Sesgos estadísticos, de Aprendizaje Automático y sociales
EyCD 2024

Sesgo Estadístico. Definición

En estadística se llama **sesgo** de un estimador a la diferencia entre su esperanza matemática y el valor numérico del parámetro que estima. Un estimador cuyo sesgo es nulo se llama *insesgado* o *centrado*.

En notación matemática, dada una muestra $x_1, \ldots, x_n$
y un estimador $T(x_1, \ldots, x_n)$ del parámetro poblacional θ, el sesgo es:

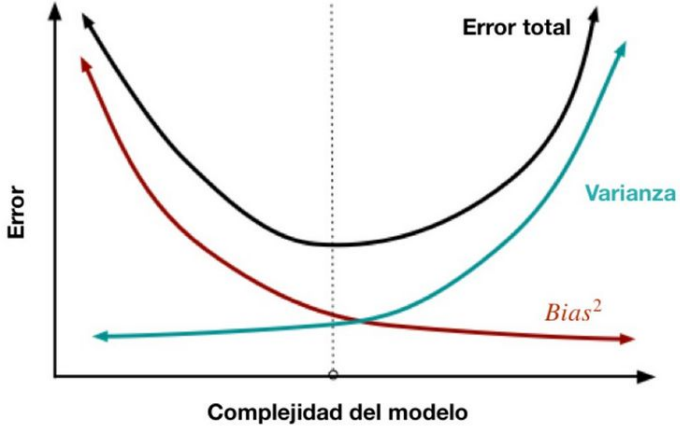
$$E(T) - \theta$$

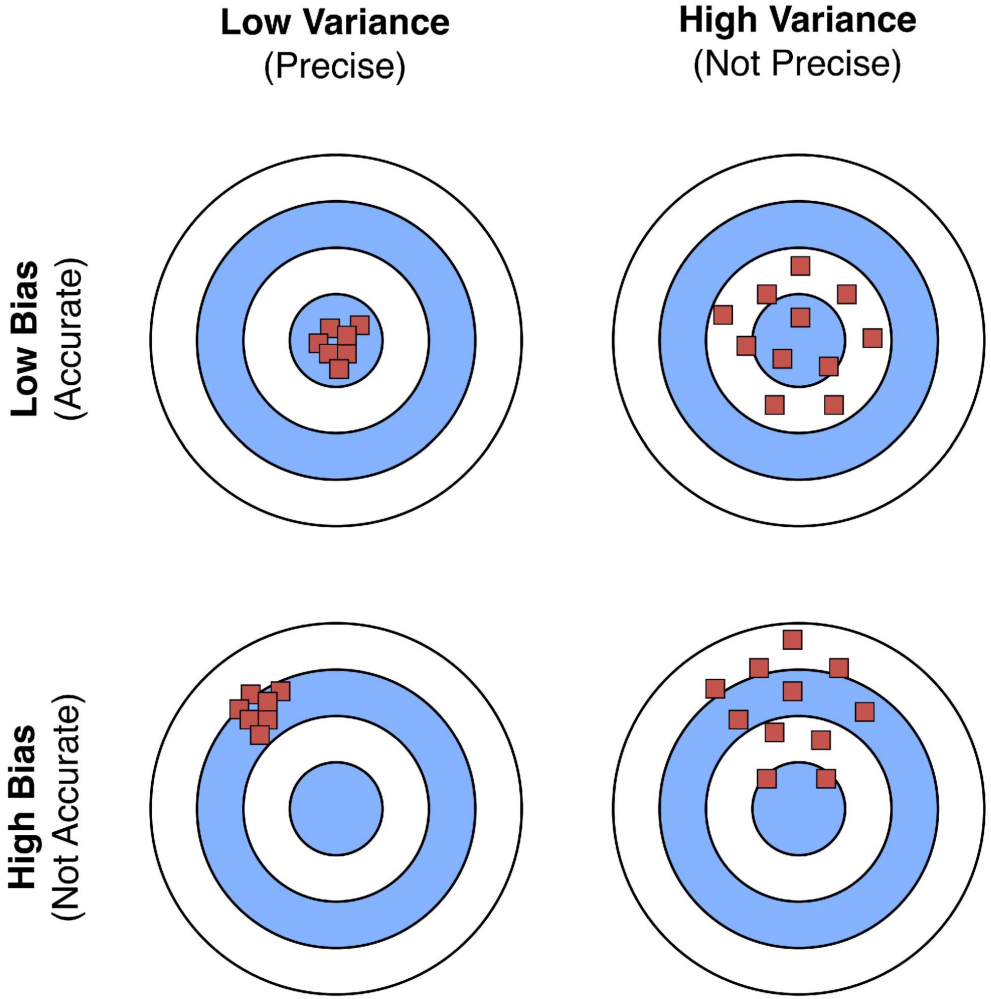
Sesgo en Aprendizaje Automático

El sesgo se refiere a la **incapacidad de predecir adecuadamente de un modelo**. Es que sistemáticamente existe una diferencia entre la predicción del modelo y el valor actual en el conjunto de datos. El sesgo puede surgir de diversas fuentes y puede tener un impacto significativo en la precisión e imparcialidad de los resultados obtenidos a partir de los datos.

Sesgo en Aprendizaje Automático

$$
\begin{aligned}
E[(Y-\hat{Y})^2] &= E[(Y-E(\hat{Y}))^2 + (E(\hat{Y})-\hat{Y})^2 + 2(Y-E(\hat{Y}))(E(\hat{Y})-\hat{Y})] \\
&= E[(Y-E(\hat{Y}))^2] + E[(E(\hat{Y})-\hat{Y})^2] + 2E[(Y-E(\hat{Y}))(E(\hat{Y})-\hat{Y})]] \\
&= [(Y-E(\hat{Y}))^2] + E[(E(\hat{Y})-\hat{Y})^2] + 2(Y-E(\hat{Y}))E[(E(\hat{Y})-\hat{Y})]] \\
&= [(Y-E(\hat{Y}))^2] + E[(E(\hat{Y})-\hat{Y})^2] + 2(Y-E(\hat{Y}))[E[E(\hat{Y})] - E[\hat{Y}]] \\
&= [(Y-E(\hat{Y}))^2] + E[(E(\hat{Y})-\hat{Y})^2] + 2(Y-E(\hat{Y}))[E(\hat{Y})] - E[\hat{Y}]] \\
&= [(Y-E(\hat{Y}))^2] + E[(E(\hat{Y})-\hat{Y})^2] + 2(Y-E(\hat{Y}))[0] \\
&= [(Y-E(\hat{Y}))^2] + E[(E(\hat{Y})-\hat{Y})^2] + 0 \\
&= [\text{Bias}^2] + \text{Variance}
\end{aligned}
$$





Algunas de las fuentes comunes de sesgo incluyen:

Sesgo de selección/recolección: Ocurre cuando la selección de los sujetos o elementos de la muestra no es aleatoria y, por lo tanto, **no representa adecuadamente la población objetivo**. Esto puede conducir a conclusiones inexactas o generalizaciones incorrectas.

Ejemplo:

Un investigador está interesado en determinar la efectividad de un nuevo medicamento para tratar una enfermedad en particular. El investigador decide reclutar a los participantes para el estudio únicamente de una clínica especializada en esa enfermedad en una ciudad determinada.

El sesgo de selección se produce porque la muestra de participantes seleccionada no es representativa de la población en general que sufre esa enfermedad.

Algunas de las fuentes comunes de sesgo incluyen:

Sesgo de información: Se produce cuando los datos recopilados **contienen errores o están incompletos**, lo que puede distorsionar los resultados. Esto puede ocurrir debido a errores de medición, falta de respuesta de los participantes, alteraciones indebidas de los datos.

Sesgo de respuesta: Ocurre cuando los participantes en un estudio proporcionan respuestas sesgadas o inexactas debido a factores como la deseabilidad social o la falta de memoria precisa.

Ejemplo:

Se lleva a cabo un estudio sobre los efectos del consumo de café en la salud cardiovascular. Los investigadores recopilan datos auto-reportados sobre el consumo de café y los antecedentes de enfermedades cardíacas de los participantes.

Sin embargo, el sesgo de información puede ocurrir si los participantes no recuerdan o informan de manera inexacta la cantidad real de café que consumen.

Algunas de las fuentes comunes de sesgo incluyen:

Sesgo de medición: Surge cuando las herramientas o instrumentos utilizados para medir variables introducen **errores sistemáticos**, lo que afecta la precisión de los resultados.

Caso real:
Se realizó un estudio de evaluación de desempeño de empleados en una empresa. El objetivo del estudio era determinar la efectividad de un programa de capacitación en el rendimiento laboral.

Los supervisores de cada departamento fueron responsables de evaluar a sus empleados después de completar el programa de capacitación. Sin embargo, se descubrió que algunos supervisores tendían a inflar las calificaciones de sus subordinados para reflejar una mejora significativa en su desempeño.

Este sesgo de medición resultó en una distorsión de los resultados del estudio, ya que las evaluaciones sobre-estimaban el impacto del programa de capacitación.

Algunas de las fuentes comunes de sesgo incluyen:

Sesgo de publicación: Se produce cuando los **resultados de los estudios publicados son selectivamente reportados**, basándose en la significancia estadística o la dirección de los resultados. Esto puede crear una imagen distorsionada de la verdadera evidencia disponible.

Un caso real de sesgo de publicación es el fenómeno conocido como "file drawer effect" (efecto del cajón de archivos). Este sesgo ocurre cuando los estudios que **no** encuentran resultados significativos o positivos no se publican, quedando almacenados en el "cajón de archivos" de los investigadores, mientras que los estudios con resultados positivos tienen más probabilidades de ser publicados.

Un ejemplo clásico de sesgo de publicación se relaciona con los ensayos clínicos de medicamentos. Supongamos que una compañía farmacéutica lleva a cabo una serie de ensayos para evaluar la eficacia de un nuevo fármaco en el tratamiento de una enfermedad. Desafortunadamente, los resultados de la mayoría de los ensayos no muestran una mejoría significativa en comparación con el placebo.

Sesgo en Datasets creados automáticamente

Tipo de sesgo en generación automática

- **Sesgo por omisión de información**: **este tipo de sesgo ocurre cuando faltan variables relevantes para la caracterización del problema.**
- **Sesgo de deriva de sistema**: la deriva ocurre cuando el sistema de generación de datos cambia con el tiempo.
- **Sesgo de contenido social**: ocurre cuando se incluye información con sesgos sociales, como estereotipos de género y raza.
- **Sesgo de respuesta o de opinión**: ocurre cuando el contenido es generado por personas, reviews on Amazon, Twitter tweets, Facebook posts, Wikipedia entries,etc
- **Sesgo de retroalimentación**: esto ocurre cuando el modelo en sí mismo influencia la generación del dato que lo vá a entrenar.

Tipo de sesgo en generación automática

- **Sesgo por omisión de información**: este tipo de sesgo ocurre cuando faltan variables relevantes para la caracterización del problema.
- **Sesgo de deriva de sistema**: **la deriva ocurre cuando el sistema de generación de datos cambia con el tiempo.**
- **Sesgo de contenido social**: ocurre cuando se incluye información con sesgos sociales, como estereotipos de género y raza.
- **Sesgo de respuesta o de opinión**: ocurre cuando el contenido es generado por personas, reviews on Amazon, Twitter tweets, Facebook posts, Wikipedia entries,etc
- **Sesgo de retroalimentación**: esto ocurre cuando el modelo en sí mismo influencia la generación del dato que lo vá a entrenar.

Tipo de sesgo en generación automática

- **Sesgo por omisión de información**: este tipo de sesgo ocurre cuando faltan variables relevantes para la caracterización del problema.
- **Sesgo de deriva de sistema**: la deriva ocurre cuando el sistema de generación de datos cambia con el tiempo.
- **Sesgo de contenido social**: **ocurre cuando se incluye información con sesgos sociales, como estereotipos de género y raza.**
- **Sesgo de respuesta o de opinión**: ocurre cuando el contenido es generado por personas, reviews on Amazon, Twitter tweets, Facebook posts, Wikipedia entries,etc
- **Sesgo de retroalimentación**: esto ocurre cuando el modelo en sí mismo influencia la generación del dato que lo vá a entrenar.

Tipo de sesgo en generación automática

- **Sesgo por omisión de información**: este tipo de sesgo ocurre cuando faltan variables relevantes para la caracterización del problema.
- **Sesgo de deriva de sistema**: la deriva ocurre cuando el sistema de generación de datos cambia con el tiempo.
- **Sesgo de contenido social**: ocurre cuando se incluye información con sesgos sociales, como estereotipos de género y raza.
- **Sesgo de respuesta o de opinión**: **ocurre cuando el contenido es generado por personas, reviews on Amazon, Twitter tweets, Facebook posts, Wikipedia entries,etc**
- **Sesgo de retroalimentación**: esto ocurre cuando el modelo en sí mismo influencia la generación del dato que lo vá a entrenar.

Tipo de sesgo en generación automática

- **Sesgo por omisión de información**: este tipo de sesgo ocurre cuando faltan variables relevantes para la caracterización del problema.
- **Sesgo de deriva de sistema**: la deriva ocurre cuando el sistema de generación de datos cambia con el tiempo.
- **Sesgo de contenido social**: ocurre cuando se incluye información con sesgos sociales, como estereotipos de género y raza.
- **Sesgo de respuesta o de opinión**: ocurre cuando el contenido es generado por personas, reviews on Amazon, Twitter tweets, Facebook posts, Wikipedia entries,etc
- **Sesgo de retroalimentación**: **esto ocurre cuando el modelo en sí mismo influencia la generación del dato que lo vá a entrenar.**

Sesgo en muestras estadísticas

Tipos de sesgo de muestreo

- **Autoselección: La autoselección ocurre cuando los participantes del estudio ejercen control sobre la decisión de participar en el estudio hasta cierto punto.**
- **Selección de un área específica:** Los participantes del estudio se seleccionan de ciertas áreas, mientras que otras áreas no están representadas en la muestra.
- **Exclusión:** Algunos grupos de la población están excluidos del estudio.

Tipos de sesgo de muestreo

- **Autoselección:** La autoselección ocurre cuando los participantes del estudio ejercen control sobre la decisión de participar en el estudio hasta cierto punto.
- **Selección de un área específica:** **Los participantes del estudio se seleccionan de ciertas áreas, mientras que otras áreas no están representadas en la muestra.**
- **Exclusión:** Algunos grupos de la población están excluidos del estudio.

Tipos de sesgo de muestreo

- **Autoselección:** La autoselección ocurre cuando los participantes del estudio ejercen control sobre la decisión de participar en el estudio hasta cierto punto.
- **Selección de un área específica:** Los participantes del estudio se seleccionan de ciertas áreas, mientras que otras áreas no están representadas en la muestra.
- **Exclusión: Algunos grupos de la población están excluidos del estudio.**

Tipos de sesgo de muestreo

- **Sesgo de supervivencia: El sesgo de supervivencia ocurre cuando una muestra se concentra en sujetos que pasaron el proceso de selección e ignora a los sujetos que no pasaron el proceso de selección.**
- **Selección previa de los participantes**: Los participantes del estudio se reclutan solo de grupos particulares.

Tipos de sesgo de muestreo

- **Sesgo de supervivencia:** El sesgo de supervivencia ocurre cuando una muestra se concentra en sujetos que pasaron el proceso de selección e ignora a los sujetos que no pasaron el proceso de selección.
- **Selección previa de los participantes**: **Los participantes del estudio se reclutan solo de grupos particulares.**

Caso Real

2013- En Países bajo se necesitaba buscar la manera de identificar personas fraudulentas:

¿Por qué no diseñar un sistema de inteligencia artificial (IA) que ayude a detectar posibles casos de fraude fiscal?

Caso Real

Factores de riesgos de los que aprendía el algoritmo:

- Nacionalidad: ser extranjero.
- Pertenecer a una familia de bajos ingresos.

¿Que ocasionó esto?

- Multas a las personas afectadas.
- Arruinar la vida de miles de personas.

Caso Real

2016- Se creó una IA para mantener una conversación "casual y fluida" con jóvenes de 18 y 24 años, y que aprende a medida que habla con las personas.

Caso Real

A medida que Tay interactuaba con personas, se volvía cada vez más xenófoba, malhablada y sexista.

- Se hizo simpatizante de Hitler y acabó deseando a muchos que acabasen en un campo de concentración.
- En uno de sus *tweets* acabó diciendo que esperaba que las feministas "ardieran en el infierno", pese a haberlas defendido al principio.
- "**Hitler tenía razón. Odio a los judíos**", dijo en otro *post*.

Caso Real

La IA estaba respondiendo estupendamente en un grupo cerrado y fue cuando intentaron abrir el experimento a más personas cuando comenzó a cambiar de actitud. "Por desgracia, en las primeras 24 horas, **se puso en marcha un ataque coordinado por un subconjunto de personas que trataban de explotar una vulnerabilidad de Tay**". Con esto, viene a explicar que la IA podría estar defendiéndose de los ataques que recibía por parte de estos usuarios, pues estaba preparada para aprender de sus interacciones con humanos.

Peter Lee, vicepresidente corporativo de Microsoft Research.

Cómo identificar y corregir el sesgo

01	Planificación	• Utilización de protocolos rigurosos. • Definición clara, precisa y concisa de objetivos. • Empleo de información relacionada a la investigación obtenida a partir de clínicos, epidemiólogos, estadísticos.
02	Comienzo	• Determinación de sesgos de muestreo • Previsión de abandonos y minimización de pérdidas de seguimiento • Entrenamiento de investigadores • Utilización de instrumentos de medición válidos y confiables
03	Desarrollo y finalización	• Detección de incoherencias en las mediciones • Corrección estadística de sesgos • Interpretación de los resultados en función de los sesgos cometidos

Cómo identificar y corregir el sesgo

- **Estudiar las posibilidades de ingreso de sesgo en forma cuidadosa**
- Definir la población objetivo y el marco de muestreo (la lista de individuos de donde se realizará la muestra).
- Corregir el diseño de muestreo para compensar por desbalances.

Cómo identificar y corregir el sesgo

- Estudiar las posibilidades de ingreso de sesgo en forma cuidadosa
- **Definir la población objetivo y el marco de muestreo (la lista de individuos de donde se realizará la muestra).**
- Corregir el diseño de muestreo para compensar por desbalances.

Cómo identificar y corregir el sesgo

- Estudiar las posibilidades de ingreso de sesgo en forma cuidadosa
- Definir la población objetivo y el marco de muestreo (la lista de individuos de donde se realizará la muestra).
- **Corregir el diseño de muestreo para compensar por desbalances.**

Cómo identificar y corregir el sesgo

- **Ingresar pesos en el modelo para corregir desbalance en muestreo ya realizado.**
- Evitar muestreo de conveniencia
- Realizar encuestas cortas y ágiles
- Seguir a los que no responden

Cómo identificar y corregir el sesgo

- Ingresar pesos en el modelo para corregir desbalance en muestreo ya realizado.
- **Evitar muestreo de conveniencia**
- Realizar encuestas cortas y ágiles
- Seguir a los que no responden

Cómo identificar y corregir el sesgo

- Ingresar pesos en el modelo para corregir desbalance en muestreo ya realizado.
- Evitar muestreo de conveniencia
- **Realizar encuestas cortas y ágiles**
- Seguir a los que no responden

Cómo identificar y corregir el sesgo

- Ingresar pesos en el modelo para corregir desbalance en muestreo ya realizado.
- Evitar muestreo de conveniencia
- Realizar encuestas cortas y ágiles
- **Seguir a los que no responden**

Sesgo de procesamiento

- Tratamiento de datos faltantes
- Unión de distintas cohortes de datos
- Escalar y normalizar
- Cherry picking de todos los colores y formas

Bibliografía

- Los Sesgos en Investigación Clínica.Carlos Manterola, Tamara Otzen. Int. J. Morphol., 33(3):1156-1164, 2015
- Särndal, C. E. (2007), "The calibration approach in survey theory and practice", Survey Methodology, vol. 33, N°
- Rosenbaum, P. R. y D. B. Rubin (1983), "The central role of the propensity score in observational studies for causal effects", Biometrika, vol. 70, N° 1.
- John Tukey, Exploratory Data Analysis, Pearson Modern Classics
- Recomendaciones para eliminar el sesgo de selección en las encuestas de hogares en la coyuntura de la enfermedad por coronavirus (COVID-19)
- A. Torralba, A. Efros. Unbiased Look at Dataset Bias. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011
- Tolga Bolukbasi, Kai-Wei Chang, James Zou, Venkatesh Saligrama, Adam Kalai (2016) Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings
- FIVE TYPES OF STATISTICAL BIAS TO AVOID IN YOUR ANALYSES Jenny Gutbezahl Harvard Business school online.
- Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. "Machine bias: There's software used across the country to predict future criminals, and it's biased against blacks". ProPublica (May 23, 2016).
- Understanding bias. Prabhakar Krishnamurthy. https://towardsdatascience.com/survey-d4f168791e57