

VALORES SHAPLEY

Concepto desarrollado en 1951 por Lloyd Shapley para la Teoría de Juegos cooperativos. Ganó el Premio Nobel Memorial en Ciencias Económicas en 2012 por ello.

- Para cada juego cooperativo, el Valor Shapley asigna una distribución única y justa (entre los jugadores) de la ganancia total generada por la coalición de todos los jugadores. Primero se aplicó a juegos finitos y luego se extendió a juegos infinitos.
- Tiene propiedades muy interesantes. Algunas son:
 - * **Eficiencia:** la suma de los Valores Shapley de todos los agentes es igual al valor de la gran coalición (la ganancia se distribuye entre todos los agentes)
 - * **Simetría:** si los actores son equivalentes en el juego, su Valor Shapley será igual
 - * **Linealidad:** si dos juegos de coalición se combinan, cada uno con su función de ganancia propia, las ganancias derivadas de la combinación será la suma de todas las funciones de ganancia individuales
 - * **Jugador nulo:** el Valor Shapley del jugador nulo es cero (el jugador nulo es el que no aporta ganancia a la coalición)
 - * **Marginalismo:** el valor Shapley se puede definir como una función que usa solamente la contribución marginal del jugador 'i' como argumento

- Se pueden aplicar los valores Shapley como método estadístico para medir la contribución individual de cada predictor en un modelo de regresión o clasificación.
- En este contexto, los jugadores son las variables predictoras o características del modelo y la ganancia es la varianza total explicada o poder predictivo del modelo.
- Los valores Shapley aseguran una distribución justa de la ganancia total entre los predictores, donde esos valores representan la contribución de cada característica a la performance del modelo.

Sirve para explicar el aporte de cada característica a la performance del modelo, es decir asigna la importancia de la misma para dicho modelo →

Feature Importance: los valores Shaply altos implican características importantes.

- Es el único método que tiene una teoría sólida detrás.
- Es un método agnóstico frente a los modelos.
- Los cálculos son computacionalmente muy costosos. Casi siempre se llega a una solución aproximada.
- Se presta a mala interpretación: El valor de Shapley es la contribución promedio del valor de una característica a la predicción en diferentes coaliciones. El valor de Shapley NO es la diferencia en la predicción cuando eliminamos la característica del modelo.
- Una alternativa que aproxima los valores Shapley es SHAP, con dos propuestas TreeSHAP y KernelSHAP (Lundberg et al, 2018)
- En Python tenemos la biblioteca 'shap' que implementa la aproximación TreeSHAP a los valores Shapley. Está integrada en las bibliotecas 'xgboost' y 'LightGBM'

•

FUENTES:

- * https://en.wikipedia.org/wiki/Shapley_value
- * recursohumano.cl/post/el-juego-infinito-simon-sinek

* INTERPRETABLE MACHINE LEARNING:

<https://christophm.github.io/interpretable-ml-book/shapley.html>

<https://christophm.github.io/interpretable-ml-book/shap.html>