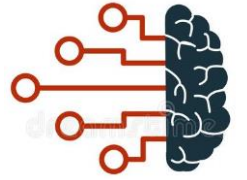


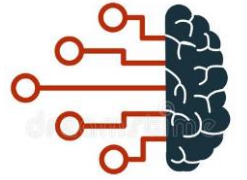
Validación en Machine Learning



Validez del Modelo

Capacidad de generalización:

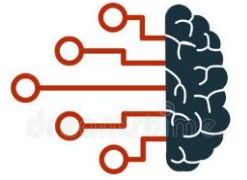
- El objetivo de Machine Learning es que el modelo obtenido después del aprendizaje tenga un buen rendimiento en nuevas muestras de datos
- Un buen modelo no sólo en las muestras utilizadas para el entrenamiento, sino para muestras nuevas
- La capacidad de aplicar un modelo a nuevas muestras de datos se denomina generalización o robustez.



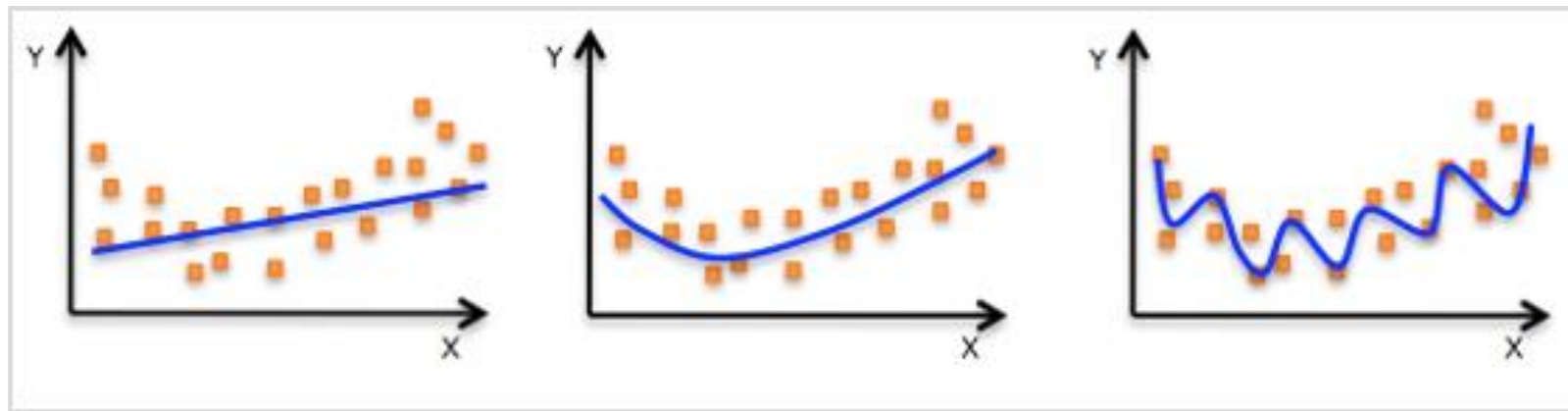
Validez del Modelo

- **Error:** diferencia entre el resultado de la muestra predicho por el modelo obtenido después del aprendizaje, y el resultado de la muestra real.
 - Error de entrenamiento: error que se obtiene al ejecutar el modelo en los datos de entrenamiento.
 - Error de generalización: error que se obtiene al ejecutar el modelo en nuevas muestras. Obviamente, preferimos un modelo con un error de generalización menor.
- **Subajuste:** se produce cuando el modelo o el algoritmo no se ajusta a los datos lo suficientemente bien.
- **Sobreajuste:** se produce cuando el error de formación del modelo obtenido después del aprendizaje es pequeño pero el error de generalización es grande (capacidad de generalización deficiente).

Validez del Modelo



- **Capacidad del modelo:** de ajustar funciones, que también se denomina complejidad del modelo.
- Cuando la capacidad se adapta a la complejidad de la tarea y a la cantidad de datos de entrenamiento proporcionados, el efecto de algoritmo suele ser óptimo.
- Los modelos con capacidad insuficiente no pueden resolver tareas complejas y pueden producirse subajustes
- Un modelo de alta capacidad puede resolver tareas complejas, pero puede producirse un sobreajuste si la capacidad es mayor que la requerida por una tarea.



Subajuste

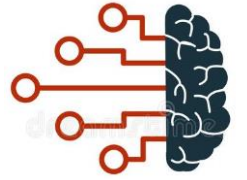
No todas las características son aprendidas.

Buen ajuste

Sobreajuste

Los ruidos se aprenden.

Sobreajuste Causa - Error



Error total de predicción final = Bias² + Varianza + Error Irreducible

El error de predicción se puede dividir en dos tipos:

- Error causado por "bias" (sesgo)

- Error causado por "variancia"

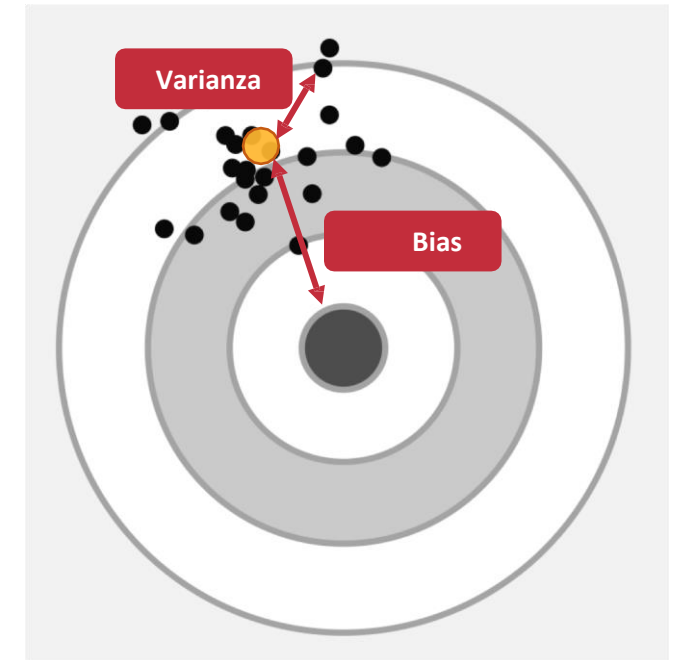
Varianza:

Desfase del resultado de la predicción a partir del valor promedio

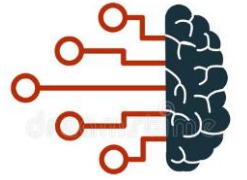
Error causado por la sensibilidad del modelo a pequeñas fluctuaciones en el conjunto de entrenamiento

Bias (sesgo):

Diferencia entre el valor previsto (o promedio) de predicción y el valor correcto que estamos intentando predecir.



Varianza y Bias (sesgo)



Las combinaciones de variación y sesgo son las siguientes:

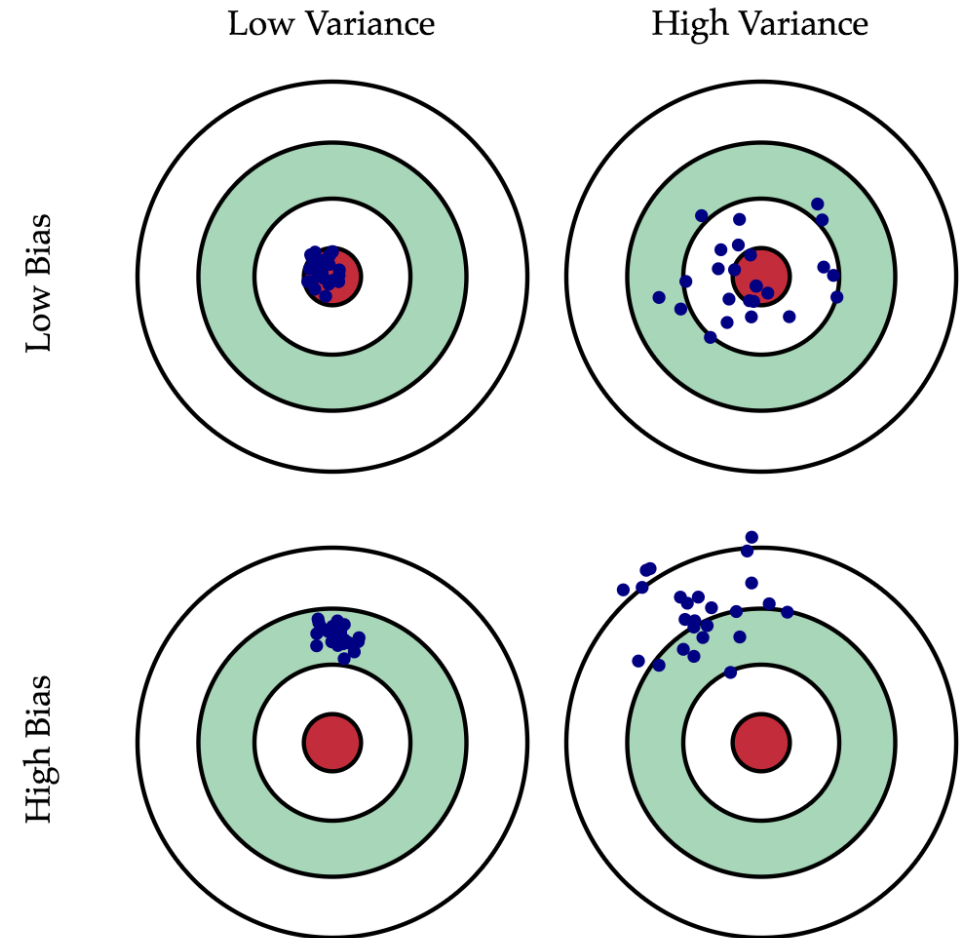
Bajo sesgo y baja varianza → Buen modelo

Bajo sesgo y alta varianza

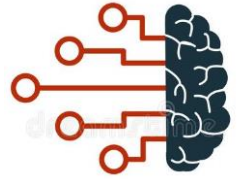
Sesgo alto y baja varianzas

Sesgo alto y alta varianza → Mal modelo

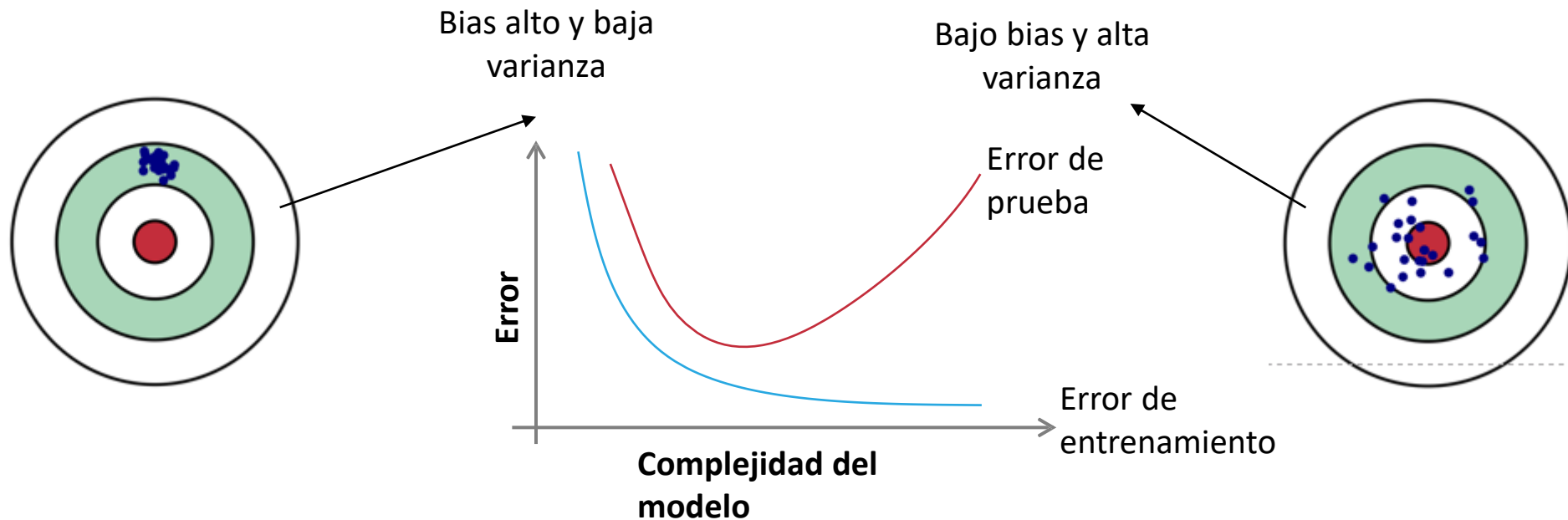
Idealmente, queremos un modelo que pueda capturar con precisión las reglas en los datos de entrenamiento y resumir los datos invisibles (nuevos datos). Pero por lo general es imposible para el modelo completar ambas tareas al mismo tiempo.



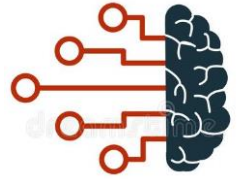
Complejidad y Error del modelo



A medida que aumenta la complejidad del modelo, disminuye el error de entrenamiento. A medida que aumenta la complejidad del modelo, el error de prueba disminuye a un punto determinado y luego aumenta en la dirección inversa, formando una curva convexa.



Evaluación del desempeño - Regresión



Cuanto más cerca esté el Mean Absolute Error (MAE) de 0, mejor se ajustará el modelo a los datos de entrenamiento.

$$MAE = \frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i|$$

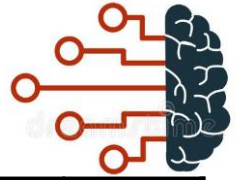
Mean Square Error (MSE)

$$MSE = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2$$

El rango de valores de R^2 es $(-\infty, 1]$. Un valor mayor indica que el modelo puede ajustarse mejor a los datos de entrenamiento. TSS indica la diferencia entre muestras. RSS indica la diferencia entre el valor predicho y el valor de muestra.

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_{i=1}^m (y_i - \hat{y}_i)^2}{\sum_{i=1}^m (y_i - \bar{y}_i)^2}$$

Evaluación del rendimiento - clasificación



Terminos y definiciones:

P: Positivo, indicando el número de casos reales positivos en los datos

N: Negativo, indicando el número de casos reales negativos en los datos

TP: True positive (positivo real), indica el numero de casos positivos que son clasificados correctamente por el clasificador

TN: True Negative (negativo real), indica el numero de casos negativos que son clasificados correctamente por el clasificador

FP: Falso Positivo, indica el número de casos positivos que son clasificados incorrectamente por el clasificador

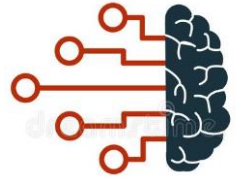
FN: Falso Negativo, indica el número de casos negativos que son clasificados incorrectamente por el clasificador

Cantidad estimada \ Cantidad real	yes	no	Total
Sí	TP	FN	P
No	FP	TN	N
Total	P'	N'	$P + N$

Matriz de confusión: al menos una tabla $m \times m$. $CM_{i,j}$ de las primeras filas m y columnas m indica el número de casos que realmente pertenecen a la clase i pero que están clasificados en la clase j por el clasificador.

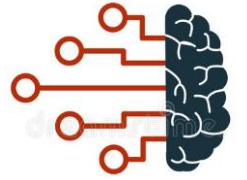
Idealmente, para un clasificador de alta precisión, la mayoría de los valores de predicción deben ubicarse en la diagonal de $CM_{1,1}$ a $CM_{(m,m)}$ de la tabla, mientras que los valores fuera de la diagonal son 0 o cerca de 0. Es decir, FP y FN son cerca de 0.

Evaluación del rendimiento - clasificación



Medición	Proporción
Exactitud y tasa de reconocimiento	$\frac{TP + TN}{P + N}$
Tasa de error y tasa de clasificación errónea	$\frac{FP + FN}{P + N}$
Sensibilidad, tasa positiva real, y recordatorio	$\frac{TP}{P}$
Especificidad y tasa negativa real	$\frac{TN}{N}$
Precisión	$\frac{TP}{TP + FP}$
F_1 , harmonic mean of the recall rate and precision	$\frac{2 \times precision \times recall}{precision + recall}$
F_β , where β is a non-negative real number	$\frac{(1 + \beta^2) \times precision \times recall}{\beta^2 \times precision + recall}$

Ejemplo de evaluación del rendimiento



Hemos entrenado un modelo de M.L. para identificar si el objeto de una imagen es un gato. Ahora usamos 200 imágenes para verificar el desempeño del modelo. Entre las 200 imágenes, los objetos en 170 imágenes son gatos, mientras que otros no lo son.

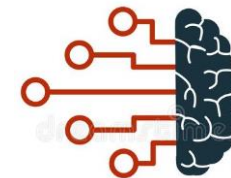
El resultado de identificación del modelo es que los objetos en 160 imágenes son gatos, mientras que otros no lo son.

$$\text{Precisión: } P = \frac{TP}{TP+FP} = \frac{140}{140+20} = 87.5\%$$

$$\text{Recall: } R = \frac{TP}{P} = \frac{140}{170} = 82.4\%$$

$$\text{Exactitud: } ACC = \frac{TP+TN}{P+N} = \frac{140+10}{170+30} = 75\%$$

Cantidad estimada Cantidad real	<i>yes</i>	<i>no</i>	Total:
<i>yes</i>	140.	30.	170.
<i>no</i>	20.	10.	30.
Total:	160.	40	200.



Gracias.....