

Compensación Sesgo-Varianza

Compensación Sesgo-Varianza

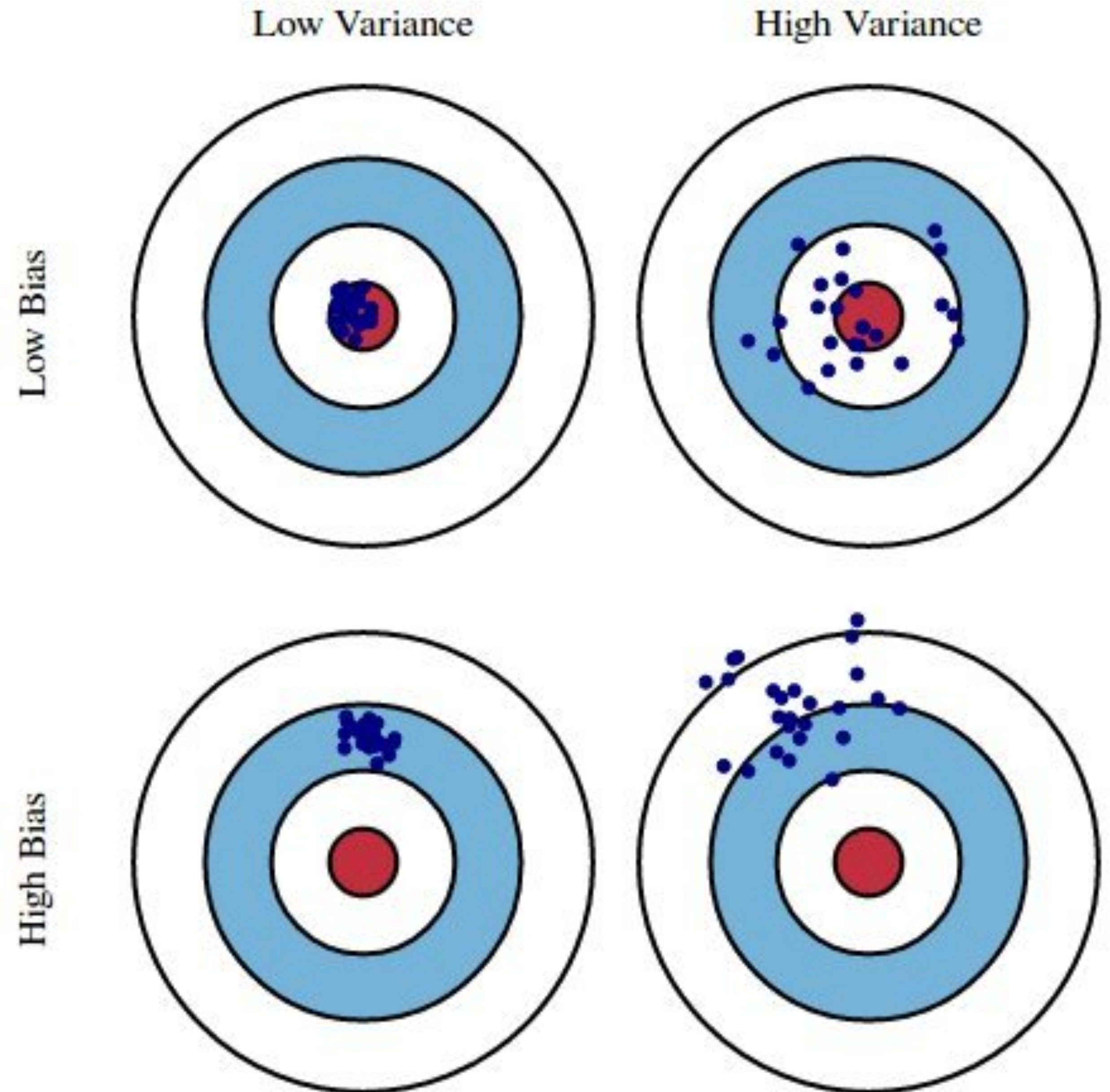
La compensación, o solución intermedia, entre el sesgo y la varianza es un tema de vital importancia en el entendimiento del rendimiento, y la evaluación, de un modelo.

Compensación Sesgo-Varianza

- La compensación, o solución intermedia, entre el sesgo y la varianza es el punto donde simplemente le estamos agregando ruido a nuestro modelo, al ir agregando complejidad (flexibilidad) a nuestro modelo.
- El error del entrenamiento baja, como debe ser, pero el error de la prueba empieza a subir.
- El modelo, luego de la compensación, empieza a sobre-ajustarse (overfit) a los datos.

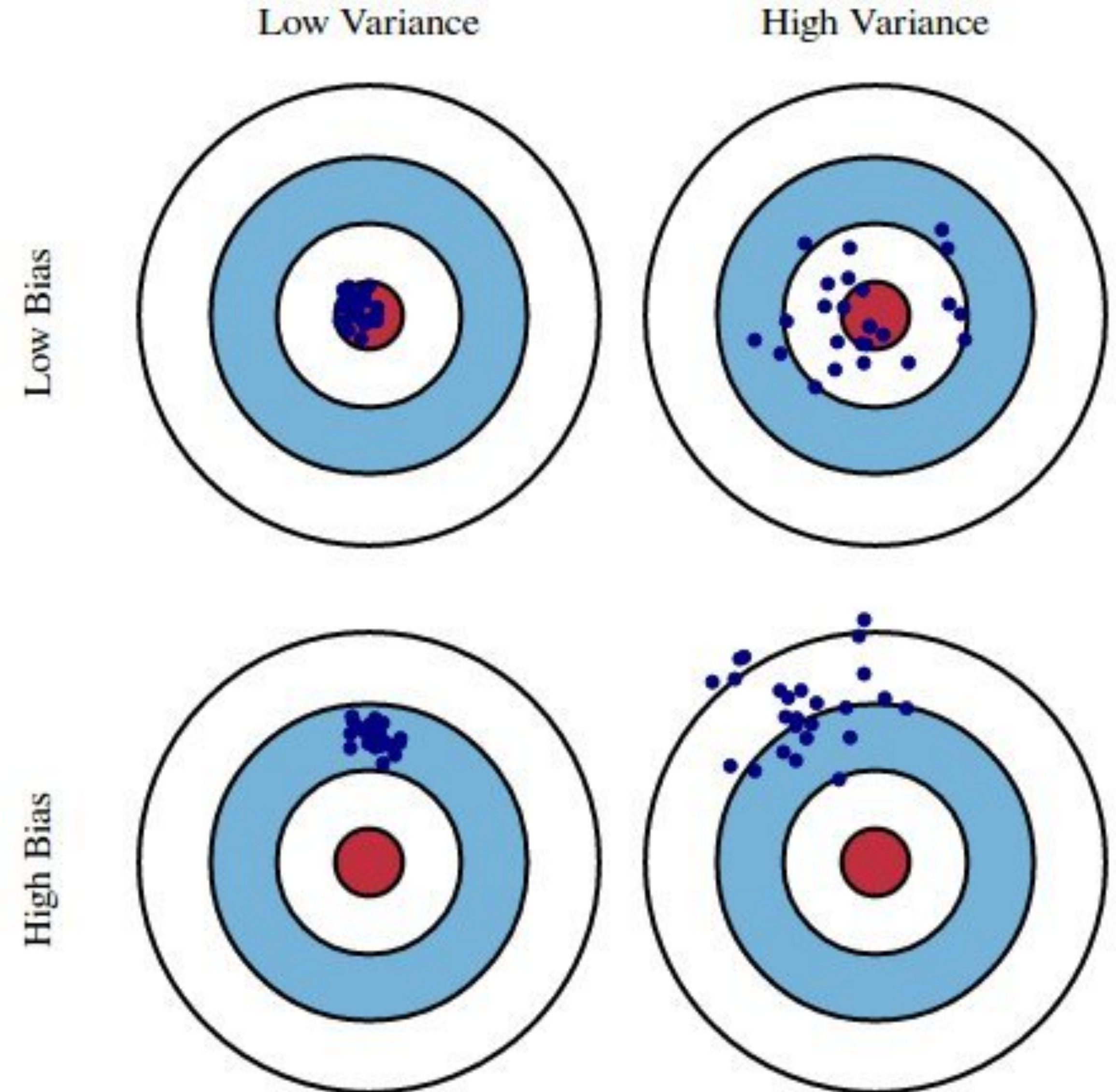
Compensación Sesgo-Varianza

- Imaginemos que el centro de un blanco es un modelo que predice perfectamente los valores correctos.
- Conforme nos apartamos del centro, nuestras predicciones se van empeorando.



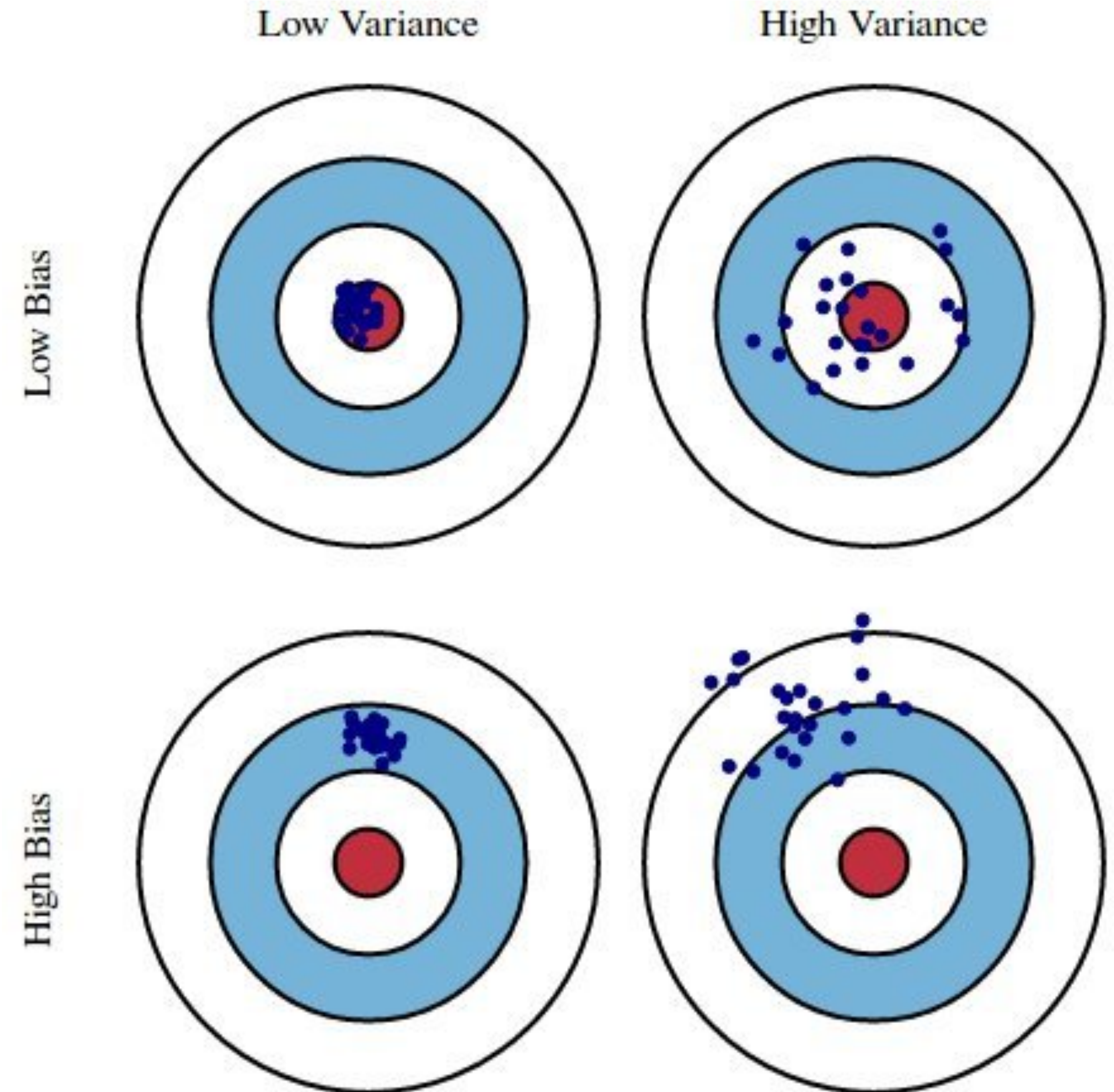
Compensación Sesgo-Varianza

- Imaginemos que podemos repetir el proceso de construir el modelo para obtener un número de intentos distintos.
- Cada intento representa una realización distinta del modelo, dado la variabilidad al azar en los datos de entrenamiento que obtenemos.



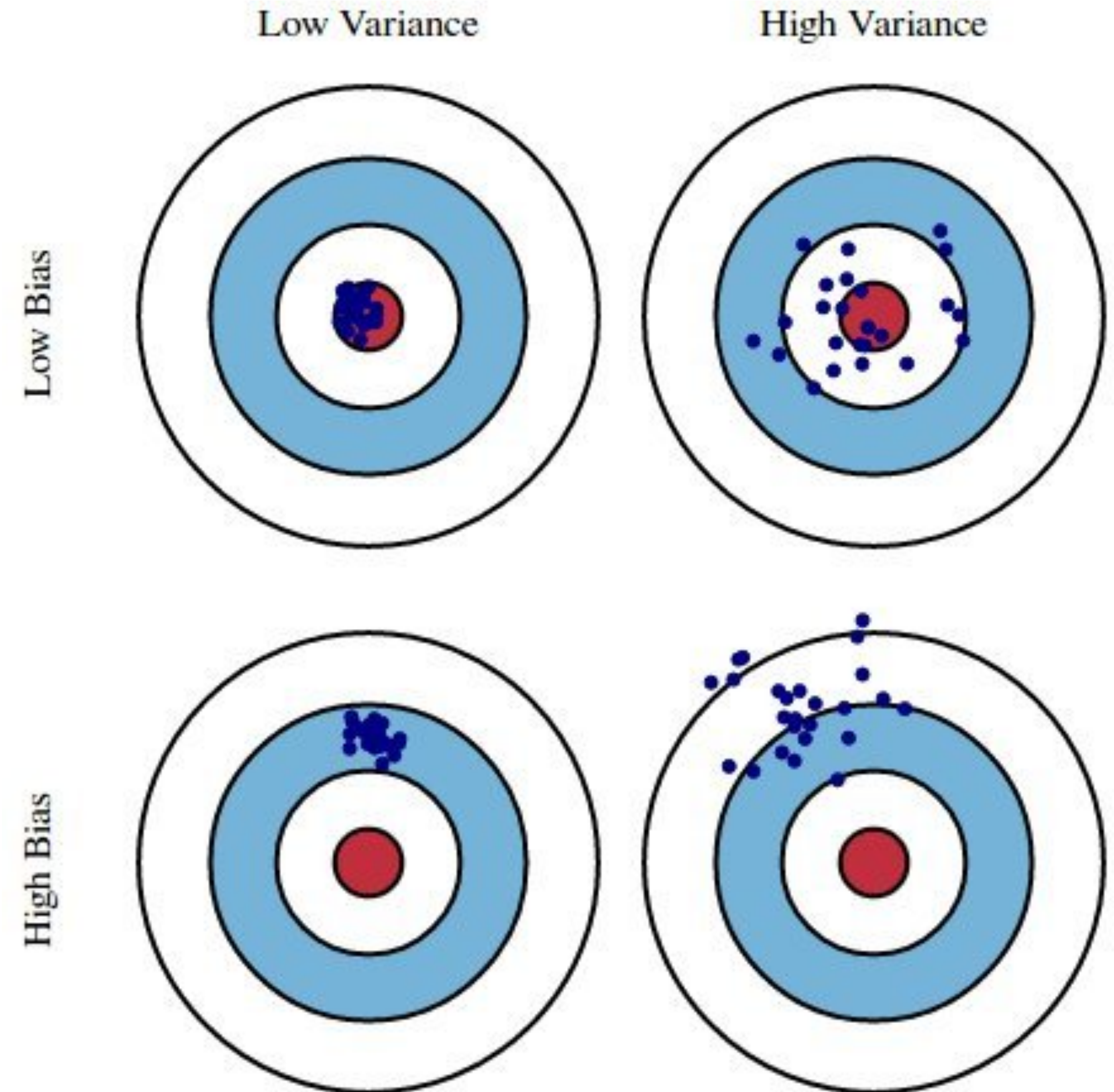
Compensación Sesgo-Varianza

- A veces tendremos una buena distribución de datos de entrenamiento por lo que predecimos bastante bien y estamos cerca del centro. Otras veces nuestros datos de entrenamiento pueden estar llenos de datos atípicos, o valores no estándar, por lo que resultan predicciones más pobres.



Compensación Sesgo-Varianza

- Estas realizaciones resultan en una distribución de hits en el blanco.
- A lo que queremos llegar es a algo con sesgo bajo y varianza pequeña.
- Pero este es un proceso de “quita y pone” por lo que tendremos que llegar a un acuerdo.



Compensación Sesgo-Varianza

- Una tentación común para principiantes es la de continuamente agregar complejidad al modelo hasta que se ajuste muy bien a los datos de entrenamiento.

Compensación Sesgo-Varianza

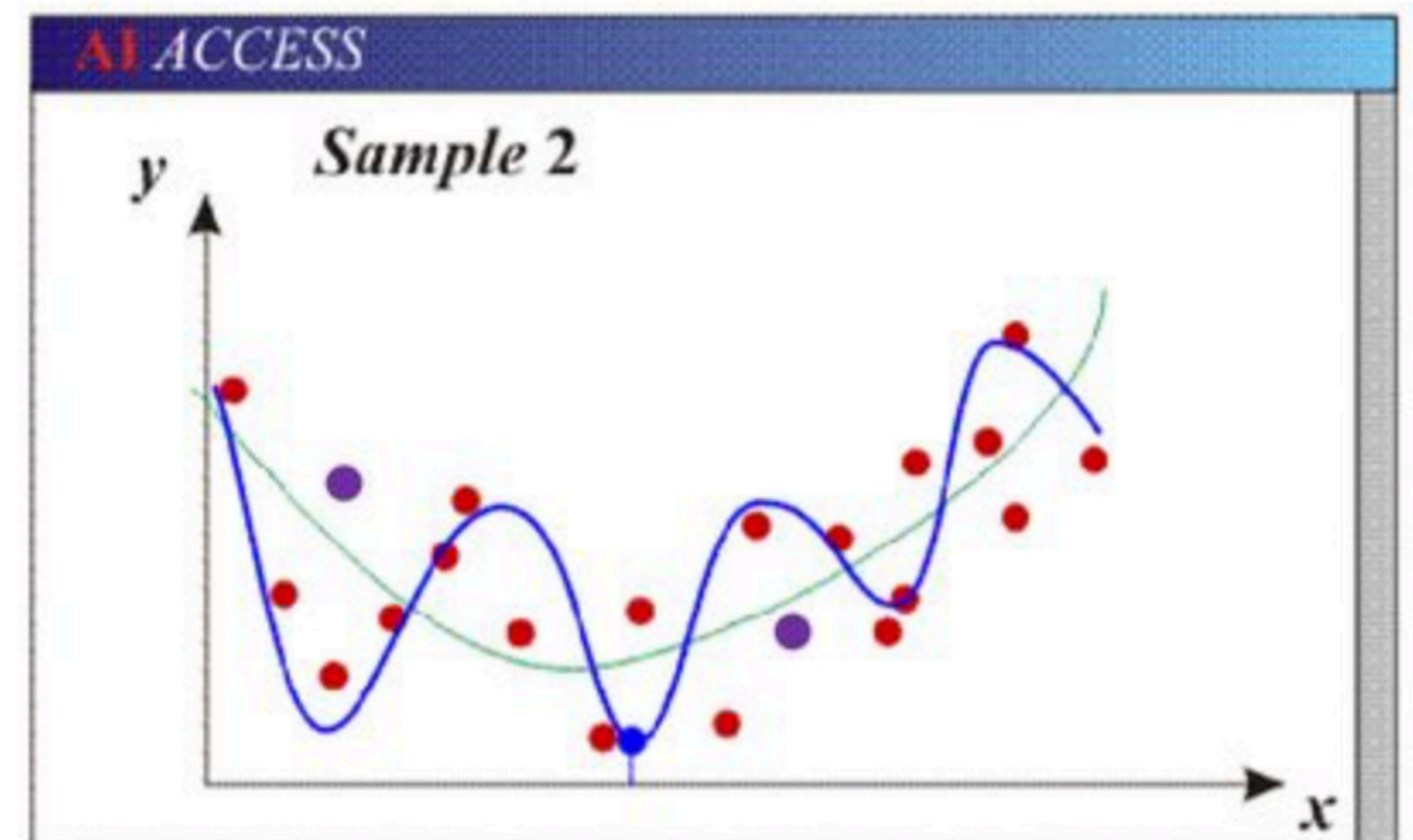
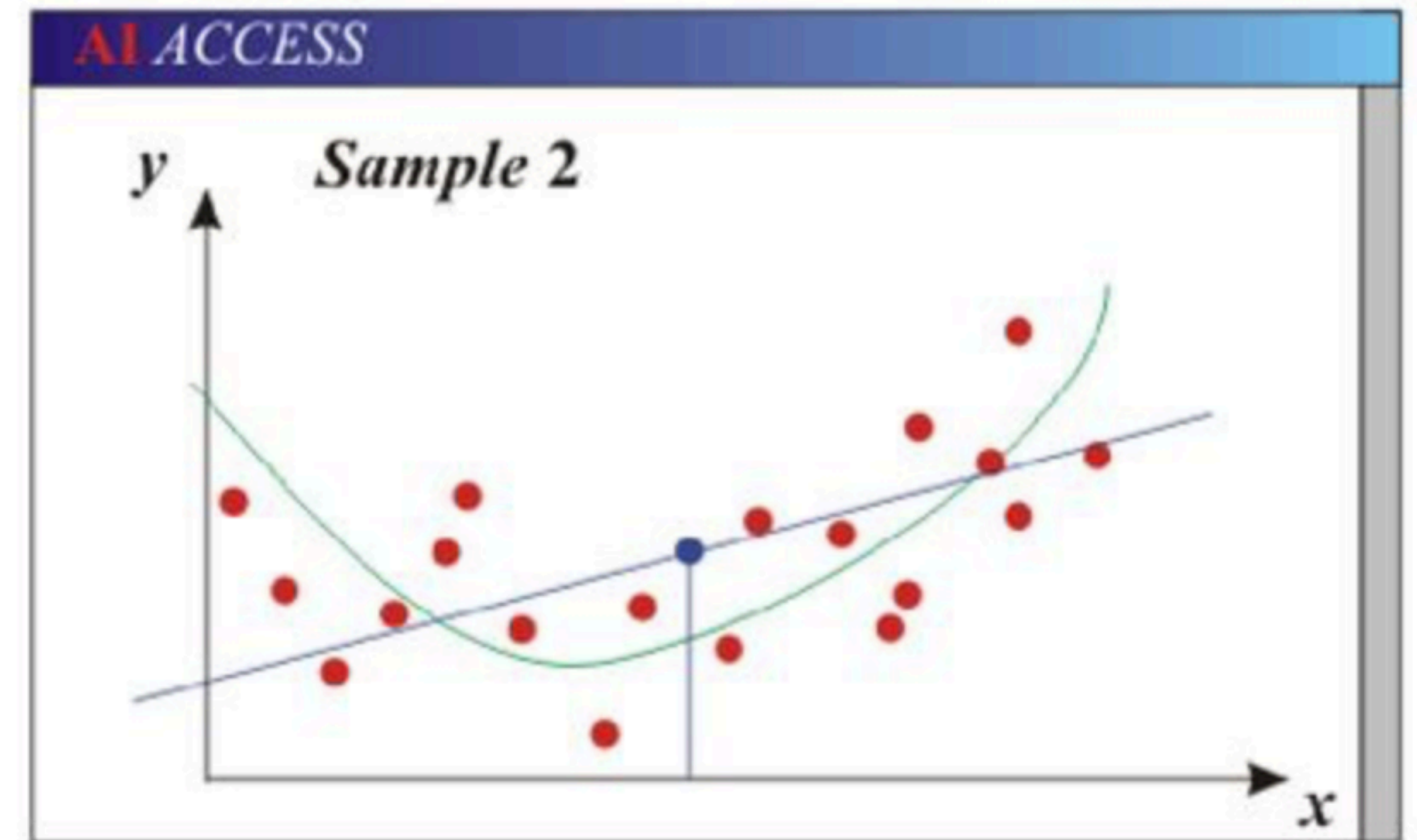
- Los modelos con demasiados pocos parámetros son inexactos debido a un sesgo alto (no hay suficiente flexibilidad)
- Los modelos con demasiados parámetros son inexactos debido a una varianza alta (demasiada sensibilidad a los datos de entrenamiento)

Puntos rojos - datos de entrenamiento (lo que vemos antes de entregar el modelo)

Curva verde - el modelo subyacente verdadero

Curva azul - nuestro modelo predicho/ajuste

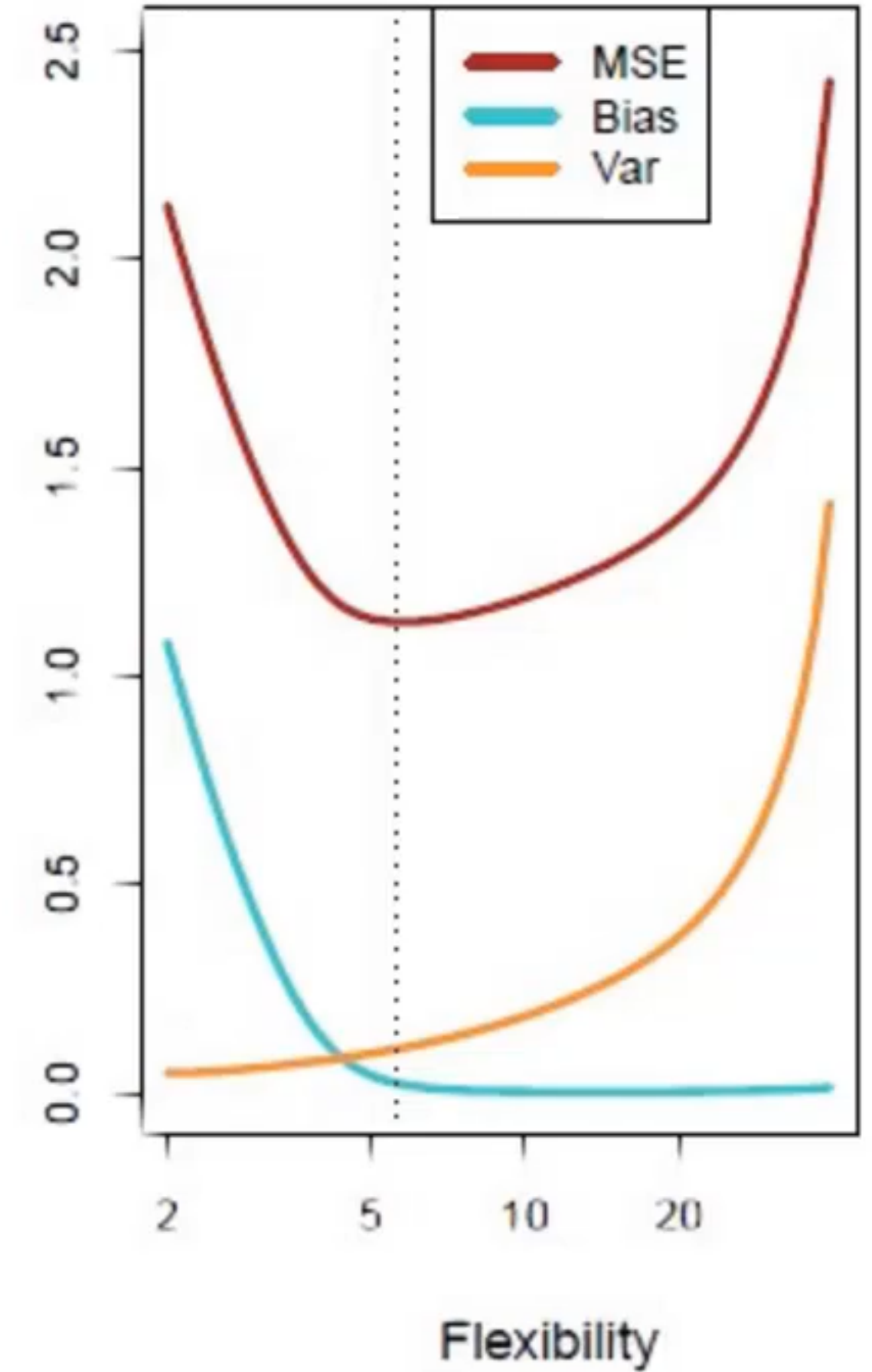
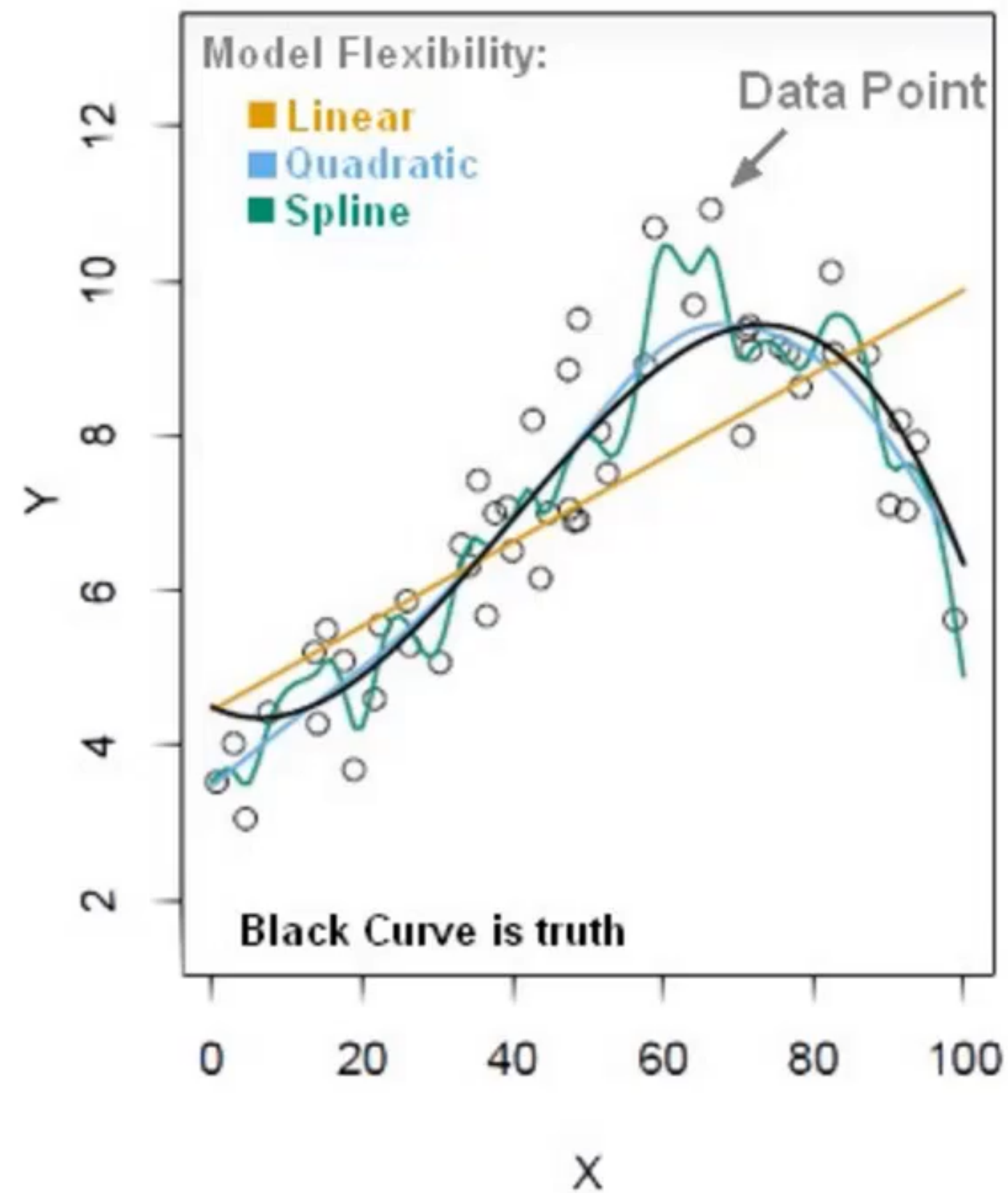
Puntos morados - puntos posibles de prueba



Compensación Sesgo-Varianza

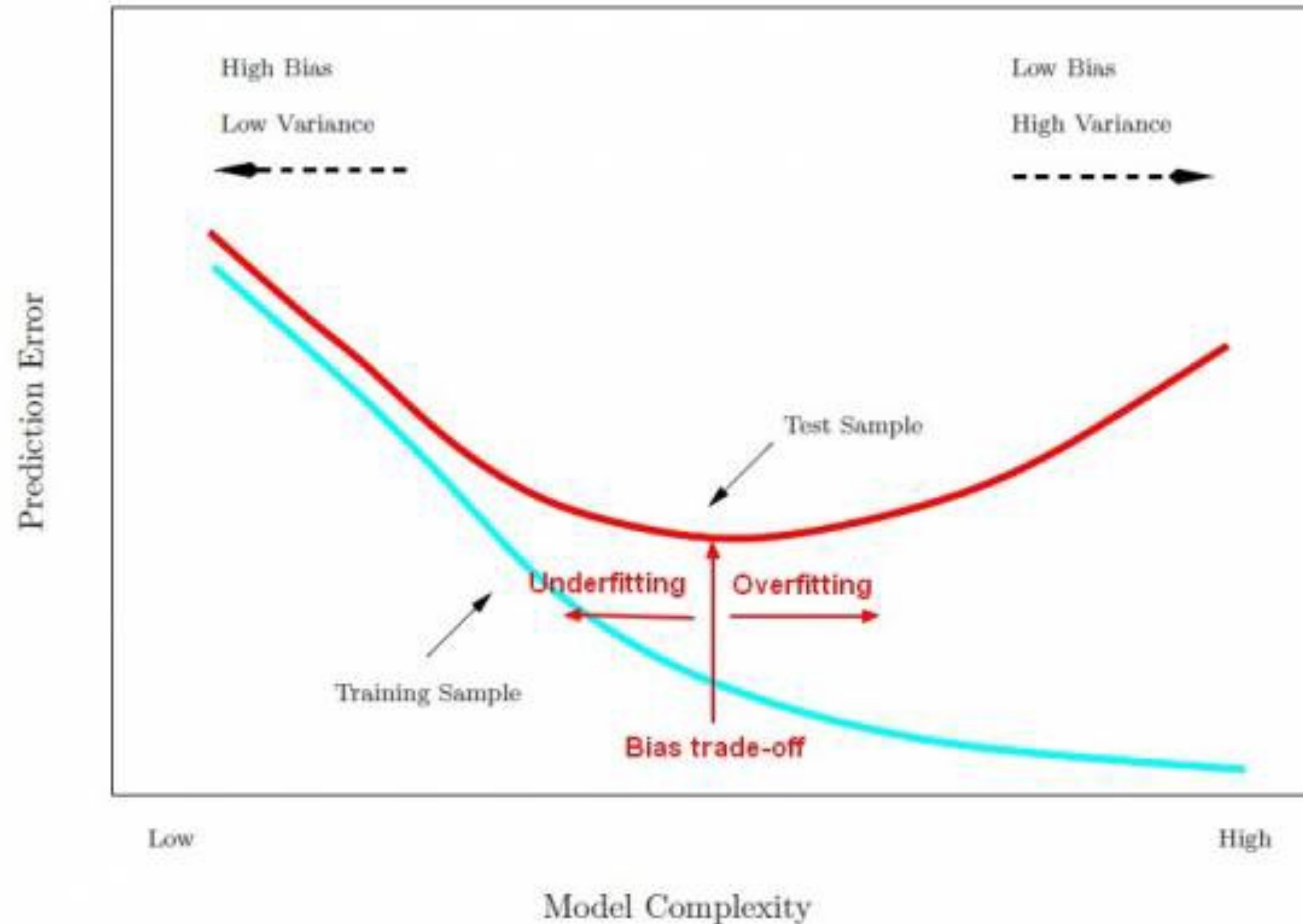
- Tratar de hacer el modelo complejo puede causar que se sobre-ajuste (overfit) a los datos de entrenamiento, causando errores grandes cuando hay datos nuevos, tales como los datos de prueba
- Veamos un ejemplo de modelo con el que podemos ver ocurrir el sobre-ajuste, visto desde el punto de vista de error, al usar datos de prueba.
- La curva negra con algunos puntos de “ruido” representa la forma verdadera de los datos.

Compensación Sesgo-Varianza



Por ej. Grados del polinomio

Compensación Sesgo-Varianza



Compensación Sesgo-Varianza

Esto fue solo un poco de teoría, más bien intuición, sobre un tema importantísimo cuando se están usando datos del mundo real. Debemos tenerlo en mente, siempre!

Todas las gráficas en esta presentación fueron obtenidas con una búsqueda en Google, de la expresión “bias variance tradeoff machine learning”

[https://www.google.com/search?
q=bias+variance+tradeoff+machine+learning&tbm=isch&ved=2ahUKEwiogfnK5JLy
AhVLtVMKHW4QAacQ2-](https://www.google.com/search?q=bias+variance+tradeoff+machine+learning&tbm=isch&ved=2ahUKEwiogfnK5JLyAhVLtVMKHW4QAacQ2-)