**Guía definitiva a Bias-Variance tradeoff**

Uno de los conceptos más importantes para poder obtener un óptimo resultado de nuestro modelo es el Bias-Variance tradeoff, por lo que es importante comprenderlo.

*Nivel: básico*



Este concepto también es conocido como Compensación Bias-Varianza.

Cuando creamos un modelo de machine learning deseamos que sea lo más eficiente posible para obtener los mejores resultados, pero para lograrlo tenemos que lograr un balance entre el Bias y Variance. Esto nos lleva a la compensación Bias-Variance.

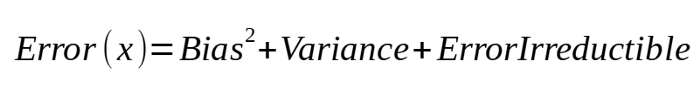
Tanto Bias como Variance son errores en nuestro modelo y buscamos reducirlos, pero no resulta tan sencillo. Cuando logramos reducir Bias notaremos que Variance incrementa, si reducimos variance entonces Bias incrementa. Esto nos lleva a una situación de balance donde ambos deben de ser lo mas pequeños posibles.

El Bias o sesgo en un modelo de machine learning es muy fácil de entender, simplemente es un tipo de error que indica la diferencia que existe entre la predicción del modelo y el valor actual. Si lo vemos desde la perspectiva de la estadística es la tendencia a sobrestimar o subestimar un parámetro. Si el modelo tiene un Bias alto significa que le presta poca atención a los datos y sobre simplifica el modelo. Esto nos lleva a tener un error alto tanto en training como en test.

Variance o varianza también se considera un error en el modelo de machine learning que tenemos. Este error lo podemos entender como que tan sensible es nuestro modelo a los datos. Si tenemos exceso de sensibilidad el modelo puede creer ver patrones que realmente no existen ahí. En el contexto de estadística es una medida de dispersión de los datos, es la distancia de cada variable a la media de todas las variables. Si tenemos un valor de Variance alto esto significa que el modelo le presta mucha atención a los datos de entrenamiento y no va a generalizar bien en datos que no ha visto. Esto lo podemos ver cuando en el entrenamiento el modelo funciona muy bien pero tiene mucho error en el test.

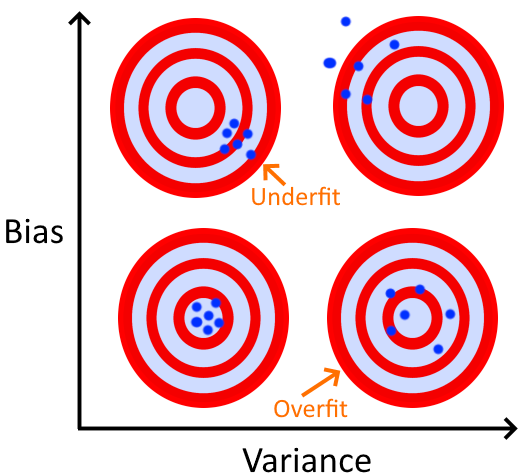
No es posible seleccionar uno sobre otro para reducir, ya que si nos concentramos en uno en particular el modelo puede generar overfit o underfit. En ambos casos nuestro modelo va a tener problemas. La única opción que tenemos es encontrar el punto donde ambos puedan ser lo más pequeños posible.

Tenemos que tomar en cuenta que existe otro tipo de error que no puede ser reducido al crear modelos, sin importar que tan buenos sean. Este error se conoce como error irreductible y simplemente es un medida de la cantidad de ruido que tenemos en nuestros datos. El error total el modelo será:



Cálculo del error total.

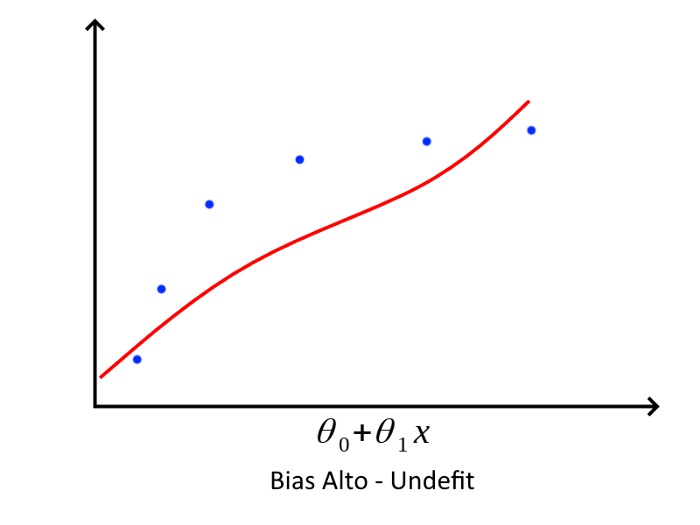
Una imagen clásica para explicar este concepto consiste de cuatro blancos (dianas) donde podemos ver el efecto de Bias y Variance.



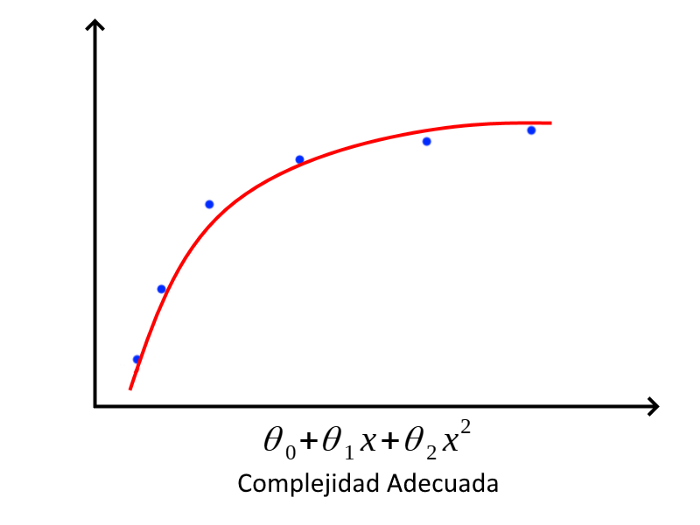
Representación de Bias y Variance

Al observar la imagen podemos entender el comportamiento de cada tipo de error. Si el Bias es pequeño nos encontramos cerca del centro, pero si se incrementa nos alejamos de él. Con la varianza podemos observar que si el valor es pequeño los puntos se encuentran cercanos entre sí, pero al aumentar los puntos se encuentran dispersos. Solamente si bajamos el error en ambos podemos tener los mejores resultados.

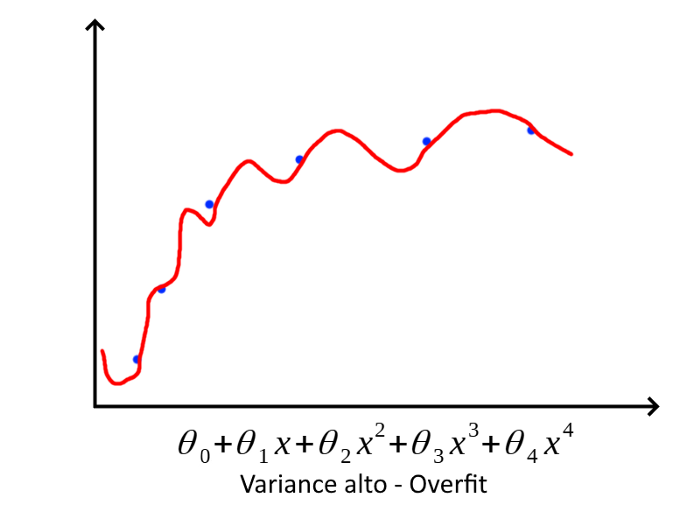
Al bajar el Bias y Variance reducimos el error total del modelo, lo que nos acerca a tener un modelo óptimo. Al lograr esto tenemos un modelo que no es muy complejo o muy sencillo. ¿Por qué es esto importante? Si tenemos un modelo complejo tendemos a tener overfitting, pero si tenemos un modelo sencillo lo que lograremos es underfitting.



Comportamiento del modelo con Bias alto

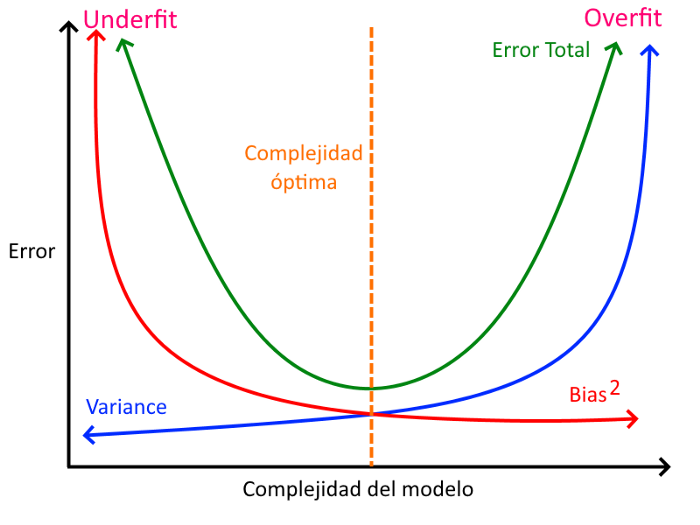


Comportamiento del modelo con una complejidad adecuada



Comportamiento del modelo con un Variance alto

Podemos entender esto más fácilmente en la siguiente imagen.



Relación entre Bias, Variance y Error total

La curva verde nos indica como se modifica el valor del error al cambiar el valor de Bias o de Variance. Si observamos hay un punto donde el error total del modelo es más bajo. En ese punto el valor de Bias y Variance nos permite tener la complejidad óptima del modelo, lo que significa que tendremos el menor error. Todo este concepto es la compensación Bias-Variance o Bias-Variance tradeoff.

Es importante conocer que puede producir el underfitting y el overfitting ahora que ya tenemos los conceptos de Bias y Variance claros.

**¿Cuando sucede el underfitting?**

Puede suceder cuando tenemos una cantidad baja de parámetros en el modelo y no se puede capturar la esencia de los datos de entrenamiento.

Cuando tenemos un Bias alto y Variance bajo.

También puede ocasionarse cuando tenemos tan pocos datos que no se puede construir un modelo adecuado.

Si intentamos construir un modelo lineal cuando los datos no son lineales también surge este problema.

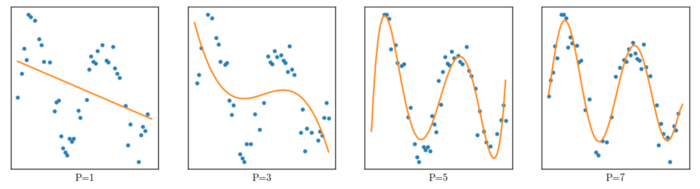
**¿Cuando sucede el overfitting?**

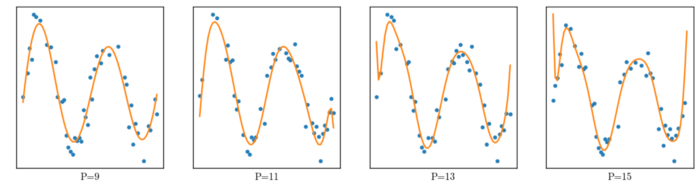
Cuando tenemos muchos predictores o parámetros lo cual ocasiona que se capture el ruido junto con el patrón de los datos. Esto hace que el modelo se ajuste mucho a los datos de entrenamiento perdiendo generalidad para poder trabajar con datos no vistos.

Si tenemos un valor bajo de Bias y alto de Variance.

Si tenemos un modelo muy complejo para representar datos más sencillos.

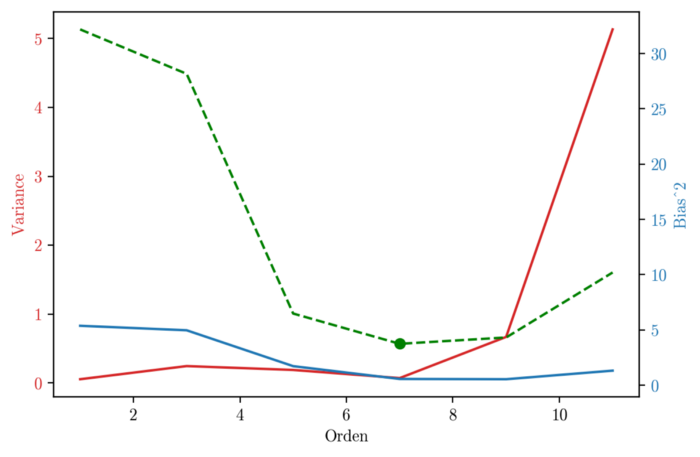
Observemos este comportamiento en un modelo real.





Modelos con diferentes complejidades para los mismos datos

El valor de P indica el grado del polinomio que estamos usando para crear el modelo. Podemos observar que para P=1 y P=2, el modelo no se ajusta bien a los datos, tenemos underfitting. En el caso de P=11 hasta P=15, el modelo trata de exagerar el ajuste a los datos lo que nos indica overfitting. Los demás modelos parecen tener un ajuste adecuado, pero ¿cuál es el mejor?, eso lo podemos saber si observamos el error total de todos los modelos.



Comportamiento de Bias, Variance y Error total para los modelos

En color azul vemos el comportamiento de Bias, tiende a disminuir conforme el modelo se hace más complejo. La varianza para estos datos incrementa al tener un modelo más complejo, se nota un incremento muy fuerte al pasar de P=9 a P=10. El error total del modelo esta marcado en verde, notamos que disminuye y después empieza a incrementar. El punto más bajo de error es cuando P=7, lo que nos lleva a considerar ese orden como la complejidad óptima del modelo.

Los conceptos aprendidos funcionan cuando trabajamos con aprendizaje supervisado. Para aprendizaje reforzado los conceptos cambian, por lo que es un tema aparte. Si te gustaría aprender sobre como utilizar Bias y Variance para aprendizaje reforzado, escribe un comentario y comparte este artículo.

**Conclusiones**

Al conocer como actúan Bias y Variance podemos encontrar el nivel de complejidad adecuado para que nuestro modelo evite en lo posible problemas como everfitting y underfitting y de esta manera obtener los mejores resultados.

Para ver el video de esta lección en YouTube visita <https://youtu.be/HASbddBQp0c>

Si deseas ver el video con una explicación más a fondo, lo puedes hacer apoyándome en Patreon: <https://www.patreon.com/nicosio>