas que se pueden contar.  
1, 2, 3 empleados, 568 personas.
- **Continuas:** Números reales. El valor dado a una observación para una variable continua puede incluir valores tan pequeños como lo permita el instrumento de medición o la representación numérica.  
Altura, peso, costo, precio...

# VARIABLES CATEGÓRICAS

Son aquellas que representan números y con ellas se pueden realizar operaciones aritméticas.

- **Nominal:** Valores que toman corresponden a nombres de categorías, clases o estados de las cosas.  
Estado civil (soltero, casado, divorciado), Spam en e-mails (Binario, es spam o no). Tipo de cerveza (Ale, Pale, Stout, etc.)
- **Ordinal:** Similar al nominal, con la diferencia de poder aplicar un orden sobre estas categorías.  
Estado de satisfacción: Me disgusta mucho, me disgusta, neutro, me gusta, me gusta mucho.  
No tiene que haber una equidistancia entre las opciones.

The background features a series of overlapping, wavy shapes in shades of purple and dark blue, creating a sense of depth and motion.

# REGRESIÓN

# REGRESIÓN

Tal como nos indica la Wikipedia, se llama análisis 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.

Es importante notar que estamos armando un modelo puramente empírico. Es decir, nos basamos 100% en los datos medidos. En contraste con los modelos basados en propiedades fundamentales.

# REGRESIÓN VS. CLASIFICACIÓN

Regresión y clasificación son problemas muy similares entre sí. En ambos buscamos predecir una variable, la diferencia radica en que **regresión** predice una variable **numérica** y **clasificación** una **categórica**.

# REGRESIÓN



# CLASIFICACIÓN



# REGRESIÓN LINEAL

# REGRESIÓN LINEAL

El modelo de regresión lineal más simple es el que involucra una combinación lineal de las variables de entradas:

$$y(\hat{x}) = w_0 + w_1x_1 + \dots + w_dx_d$$

$$\hat{x} = (x_1, x_2, \dots, x_D)$$

Son los features de nuestras observaciones. Son todas variables numéricas

$$w_0, w_1, \dots, w_D$$

Son los coeficientes del modelo. Son números reales. Cuanto mas cerca de cero, la variable dependiente depende menos del feature que multiplica.

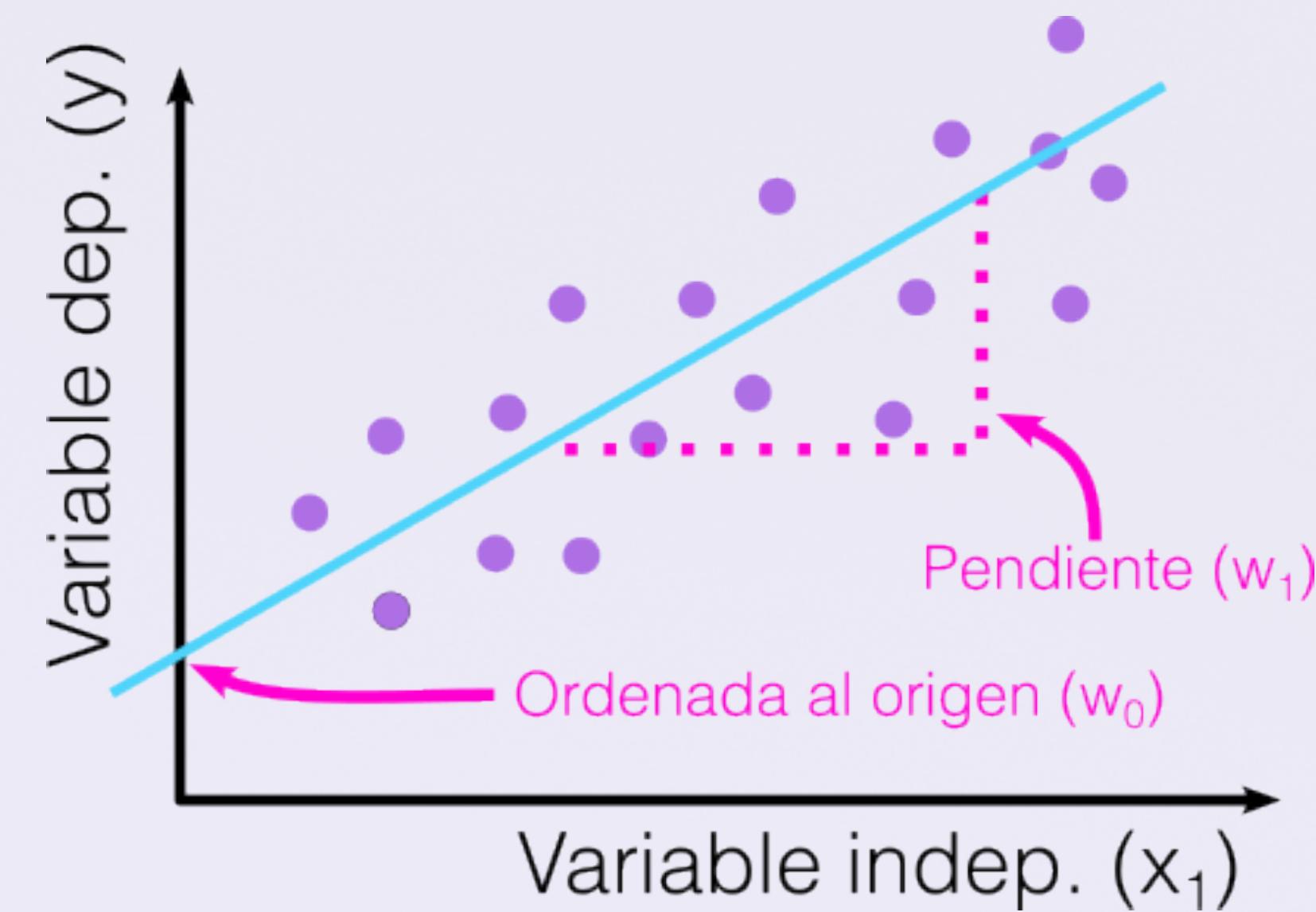
$$y$$

Es la predicción del modelo. Es con quien comparamos con el Label de la observación

# REGRESIÓN LINEAL

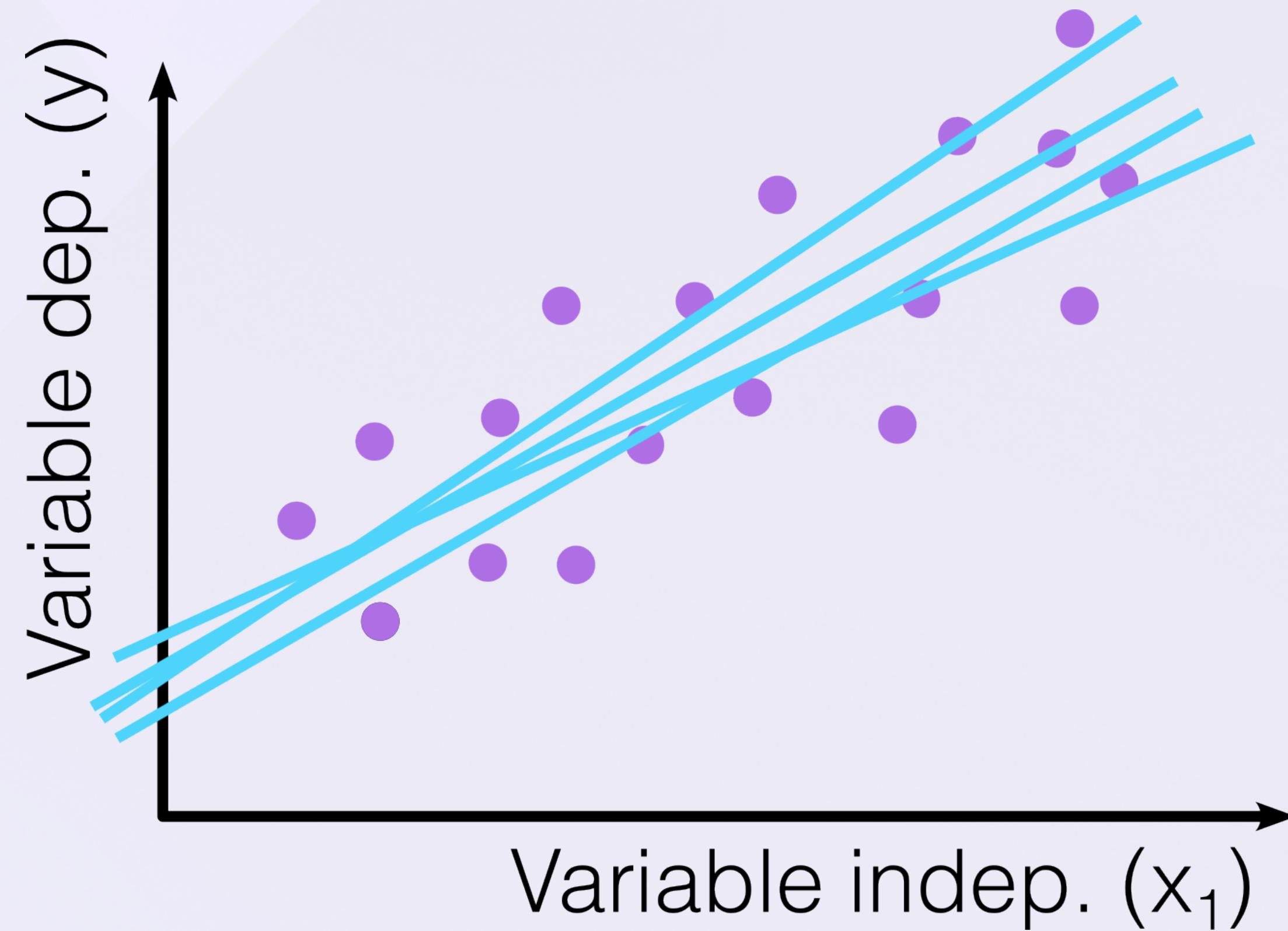
Vamos al caso más sencillo, la regresión lineal de una sola variable independiente:

$$y(x_1) = w_0 + w_1 x_1$$



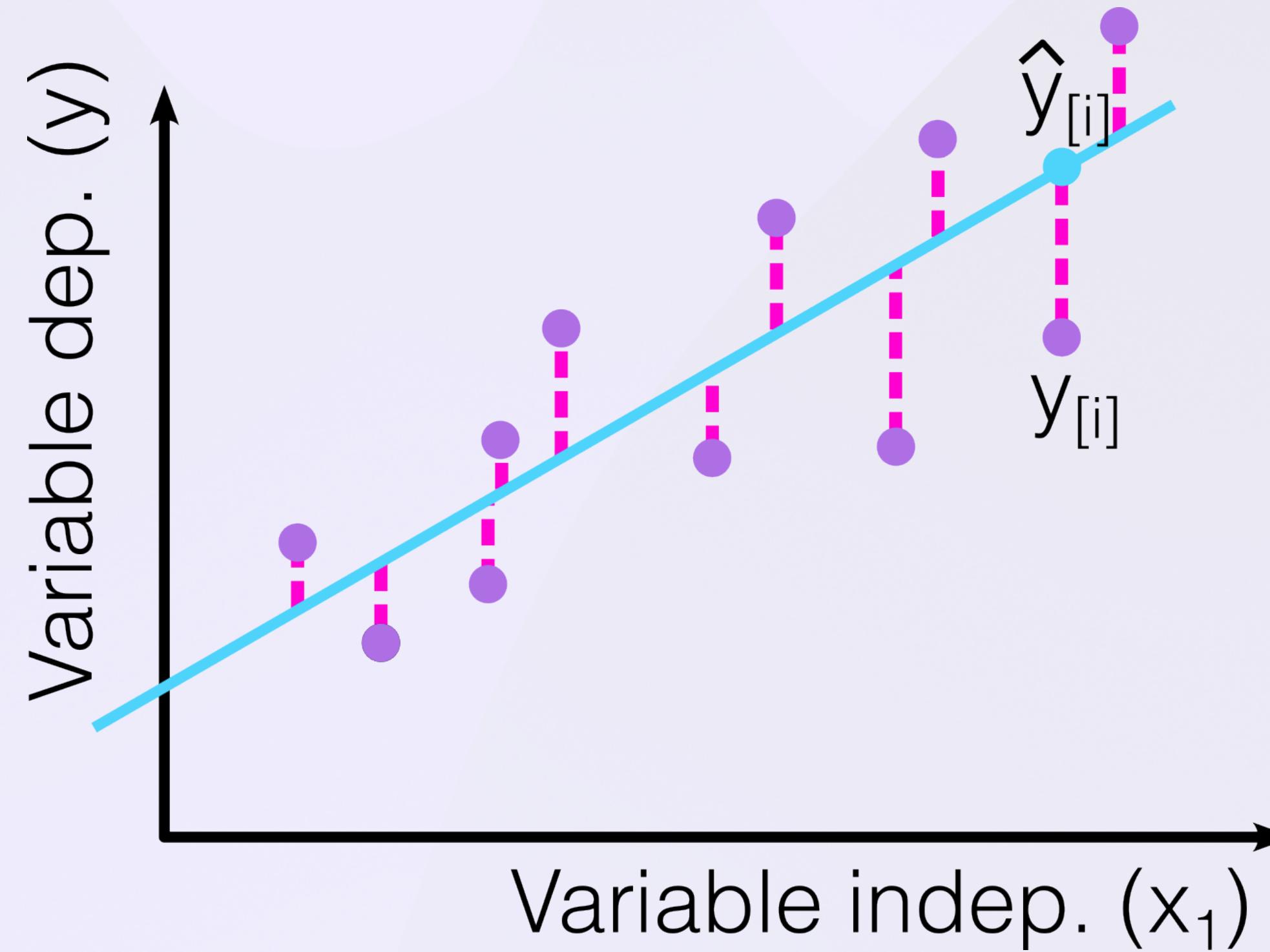
# REGRESIÓN LINEAL

Ahora cuál recta?



# REGRESIÓN LINEAL

Para encontrarla, medimos la distancia entre la recta y cada punto, que llamamos **residuos**.

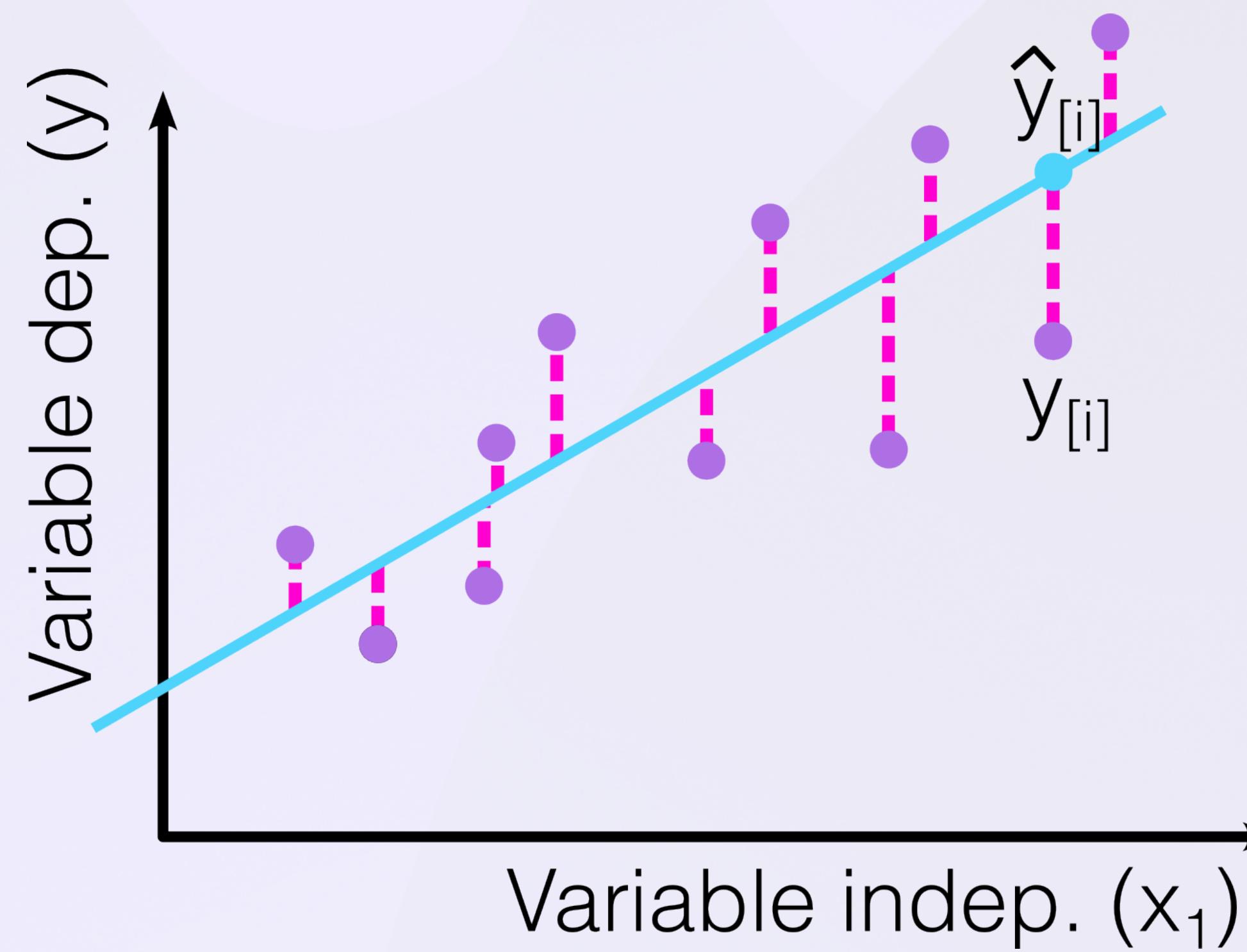


$$e_{[i]} = y_{[i]} - \hat{y}_{[i]}$$

$$y_{[i]} = w_0 + w_1 x_{1[i]} + e_{[i]}$$

# REGRESIÓN LINEAL

Buscamos minimizar el valor de los **residuos**. Para lograr esto, lo hacemos minimizando la suma de los cuadrados de los residuos



$$S_R = \sum_{i=0}^{N-1} e_{[i]}^2 = \sum_{i=0}^{N-1} (y_{[i]} - w_0 - w_1 x_{1[i]})^2$$

$$\min(S_R) = \min \left( \sum_{i=0}^{N-1} e_{[i]}^2 \right)$$

Para minimizar, solo podemos tocar los coeficientes. Lo que hacemos es ir por el gradiente

$$\frac{\partial S_R}{\partial w_0} = 0$$

$$\frac{\partial S_R}{\partial w_1} = 0$$

# REGRESIÓN LINEAL

Trabajando con las derivadas parciales, se llegan a lo que se llama ecuaciones normales, lo que nos da un sistema de ecuaciones que podemos resolver.

Lo importante de esto es que no hay un algoritmo iterativo que se debe dejar procesar. Sino que se resuelve con métodos numéricos de resolución de sistemas de ecuaciones.

# AJUSTE

Cómo hacemos para medir que tan bien se ajusta una regresión a nuestros datos?

Si medimos la varianza de la variable dependiente de los datos que hacemos la regresión:

$$S_T = \sum_{i=0}^{N-1} (y_{[i]} - \bar{y})^2$$

Esta varianza, la podemos separar en dos parte, una parte que es dada por el modelo y una parte que no.

$$S_T = S_R + S_E$$

$$S_T = \sum_{i=0}^{N-1} (\hat{y}_{[i]} - \bar{y})^2 + \sum_{i=0}^{N-1} (y_{[i]} - \hat{y}_{[i]})^2$$

# AJUSTE

Cómo hacemos para medir que tan bien se ajusta una regresión a nuestros datos?

Si medimos la varianza de la variable dependiente de los datos que hacemos la regresión:

$$S_T = \sum_{i=0}^{N-1} (y_{[i]} - \bar{y})^2$$

Esta varianza, la podemos separar en dos parte, una parte que es dada por el modelo y una parte que no.

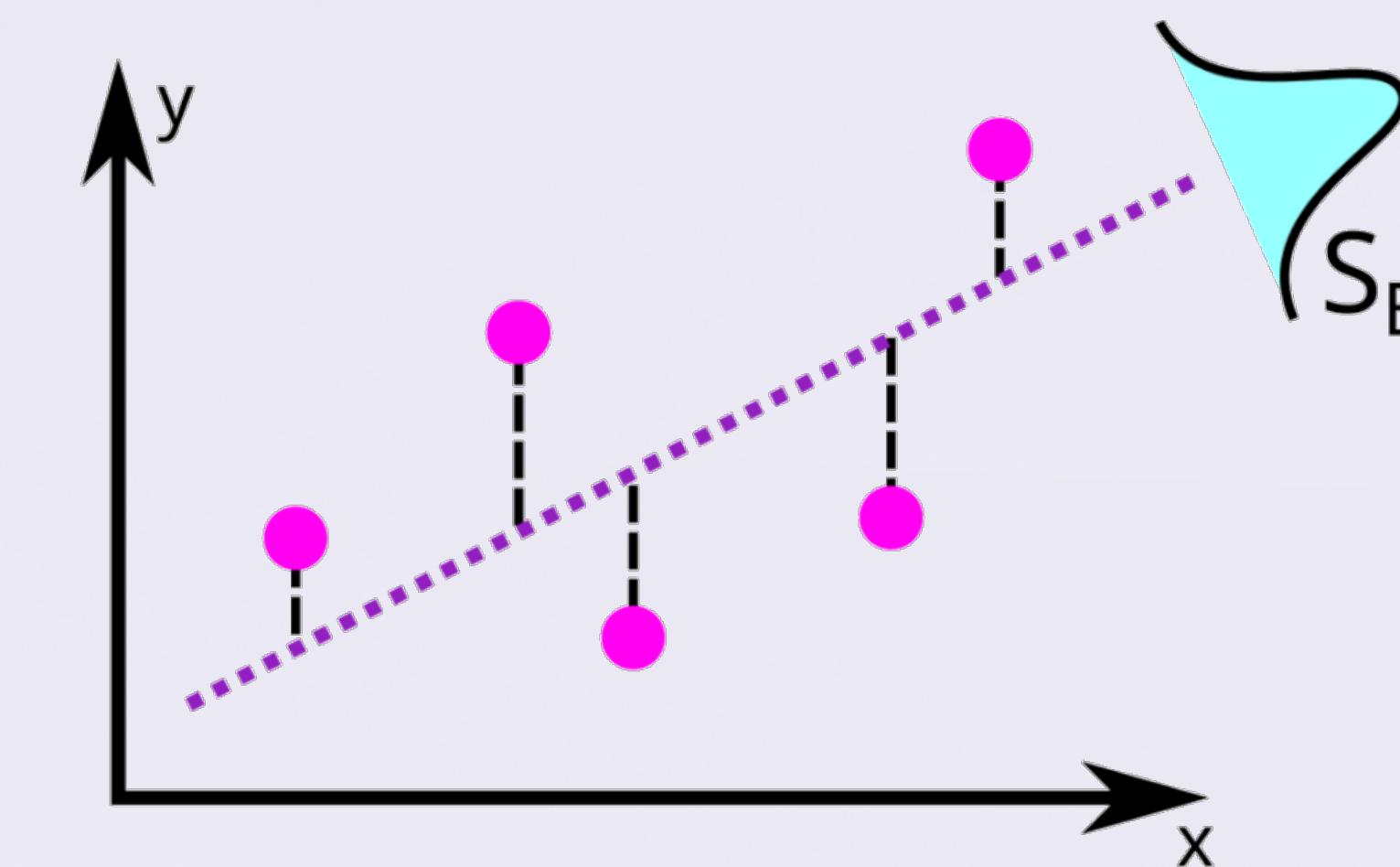
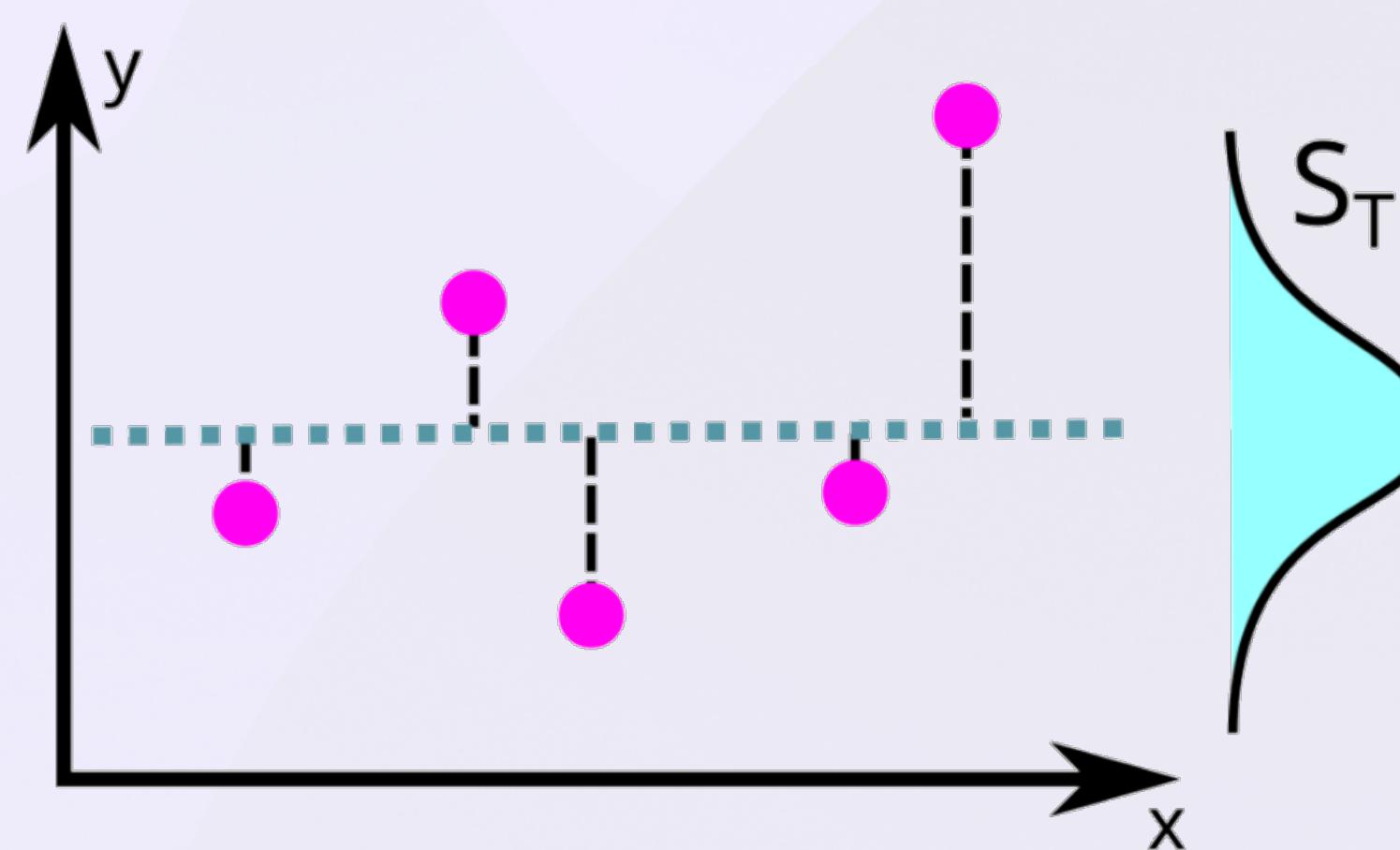
$$S_T = S_R + S_E$$

$$S_T = \sum_{i=0}^{N-1} (\hat{y}_{[i]} - \bar{y})^2 + \sum_{i=0}^{N-1} (y_{[i]} - \hat{y}_{[i]})^2$$

Parte que explica el modelo

Parte que no (residuos)

# AJUSTE



# AJUSTE

Como métricas, podemos usar:

- El cálculo del desvío estándar del modelo:

$$s_E = \sqrt{\frac{S_E}{N - 2}}$$

- Coeficiente de Pearson (cuando mas cerca a 1 mejor ajusta):

$$R^2 = \frac{S_R}{S_T} = 1 - \frac{S_E}{S_T}$$

# AJUSTE

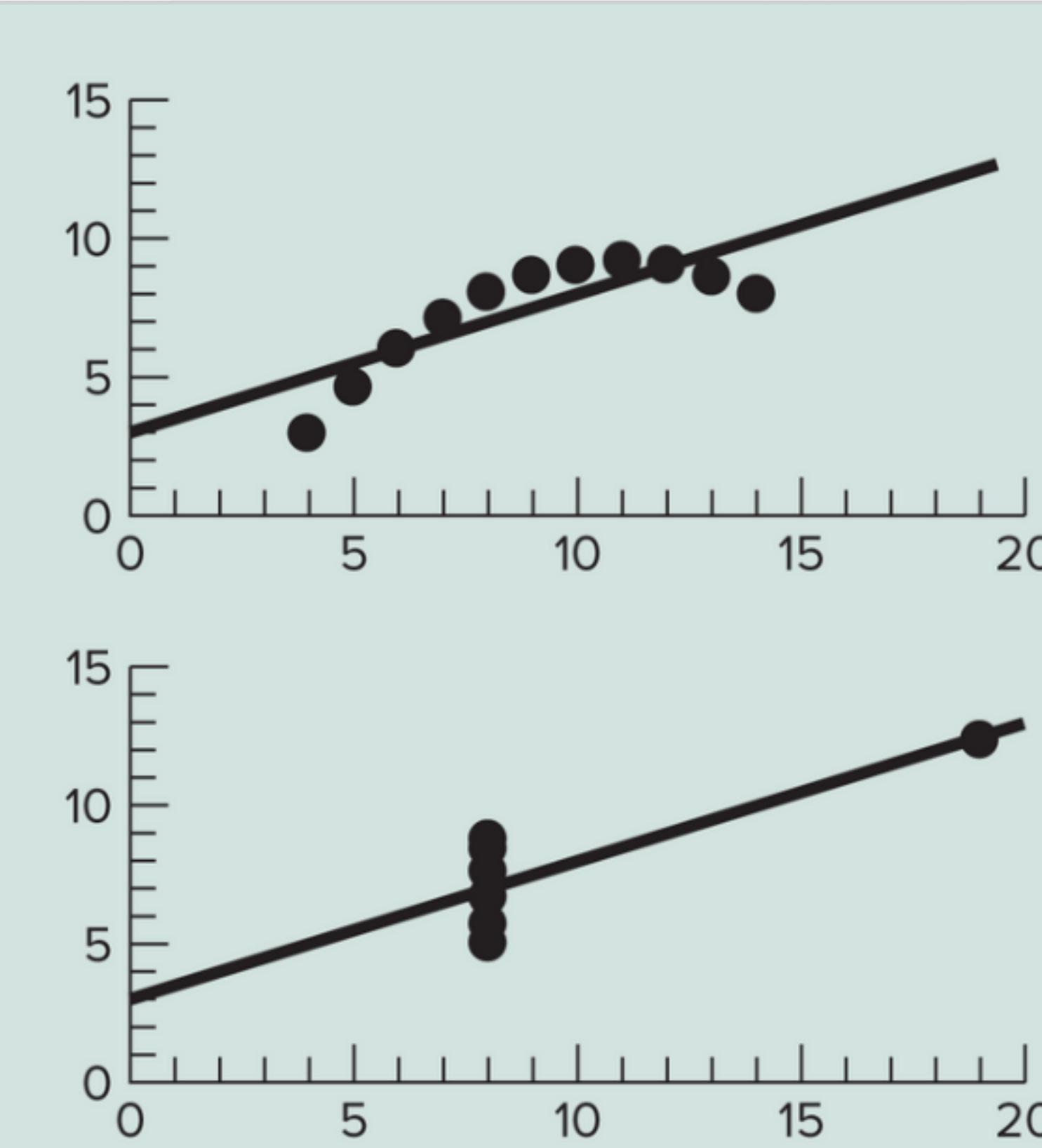
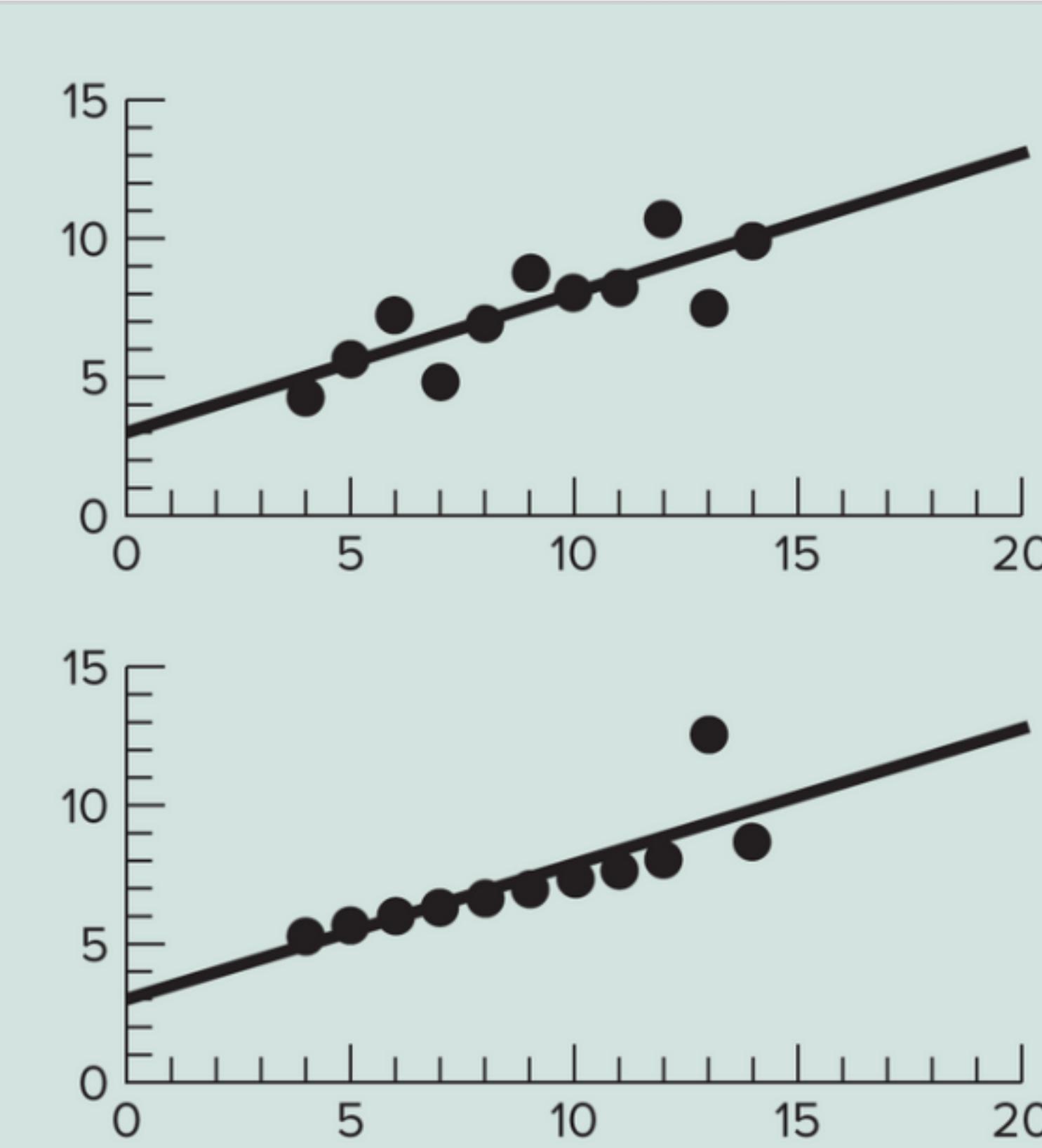
El valor de coeficiente de Pearson no es perfecto. Todo dependerá de que hace que genera el valor de residuo:

- Si la relación entre la variable independiente y dependiente es realmente lineal, el residuo  $e_{[i]} \sim N(0, \sigma)$ . Por lo que depende de como se tomaron los datos. Por ejemplo, en registro de temperatura, si el termómetro es de mala calidad, esperamos tener una mayor variación, por lo que el coeficiente podría ser bajo, pero sin ser culpa del modelo.
- Si la relación no es lineal, el valor del coeficiente de Pearson nos va a indicar que el modelo es malo.

# AJUSTE

El valor de coeficiente de determinación es el valor de residuo que genera el modelo.

- Si la relación entre los datos y el valor de residuo es lineal, el modelo es ajustado. Por ejemplo, si queremos tener una recta que genere el valor de residuo que genera el modelo, el modelo es ajustado.
- Si la relación entre los datos y el valor de residuo no es lineal, el modelo es mal ajustado.



que genera el modelo, el modelo es mal ajustado. Por ejemplo, si queremos tener una recta que genere el valor de residuo que genera el modelo, el modelo es mal ajustado.

# CONDICIONES PARA APLICAR REGRESIÓN

1. **Relación lineal:** Lógicamente, y esto muchas veces al aplicar el modelo buscamos validar.
2. **Features independientes:** Los features de entrada de la regresión deben ser independientes entre si.
3. **Homocedasticidad:** Es decir, el valor de  $S_E$  se mantiene igual en toda parte de la recta
4. **Errores independientes:** Los errores entre si no están correlacionados.

The background features a minimalist design with abstract, wavy shapes in shades of purple and dark blue. These shapes are layered and overlap, creating a sense of depth and movement across the entire frame.

**VAMOS A PRÁCTICAR UN POCO...**



# MÉTRICAS DE EVALUACIÓN

# MÉTRICAS DE EVALUACIÓN

Cuando armamos el modelo, al dataset lo separamos en un parte usada para entrenar el modelo y la parte de evaluación.

El conjunto de datos de prueba se utiliza para evaluar qué **tan bien se entrenó el algoritmo** con el conjunto de datos de entrenamiento. En los proyectos de ML, no podemos usar el conjunto de datos de entrenamiento en la etapa de prueba porque el algoritmo ya sabrá por adelantado el resultado esperado, que no es el objetivo.

El objetivo es utilizar el modelo para hacer nuevas predicciones, por lo que los datos de evaluación nos permite medir eso.

# MÉTRICAS DE EVALUACIÓN

Pero como evaluamos?

- **El coeficiente de Pearson ( $R^2$ )**. Aunque no es el mejor caso.

Podemos usar métricas mas generales, métricas que midan error de variables numéricas que se pueda aplicar también a otros tipos de casos, como por ejemplo forecasting en series de tiempo.

# ERROR ABSOLUTO MEDIO (MAE)

El **error absoluto medio (MAE)** es el calculo del valor absoluto del residuo para cada punto de datos, para que los residuos negativos y positivos no se cancelen. Luego tomamos el promedio de todos estos residuos.

$$MAE = \frac{1}{N} \sum_{i=0}^{N-1} |y_{[i]} - \hat{y}_{[i]}|$$

Debido a que utilizamos el valor absoluto del residuo, MAE no indica si el modelo sobreestima o subestima.

# ERROR CUADRÁTICO MEDIO (MSE)

El **error cuadrático medio (MSE)** es similar al MAE, pero ahora calculamos el cuadrado de los residuos. Justamente esto es lo que se usa para encontrar los coeficientes.

$$MSE = \frac{1}{N} \sum_{i=0}^{N-1} (y_{[i]} - \hat{y}_{[i]})^2$$

MSE siempre es mayor a MAE. Un detalle importante son aquellos residuos grandes (outliers), en esta métrica aporta más que en MAE. En MAE el aporte es proporcional al valor del residuo, pero aquí es cuadráticamente mas grande.

# RAÍZ CUADRADA DEL ERROR CUADRÁTICO MEDIO (RMSE)

Si al MSE le calculamos la raíz, tendremos una métrica llamada **RMSE** que tiene la misma unidad de la salida original, donde el MSE no. El RMSE es análogo al desvío estándar.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=0}^{N-1} (y_{[i]} - \hat{y}_{[i]})^2}$$

# OUTLIERS

Valores atípicos es una constante de discusión. Se incluyen o no?

La respuesta dependerá del problema en particular, de los datos disponibles y las consecuencia que hay si se consideran o no.

Si quiero tenerlo en cuenta a la hora de comparar modelos, me va a convenir usar MSE, en cambio si quiero reducir su importancia, puedo usar MAE.

Ambas son métricas de error viables, pero **describen diferentes matices** sobre los errores de predicción.

# ERROR ABSOLUTO PORCENTUAL MEDIO (MAPE)

El **error absoluto porcentual medio (MAPE)** es el calculo del error MAE pero escalado al verdadero valor, por lo que el resultado es porcentual

$$MAPE = \frac{100\%}{N} \sum_{i=0}^{N-1} \left| \frac{y_{[i]} - \hat{y}_{[i]}}{y_{[i]}} \right|$$

No es una métrica buena porque es susceptible a errores numéricos. No puede calcularse cuando  $y_{[i]}$  vale cero. Y ademas tiene sesgo para cuando la predicción subestima:

n=1  $y_{pred} = 10$   $y_{real} = 20$   
MAPE = 50%

n=1  $y_{pred} = 20$   $y_{real} = 10$   
MAPE = 100%

# ERROR PORCENTUAL MEDIO (MPE)

El **error porcentual medio (MPE)** es igual a MAPE pero ahora no calculamos el valor absoluto.

$$MPE = \frac{100\%}{N} \sum_{i=0}^{N-1} \left( \frac{y_{[i]} - \hat{y}_{[i]}}{y_{[i]}} \right)$$

Aunque la falta de valor absoluto puede ser problemático ya que puede cancelar términos, nos permite saber si:

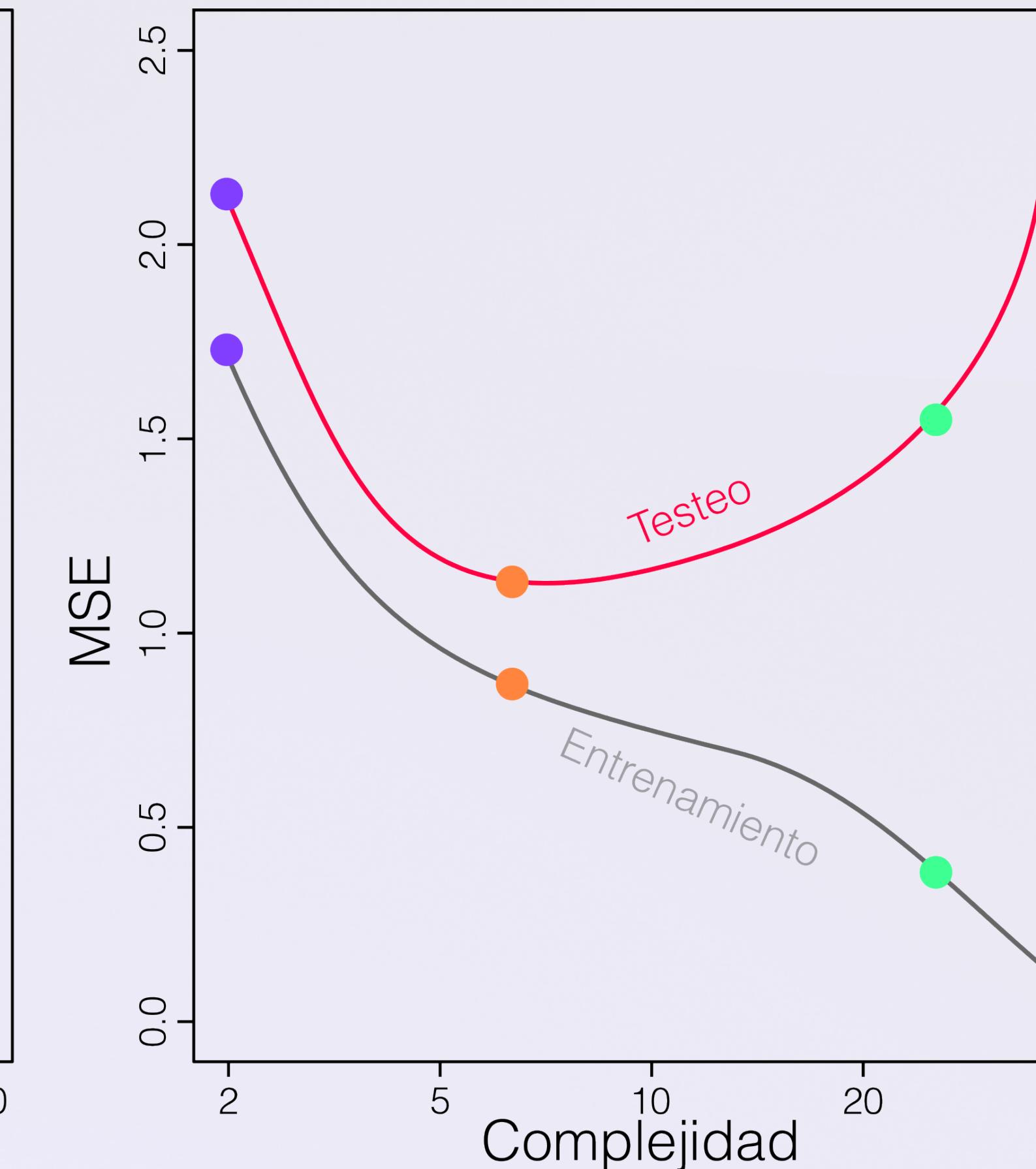
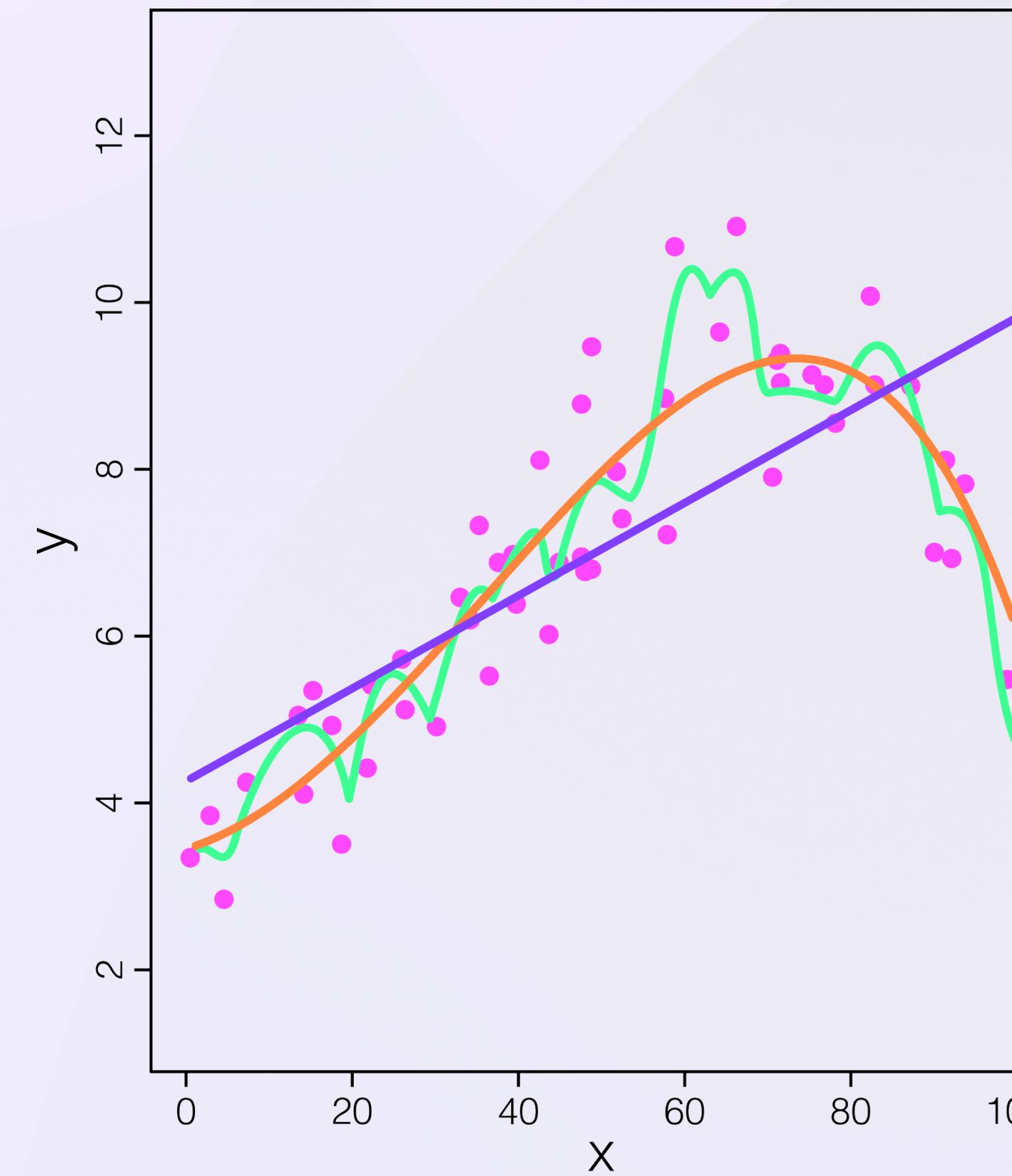
- El modelo subestima
- El modelo sobreestima

The background features a minimalist design with abstract, wavy shapes in shades of purple and dark blue. These shapes are layered and overlap, creating a sense of depth and movement across the entire frame.

**VAMOS A PRÁCTICAR UN POCO...**

# EVALUACIÓN

Veamos un ejemplo de tres regresiones, una lineal y otras dos de mayor complejidad.



# SESGO - VARIANZA

La curva en U del error MSE es producto de dos propiedades de los modelos. El error de los modelos se puede separar en tres cantidades fundamentales:

$$E(y_0 - \hat{f}(x_0))^2 = Var(\hat{f}(x_0)) + [Bias(\hat{f}(x_0))]^2 + Var(\varepsilon)$$

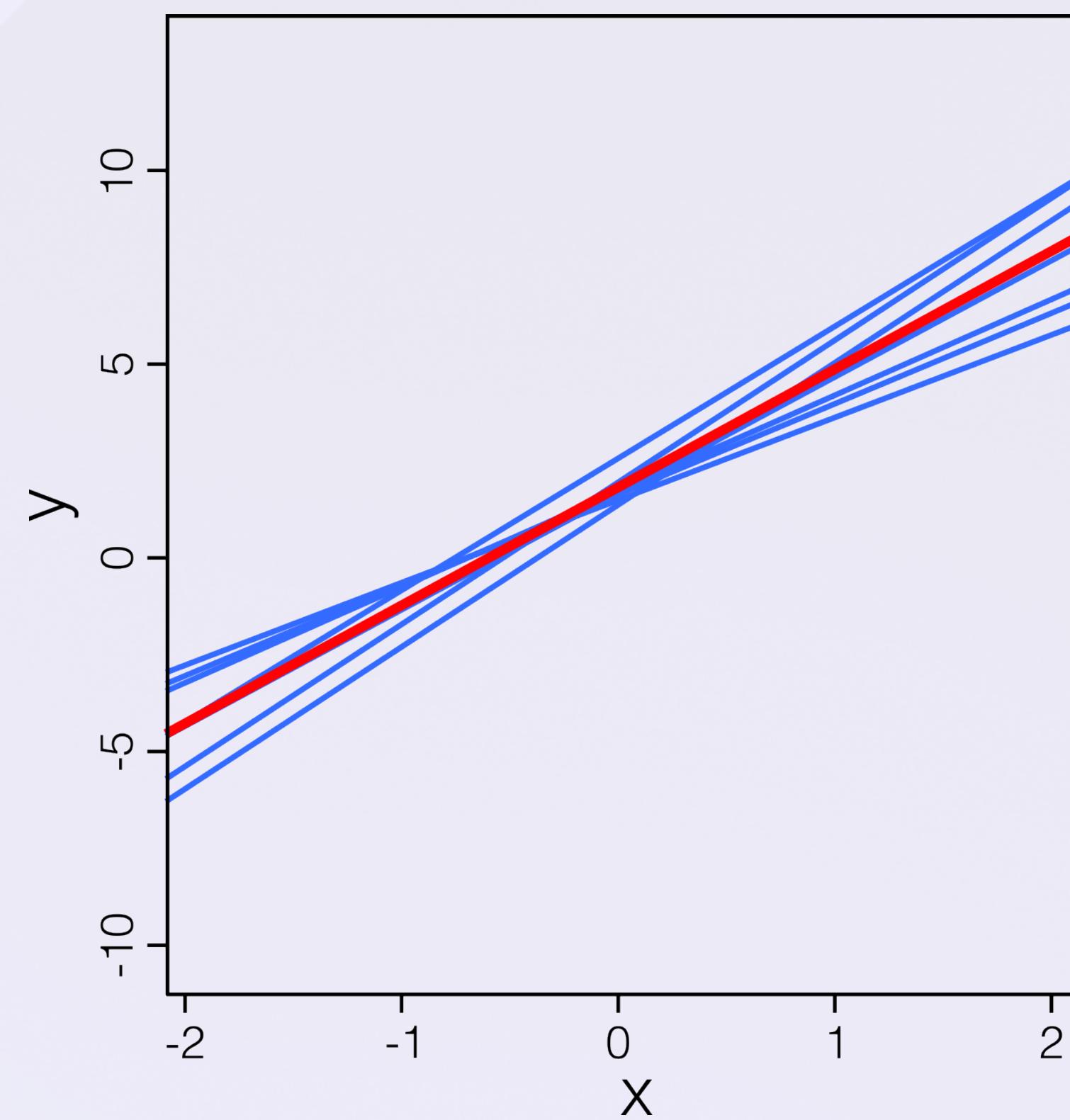
Donde:

- $E(y_0 - \hat{f}(x_0))^2$  es el MSE promedio si se estima repetidamente a f.
- $Var(\hat{f}(x_0))$  es la varianza del modelo.
- $[Bias(\hat{f}(x_0))]^2$  es el error introducido por el sesgo.
- $Var(\varepsilon)$  es la varianza que no explica el modelo.

# SESGO - VARIANZA

Pero qué es sesgo y varianza?

La varianza se refiere a la cantidad que la función estimada cambia si cambiamos el set de entrenamiento.



# SESGO - VARIANZA

Pero qué es sesgo y varianza?

La **varianza** se refiere a la cantidad que la función estimada cambia si cambiamos el set de entrenamiento. Es esperable que la estimación no varíe tanto entre datos de entrenamiento. Mas cantidad de datos para entrenar, nos mejoraría la varianza porque mas tenemos de la población.

# SESGO - VARIANZA

Pero qué es sesgo y varianza?

El **sesgo** se refiere al error introducido por aproximarnos a problemas de la vida real, que son mucho mas complicados que por ejemplo, una simple relación lineal. Es decir, dado un set de datos, si asumimos una relación lineal, es muy poco probable que un evento real tenga esa relación lineal.

Es decir, no importa que tanta observaciones agreguemos, no se van a poder obtener resultados mas exactos.

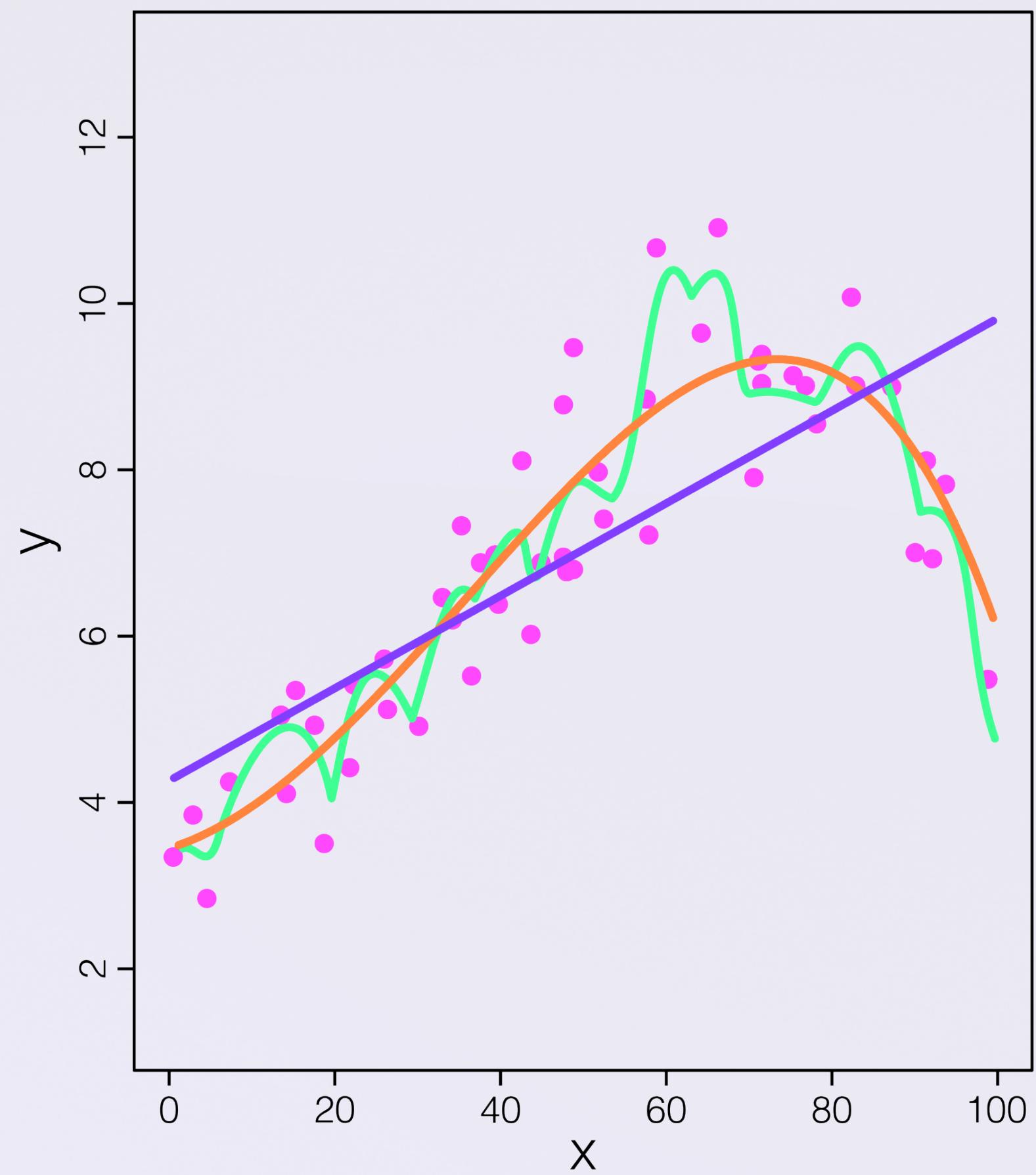
# SESGO - VARIANZA

Como regla general,

Cuando más complejo es el modelo, la varianza va a aumentar y el sesgo va a disminuir.

Cuando aumentamos la complejidad del mismo, el sesgo tiende a disminuir más rápido de lo que la variabilidad aumenta, disminuyendo el MSE.

Llega un punto en donde el efecto de la variabilidad es apreciable, aumentando el valor de MSE de nuevo.



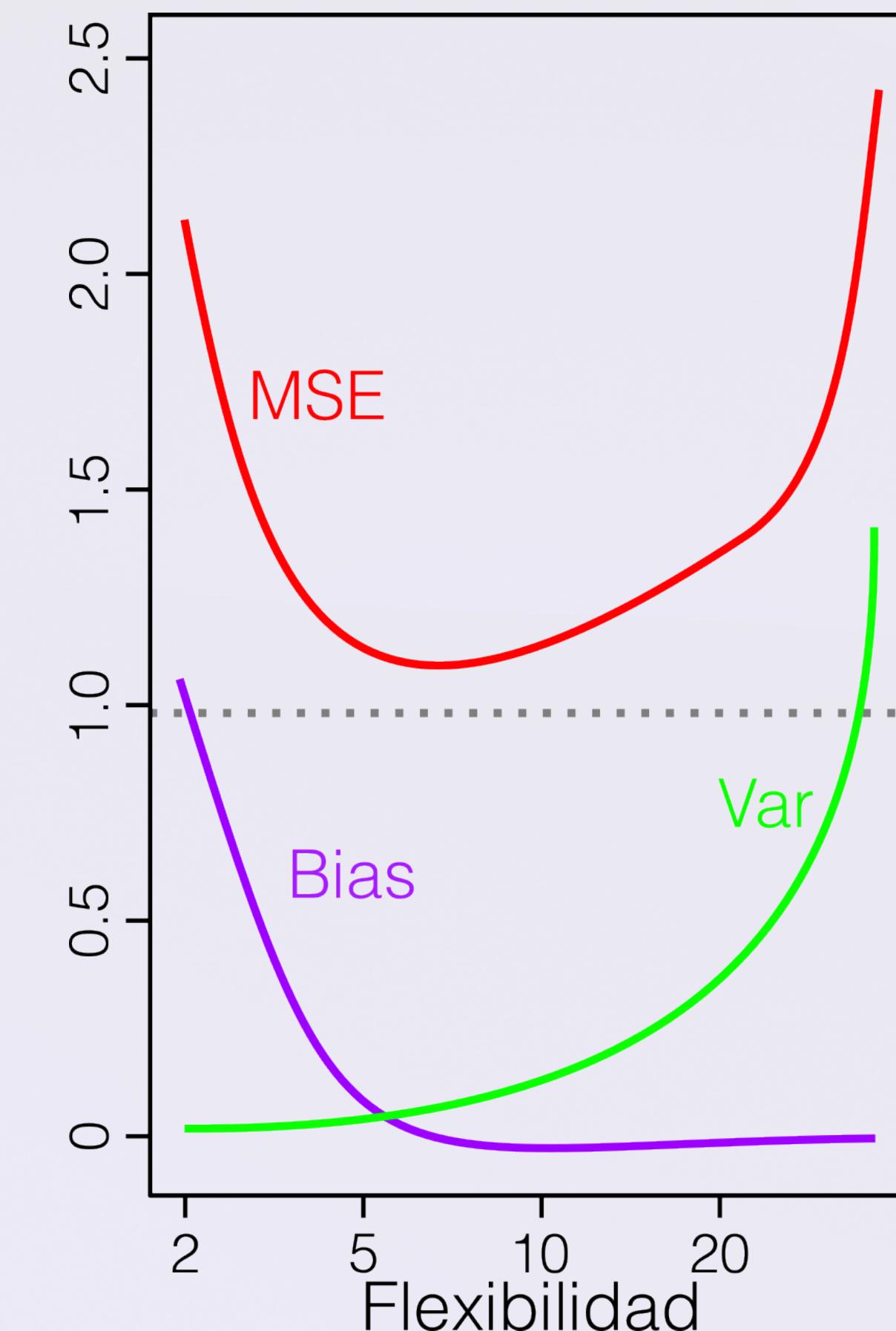
# SESGO - VARIANZA

Como regla general,

Cuando más complejo es el modelo, la varianza va a aumentar y el sesgo va a disminuir.

Cuando aumentamos la complejidad del mismo, el sesgo tiende a disminuir más rápido de lo que la variabilidad aumenta, disminuyendo el MSE.

Llega un punto en donde el efecto de la variabilidad es apreciable, aumentando el valor de MSE de nuevo.



# VARIABLES DUMMY

# VARIABLES DUMMY

Como vimos, regresión utiliza variables numéricas para predecir un valor. Como podemos hacer para usar variables categóricas?

Para poder usarlos, debemos transformarlos en numéricas mediante alguna codificación.

Cuando tenemos variables categóricas ordinales, podemos asociar un numero.

Por ejemplo, si tenemos que usar casos como : “me gusta mucho”, “me gusta poco”, “neutral”, “no me gusta poco”...

Se puede usar números enteros, teniendo en cuenta el orden para darle importancia. Estará en la creatividad de quien lo hace para determinar si las distancias son equidistantes o no.

# VARIABLES DUMMY

Ahora si tenemos casos nominales no podemos asociar números, porque al hacerlo, establecemos un orden.

Para este tipo de variable existe **one-hot encoding**.

Al usar esta codificación, creamos nuevos atributos de acuerdo con la cantidad de clases presentes en la variable categórica, es decir, si hay n número de categorías, se crearán n nuevos atributos. Estos atributos creados se denominan **variables dummy**.

# VARIABLES DUMMY

Peso	Altura	Pais
80	180	Argentina
83	177	Chile
75	169	Chile
68	155	Argentina

# VARIABLES DUMMY

Peso	Altura	País	país_0	país_1
80	180	Argentina	1	0
83	177	Chile	0	1
75	169	Chile	0	1
68	155	Argentina	1	0

# VARIABLES DUMMY

Peso	Altura	pais_0	pais_1
80	180	1	0
83	177	0	1
75	169	0	1
68	155	1	0

# VARIABLES DUMMY

Como vimos, el tipo de codificación llamado one-hot encoding nos genera un nuevo atributo por categoría, pero esto **nos genera una trampa**

Si vemos el ejemplo, las dos variables que estamos usando están 100% correlacionadas entre si.

$$x_{chile} = 1 - x_{arg}$$

$$\hat{y} = w_0 + w_1 x_{peso} + w_2 x_{altura} + w_3 x_{arg} + w_4 x_{chile}$$

$$\hat{y} = w_0 + w_1 x_{peso} + w_2 x_{altura} + w_3 x_{arg} + w_4 (1 - x_{arg})$$

$$\hat{y} = (w_0 + \textcolor{red}{w_4}) + w_1 x_{peso} + w_2 x_{altura} + (w_3 + \textcolor{red}{w_4}) x_{arg}$$

# VARIABLES DUMMY

La solución a este caso **es quitar una** de las columnas.

El procedimiento entonces es:

1. Dado N categorías
2. Creamos N nuevos atributos usando one-hot encoding de las categorías.
3. Quitamos la columna de categorías
4. Quitamos **una** de las columnas del encoding, quedándonos con N-1 columnas.

**Cual de las columnas se** quita no es importante, pero si se debe mantener consistencia en futuro usos del modelo.

# VARIABLES DUMMY

Peso	Altura	pais_0
80	180	1
83	177	0
75	169	0
68	155	1

# CONSTRUCCIÓN DE UN MODELO

# CONSTRUYENDO UN MODELO

Cuando construimos un modelo de regresión múltiple, como hacemos para elegir los features que formarán parte del modelo?

Ver la correlación entre variables es un primer paso, pero surge la pregunta, si dos variables estás correlacionadas, cual de las dos descarto?

Por lo que hay diferentes métodos de construir un modelo.

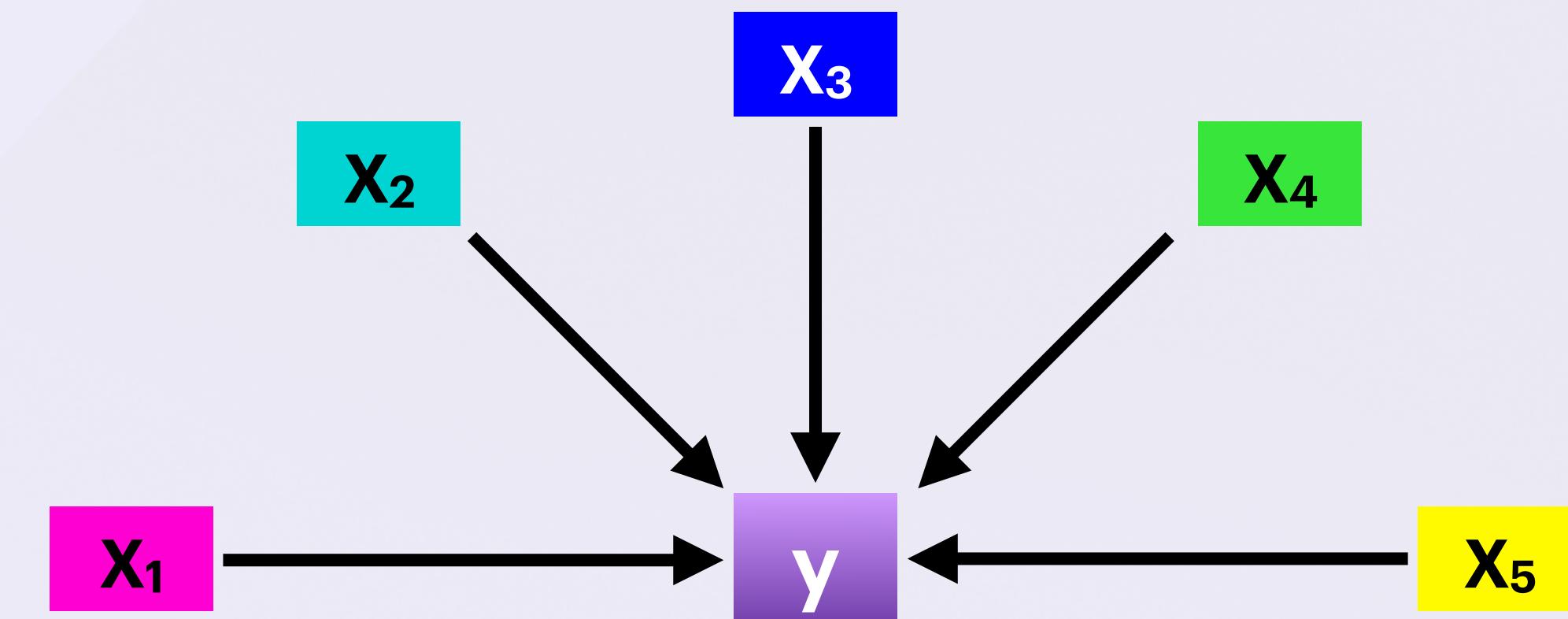
# CONSTRUYENDO UN MODELO

Podemos mencionar 4 formas:

- A. Exhaustivo
- B. Eliminación hacia atrás
- C. Selección hacia adelante
- D. Eliminación bidireccional

# EXHAUSTIVO (ALL-IN)

Este es el caso más sencillo, usamos todas las variables. En que casos conviene usar esta forma es cuando tenemos conocimiento a priori, o una necesidad específica.



# ELIMINACIÓN HACIA ATRÁS

Se arranca con un modelo completo con todas las variables.

Se va eliminando variables de entrada que menos mejoran el modelo de una por vez.

Se continua hasta que eliminar variables hasta que eliminar variables mejore significativamente el modelo.

La forma que nosotros lo vamos a hacer es usar la bondad de ajuste. El valor p de cada término evalúa la hipótesis nula de que el coeficiente es igual a cero (no hay efecto). Un valor p bajo ( $< 0,05$ ) indica que se puede rechazar la hipótesis nula, indicando que cambios en esta variable es probable que genere cambios en la respuesta.

# SELECCIÓN HACIA ADELANTE

Comienza con un modelo "vacío" con solo la ordenada al origen.

Luego se agrega las variables que más aporta al modelo (usando el criterio de ajuste) de una por vez.

Se termina una vez que agregar mas variables no genera mejor aporte.

# **ELIMINACIÓN BIDIRECCIONAL**

Es en esencia la selección hacia adelante pero dando la posibilidad de quitar variables en cada iteración cuando se observa correlación entre variables.

# OTRAS MÉTRICAS QUE NOS AYUDAN A CONSTRUIR EL MODELO

Nosotros vamos a trabajar para este caso con la bondad de ajuste. Pero contamos con otras herramientas que nos dan información cuando construimos el modelo, los cuales nos indican que tan bien es el ajuste, pero penalizando el numero de coeficientes.

Por un lado tenemos el **coeficiente de Pearson ajustado**:

$$R_{adj}^2 = 1 - (1 - R^2) \frac{N - 1}{N - d - 1}$$

Donde N es numero de observaciones y d es el numero de atributos. Esto nos permite penalizar la complejidad del modelo.

# OTRAS MÉTRICAS QUE NOS AYUDAN A CONSTRUIR EL MODELO

**Criterio de Información de Akaike (AIC)**: AIC maneja un equilibrio entre la bondad de ajuste del modelo y la complejidad del modelo. En otras palabras, AIC aborda tanto el riesgo de sobre-ajuste como el riesgo sub-ajuste.

$$AIC = 2d - 2\ln(\hat{L})$$

Donde  $\ln(\hat{L})$  es el logaritmo de la estimación de máxima similitud para una regresión lineal. Este valor se obtiene como:

$$\hat{L} = -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln\left(\frac{S_E}{N}\right) - \frac{N}{2}$$

El valor de AIC cuando es más bajo, es mejor.

# OTRAS MÉTRICAS QUE NOS AYUDAN A CONSTRUIR EL MODELO

**Criterio de información bayesiano (BIC)**: Se basa en el principio de la navaja de Occam, que establece que es preferible el modelo más simple que explique los datos. A diferencia del AIC, BIC penaliza más el modelo por su complejidad.

$$BIC = d \ln(N) - 2\ln(\hat{L})$$

Donde  $\ln(\hat{L})$  es el logaritmo de la estimación de máxima similitud para una regresión lineal

El valor de BIC cuando es más bajo, es mejor.

# OTRAS MÉTRICAS QUE NOS AYUDAN A CONSTRUIR EL MODELO

Con esta métricas podemos aplicar cualquier de las formas de construcción de modelo. Es más, si usamos **BIC** o **AIC**, podemos extender la búsqueda del modelo más optimo usando otros tipos de modelos.

Esto lo podemos usar sumados a validación cruzada, y técnica de búsqueda de hiperparámetros, para encontrar el modelo que mejor prediga en nuestro problema.

The background features a minimalist design with abstract, wavy shapes in shades of purple and dark blue. These shapes are layered and overlap, creating a sense of depth and movement across the entire frame.

**VAMOS A PRÁCTICAR UN POCO...**



# REGRESIÓN POLINÓMICA

# GENERALIZACIÓN

Si volvemos al modelo de regresión:

$$y(\hat{x}) = w_0 + w_1 x_1 + \dots + w_d x_d$$

Podemos generalizando reemplazando a los coeficientes por funciones.

$$y(\hat{x}) = w_0 + \sum_{j=0}^{N-1} w_j \phi(x_j)$$

Donde  $\Phi$  son funciones bases. Si elegimos funciones no lineales, podemos realizar regresiones no lineales. Un caso particular es la regresión polinómica.

# REGRESIÓN POLINÓMICA

Veamos el caso mas sencillo de una sola variable de entrada, el modelo quedaría:

$$y(\hat{x}) = w_0 + w_1x_1 + w_2x_1^2 + \dots + w_mx_1^m$$

El orden del polinomio es una elección que no depende de los datos (hiperparámetro).

La forma de encontrar los coeficientes sigue siendo mediante el método de cuadrados mínimos. Usando la formula general de regresión, se puede llegar a la formulación general de minimización por cuadrados mínimos.

Si agregamos mas de una entrada, cada una puede ser un polinomio diferente.

The background features a minimalist design with abstract, wavy shapes in shades of purple and dark blue. These shapes are layered and overlap, creating a sense of depth and movement across the entire frame.

**VAMOS A PRÁCTICAR UN POCO...**