Bias-Variance Tradeoff

Luis Norberto Zúñiga Morales

5 de septiembre de 2022

Bias - Variance Tradeoff

Supongan que se tiene un conjunto de datos $x_1, ..., x_n$, donde cada x_i tiene asociado un valor real y_i. Asumamos que existe una función

$$y = f(x) + \epsilon$$

donde ϵ representa el ruido, que sigue una distribución de probabilidad P con media cero y varianza σ^2.

Bias - Variance Tradeoff

En teoría, buscamos encontrar un función $\hat{f}(x; D)$ que se aproxime lo más posible a la función f que modela nuestro conjunto de datos $D = \{(x_1, y_1), ..., (x_n, y_n)\}$.

¿Cómo se logra ésto?

Bias - Variance Tradeoff

Por medio de un algoritmo o modelo de aprendizaje que utilice el conjunto de datos D que ajuste la función $\hat{f}$ lo mejor posible .

¿Cómo medir el desempeño? Por medio del error cuadrático medio entre y y $\hat{f}(x; D)$:

$$EQM = (y - \hat{f}(x; D))^2$$

por lo que se busca minimizarlo con respecto a los puntos x_i del conjunto D, pero también idealmente para puntos fuera de D.

Bias - Variance Tradeoff

Pregunta

¿Qué nos impide obtener $\hat{f}(x; D) = y$?

Bias - Variance Tradeoff

Pregunta

¿Qué nos impide obtener $\hat{f}(x; D) = y$?

Respuesta

y_i tiene un componente de ruido aleatorio que nos es imposible modelar, i.e., hay un *error irreducible*.

Bias - Variance Tradeoff

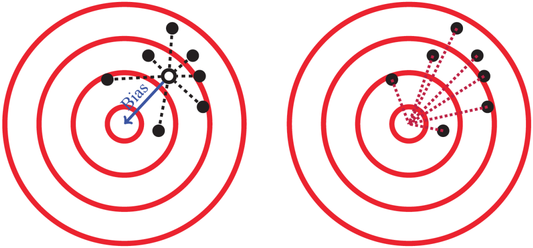
Definición: Sesgo

El sesgo (en el área del aprendizaje automático) se puede entender como la diferencia entre el promedio de las predicciones que realiza un modelo y el valor correcto que buscan predecir.
Modelos con alto sesgo no se fijan mucho en los puntos del conjunto de datos, llevando al underfitting.

Definición: Varianza

La varianza es la variabilidad de la predicción de un modelo en un punto en específico, ligado a la dispersión de nuestros datos.
Un modelo con alta varianza presta mucha atención a los elementos del conjunto de datos, perdiendo la capacidad para generalizar puntos fuera de ese conjunto, llevando al overfitting.

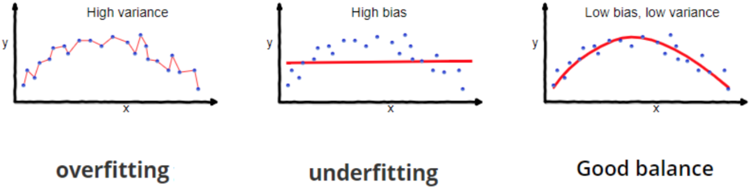
Bias - Variance Tradeoff



(a) Bias and variance. (b) Mean-squared error.

FIGURE 1. *A depiction of the bias-variance tradeoff using targets. If these points are thought of as arrows, then the goal would be for the points to be near the center of the target. (a) To the extent which the points are far from the center, they suffer from bias (solid blue line) and/or variance (dashed black lines). (b) The bias and variance combine to form mean squared error (dotted purple lines).*

Bias - Variance Tradeoff



Bias - Variance Tradeoff

El error esperado se puede representar de la siguiente forma:

$$E_{D,\varepsilon}\left[\left(y-\hat{f}(x;D)\right)^2\right] = \left(\text{Sesgo}_D\left[\hat{f}(x;D)\right]\right)^2 + \text{Var}_D\left[\hat{f}(x;D)\right] + \sigma^2$$

donde

$$\text{Sesgo}_D\left[\hat{f}(x;D)\right] = E_D\left[\hat{f}(x;D)\right] - f(x)$$

y

$$\text{Var}_D\left[\hat{f}(x;D)\right] = E_D[(E_D[\hat{f}(x;D)] - \hat{f}(x;D))^2].$$

Bias - Variance Tradeoff

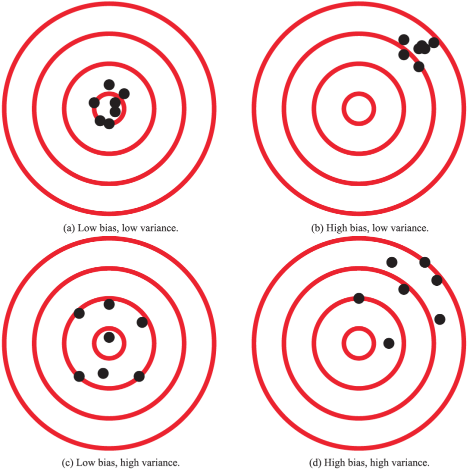


FIGURE 2. *A depiction of varying combinations of bias and variance: (a) low bias and low variance, (b) high bias and low variance, (c) low bias and high variance, and (d) high bias and high variance.*

Ejercicio (Tarea)

Demostrar que

$$EQM = \text{Sesgo}^2 + \text{Var} + \sigma^2$$

donde

$\text{Sesgo}_D\left[\hat{f}(x;D)\right] = \mathrm{E}_D\left[\hat{f}(x;D)\right] - f(x)$
y

$$\text{Var}_D\left[\hat{f}(x;D)\right] = \mathrm{E}_D[\left(\mathrm{E}_D[\hat{f}(x;D)] - \hat{f}(x;D)\right)^2].$$

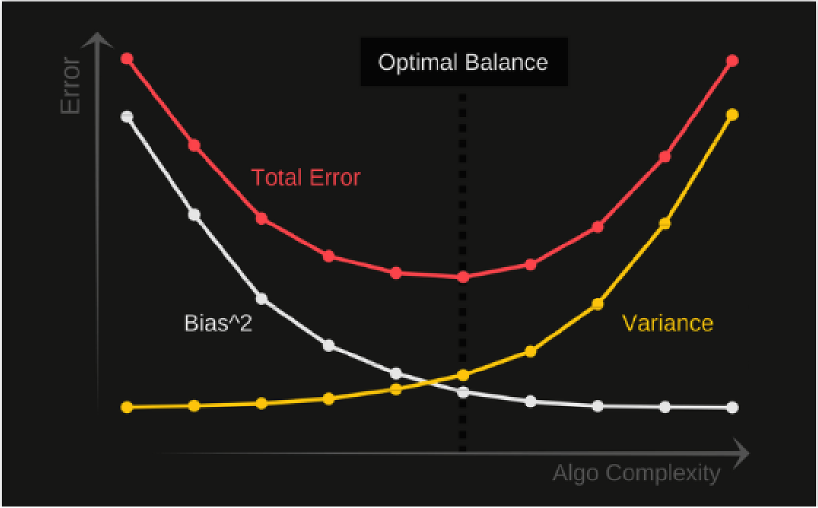
Bias - Variance Tradeoff

¿Dónde está el *tradeoff*?

- Si el modelo es muy simple y con pocos parámetros, puede presentar alto sesgo y baja varianza.
- Si el modelo tiene muchos parámetros, puede presentar bajo sesgo y alta varianza.

En resumen, un modelo no puede ser más y menos complejo al mismo tiempo.

Bias - Variance Tradeoff



Bias - Variance Tradeoff

¿Qué opciones existen para evitar el overfitting y el underfitting?

- Regularización
- Ensemble Learning