

Proyecto final

Matias Bajac

2025-11-07

Introducción

El objetivo de la inferencia conformal es cuantificar la incertidumbre de las predicción tanto para problemas de clasificación como de regresión. Fundamentalmente, lo hace convirtiendo las predicciones de un algoritmo en conjuntos de predicción, los cuales tienen fuertes propiedades de cobertura en muestras finitas.

Objetivos

Sea $(X_i, Y_i) \sim P, i = 1, \dots, n$ iid , las variables explicativas y dependiente de una distribucion P. Dado una probabilidad α llamado tasa de no cobertura, queremos encontrar una banda de predicción $CR^n XR$

$$\hat{C}_n : \mathcal{X} \rightarrow \{\text{subconjunto de } \mathcal{Y}\}$$

Con la propiedad que para un nuevo par de datos $(X_{n+1}, Y_{n+1}) \sim P$

$$P(Y_{n+1} \in \hat{C}_n(x_{n+1})) \geq 1 - \alpha \quad (1)$$

Por otra parte, si no asumimos ninguna teoría asintótica ni tampoco ninguna distribución en P, esto puede resultar bastante desafiante

Supongamos que nuestro objetivo inicial es encontrar una cola de intervalo de predicción. $\hat{C}_n = (\infty, \hat{q}_n]$

Dada esta ecuación, un punto de partida natural sería fijar \hat{q}_n , como el cuantil muestral de nivel $1 - \alpha$ de Y_1, \dots, Y_n el cual denotamos por

$$\hat{q}_n = Quantile(1 - \alpha; \frac{1}{n} \sum_{i=1}^n \delta_{Y_i})$$

donde $\frac{1}{n} \sum_{i=1}^n \delta_{Y_i}$ es la distribución empírica: pone pesos uniformes $\frac{1}{n}$ en cada dato Y_i

Mientras que $Quantile_{1-\alpha}$ devuelve el mínimo valor q tal que la función de distribución acumulada empírica alcanza al menos $1 - \alpha$

$$P(Y_{n+1} \leq \hat{q}_n) \sim 1 - \alpha$$

Esto es exacto cuando $n \rightarrow \infty$, bajo ciertas condiciones (que garantizan la convergencia del cuantil muestran al cuantil poblacional). Se puede algo que satisfaga la ecuación anterior en muestra finita?.

Usar rangos para formar quantiles ajustados

Aquí es donde comienza la primera idea detrás de la predicción conformal. Con Y_{n+1} es iid con Y_1, Y_2, \dots, Y_n , entonces

$$P(Y_{n+1} \text{esta entre los } [(1 - \alpha)(n + 1)] \text{valores mas pequeños de } Y_1, \dots, Y_{n+1} \geq 1 - \alpha)$$

que es equivalente a decir que

$$P(Y_{n+1} \text{esta entre los } [(1-\alpha)(n+1)] \text{valores mas pequeños de } Y_1, \dots, Y_n \geq 1-\alpha)$$

Método de Naive

Veamos la primera idea clave sobre regresión, donde observamos ambos X_i and $Y_i \in R$ $i = 1, \dots, n$, y queremos un conjunto de predicción para Y_{n+1} basado en X_{n+1} . Supongamos que \hat{f}_n es cualquier predictor puntual, entrenado en (X_i, Y_i) , $i = 1, \dots, n$. En otras palabras, $\hat{f}_n(x)$ predice el valor de y que esperamos observar en x

Definimos los residuos de los datos de entrenamiento,

$$e_i = |Y_i - \hat{f}_n(X_i)|, i = 1, \dots, n$$

tenemos $\hat{q}_n = [(1-\alpha)(n+1)]$ el mas chico de R_1, \dots, R_n , podemos definir el conjunto de predicción como $\hat{C}_n(x) = \{y : |y - \hat{f}_n(x)| \leq \hat{q}_n\}$

O en otras palabras:

$$\hat{C}_n(x) = [\hat{f}_n(x) - \hat{q}_n, \hat{f}_n(x) + \hat{q}_n]$$

Este método es aproximadamente válido para muestras grandes, bajo la condición de que $\hat{f}_n(x)$ sea lo suficientemente preciso, es decir, que \hat{q}_n este cerca del cuantil $1-\alpha$ de R_i . Un problema de este método es que los intervalos de predicción pueden presentar una considerable subcobertura, dado que se están empleando los residuos dentro de la muestra. Para evitar esto, se plantea la metodología de los intervalos de predicción conformales

Separación de la muestra de intervalos de predicción

En esta primera parte de esta sección, nos enfocaremos en la parte de regresión, es decir, que Y pertenece a todos los reales, luego se verá cuando la variable de respuesta es categórica (problema de clasificación).

En concreto, dividimos el conjunto de entrenamiento en 2:

T_1 , es el conjunto de entrenamiento propiamente dicho T_2 es el conjunto de calibración

Tiene sentido pensar que la intersección de ambos es vacía, $T_1 \cap T_2 = \emptyset$ y $T_1 \cup T_2 = \{1, 2, \dots, n\}$, sea $n_1 = |T_1|$ y $n_2 = |T_2|$

En el siguiente paso, entrenamos el predictor puntual usando los datos del conjunto de entrenamiento propiamente dicho (X_i, Y_i) donde $i \in T_1$ y lo denotamos como \hat{f}_{n_1} . Luego, vemos los residuos en el conjunto de calibración: $e_i = |Y_i - \hat{f}_{n_1}|$ $i \in T_2$

Por lo que llegamos a que el residuo más chico del cuantil conformal es el siguiente: $\hat{q}_{n_2} = [(1-\alpha)(n_2+1)] e_i$ $i \in R_2$. Aquí se calcula un cuantil de nivel de cobertura $1-\alpha$ de los residuos e_i , para construir un **intervalo de predicción** que tenga cobertura aproximada $1-\alpha$, es decir, primero se ordenan los residuos e_i del conjunto T_2 de menor a mayor. Luego, se busca el cuantil empírico de orden $(1-\alpha)$, es decir, el residuo en la posición $[(1-\alpha)(n_2+1)]$ donde n_2 es el tamaño del conjunto de calibración. Este valor se denota como: \hat{q}_{n_2} y se utiliza para crear un intervalo de predicción alrededor de la estimación puntual, $C(x) = [\hat{f}_{n_1}(x) - \hat{q}_{n_2}, \hat{f}_{n_1}(x) + \hat{q}_{n_2}]$

La mayor garantía que podemos obtener es que:

$$P(y_{n+1} \in \hat{C}_n(X_{n+1}) | (X_i, Y_i) i \in T_1) \in [1-\alpha, 1-\alpha + \frac{1}{n_2+1}]$$

Intervalos predictivos vía Jackknife

Aquí la idea es separar el conjunto en entrenamiento y testeo, dejando una observación aparte para testear, es decir, entrenas \hat{f}_{-i} dejando fuera la observación i , luego se calculan los residuos