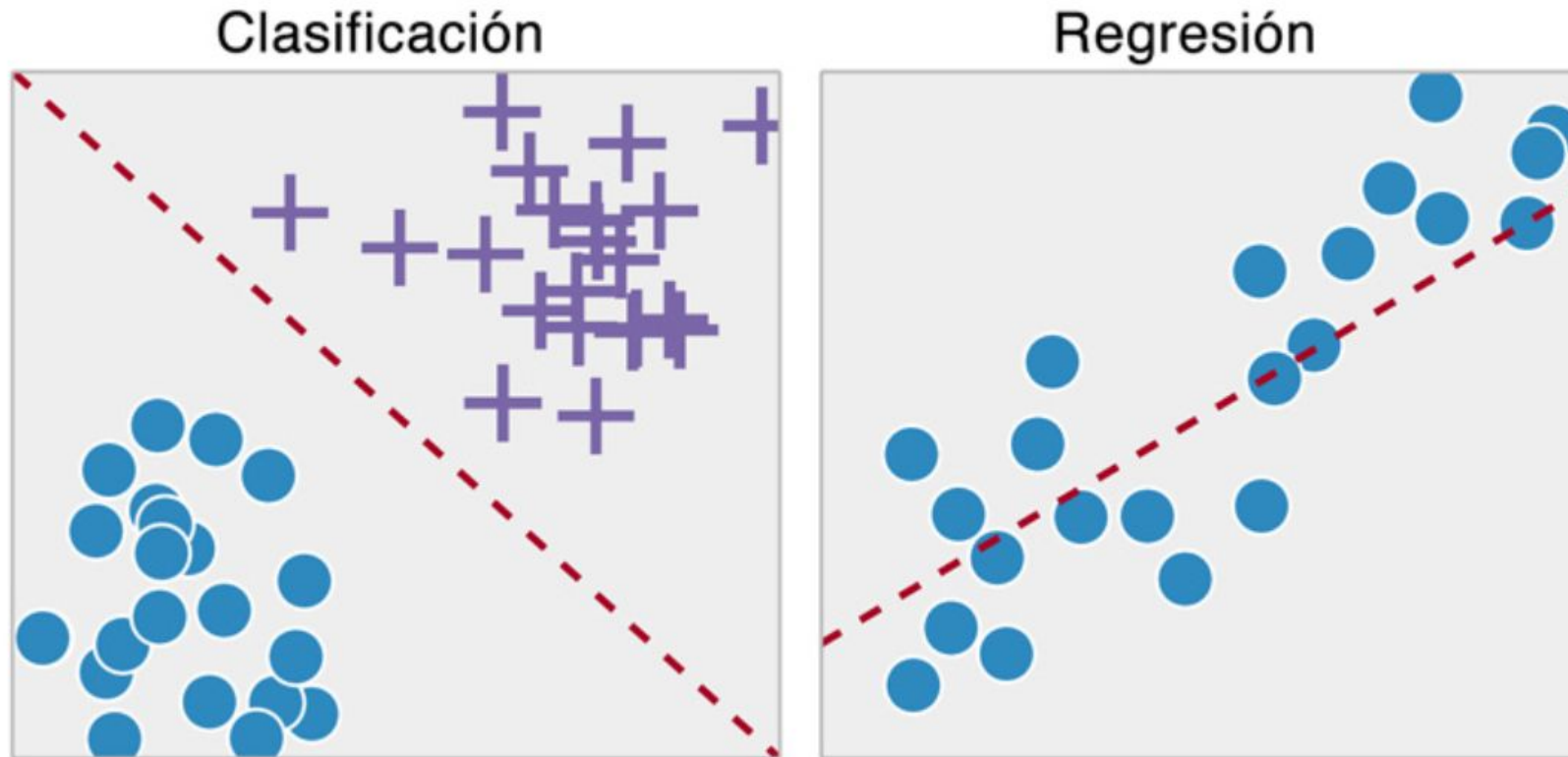



Hay dos tipos principales de problemas de aprendizaje supervisados llamados **clasificación** y **regresión**.



En la **clasificación**, el objetivo es predecir una etiqueta o clase, que es una elección de una lista de posibilidades.

- Clasificación (especies, preferencias, etc.).
 - Preguntas si/no (spam, posibilidad de compra, utilidad para cierta tarea, etc.).
- 


En la **clasificación**, el objetivo es predecir una etiqueta de clase, que es una elección de una lista predefinida de posibilidades.

- Clasificación multiclase.
- Clasificación binaria.

En **regresión**, el objetivo es predecir un número continuo o un número de punto flotante.

- Ingresos anuales de una eprsona.
- Ganancias de una empresa.
- Promedio de empleados de una empresa.


Generalización, Overfitting y Underfitting.

- Queremos construir un modelo sobre el conjunto de datos de entrenamiento que sea capaz de hacer predicciones precisas en datos nuevos con las mismas características.
 - Si un modelo es capaz de hacer esto, decimos que generaliza el conjunto de entrenamiento a un conjunto de datos nuevos o desconocidos para el modelo.
- 


Generalización, Overfitting y Underfitting.

Age	Number of cars owned	Owns house	Number of children	Marital status	Owns a dog	Bought a boat
66	1	yes	2	widowed	no	yes
52	2	yes	3	married	no	yes
22	0	no	0	married	yes	no
25	1	no	1	single	no	no
44	0	no	2	divorced	yes	no
39	1	yes	2	married	yes	no
26	1	no	2	single	no	no
40	3	yes	1	married	yes	no
53	2	yes	2	divorced	no	yes
64	2	yes	3	divorced	no	no
58	2	yes	2	married	yes	yes
33	1	no	1	single	no	no

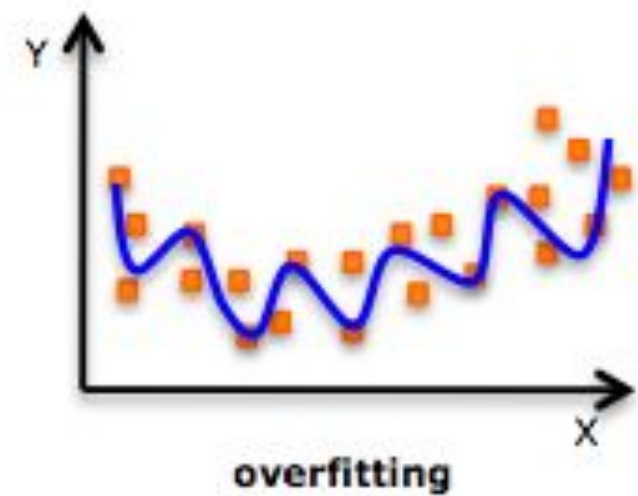
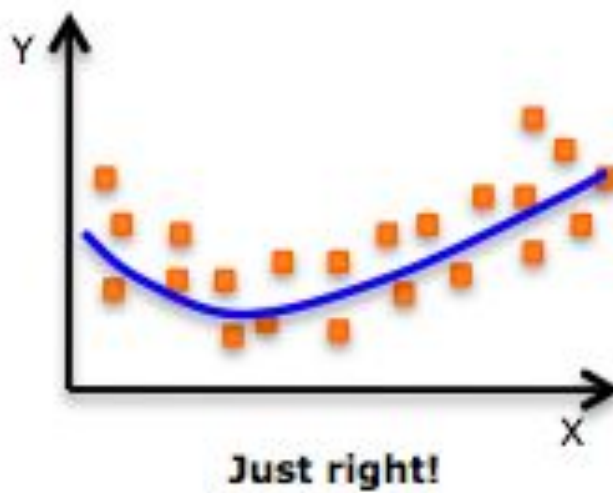
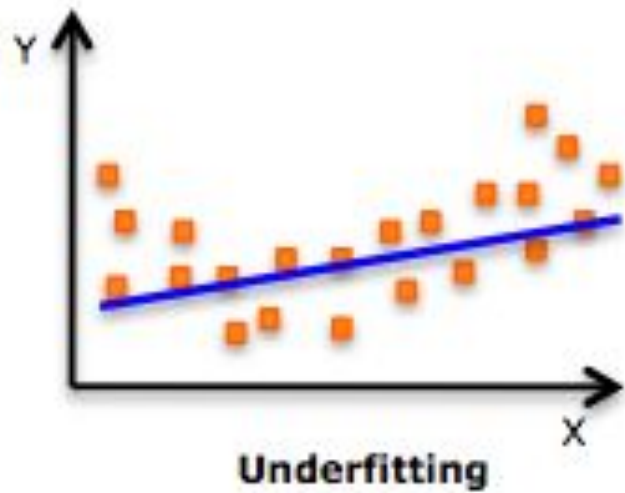
Generalización, Overfitting y Underfitting.

- El modelo propuesto, ¿es un buen modelo?
 - ¿Qué problema tiene?
 - ¿Cómo se podría mejorar?
- 

Generalización, Overfitting y Underfitting.

- Construir un modelo demasiado complejo para la cantidad de datos que tenemos es llamado **overfitting (sobre ajuste)**.
 - El overfitting ocurre cuando se usa un modelo demasiado cercano a las particularidades del conjunto de entrenamiento.
 - Por otra parte, construir un modelo muy simple es llamado **underfitting**.
- 

Generalización, Overfitting y Underfitting.



Linear Regression.

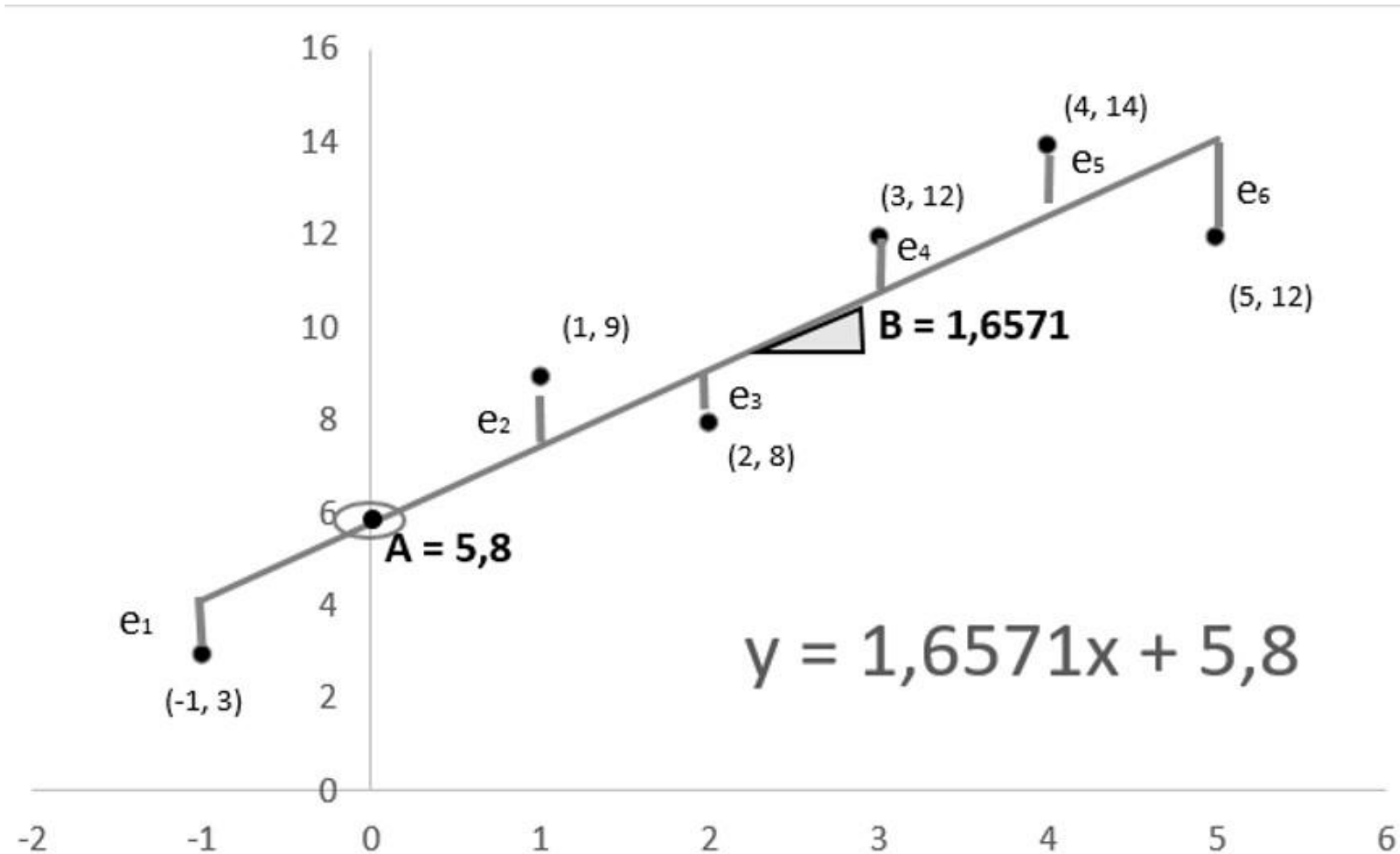
$$\hat{y} = w[0] * x[0] + w[1] * x[1] + \dots + w[p] * x[p] + b$$

$$\nabla_{\mathbf{w}} \text{MSE}_{\text{train}} = 0$$

$$\Rightarrow \nabla_{\mathbf{w}} \frac{1}{m} \|\hat{\mathbf{y}}^{(\text{train})} - \mathbf{y}^{(\text{train})}\|_2^2 = 0$$

$$\Rightarrow \frac{1}{m} \nabla_{\mathbf{w}} \|\mathbf{X}^{(\text{train})} \mathbf{w} - \mathbf{y}^{(\text{train})}\|_2^2 = 0$$

Linear Regression.



Linear Regression.

