(Un primer acercamiento a) Conformal Prediction

Bernardo Marenco

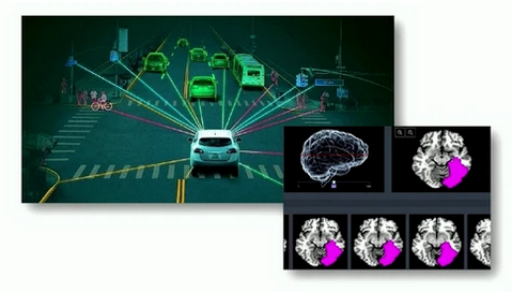
FACULTAD DE INGENIERÍA

UNIVERSIDAD DE LA REPÚBLICA URUGUAY

Seminario Optimización y Aprendizaje Automático
7 de noviembre de 2024

Conformal prediction: motivación

- Modelos de ML se utilizan en aplicaciones críticas: autos autónomos, diagnósticos médicos, etc



- Objetivo: cuantificación de incertidumbre para modelos de ML, i.e.
  - ¿Qué tan seguro estoy de mis predicciones?
  - ¿Cómo ese nivel de seguridad afecta mi decisión?
  - En la práctica, ¿puedo utilizar mi modelo con seguridad?

E. Candès, "A Taste of Conformal Prediction", *A Multiscale tour of Harmonic Analysis and Machine Learning*, IHES, abril 19-21, 2023

Introducción

- Enfoque: usar modelos de ML como caja negra y construir "intervalos de confianza" para sus predicciones
- A partir de un conjunto de entrenamiento $(X_i, Y_i) \sim P, i = 1, \ldots, n$ iid, con P distribución en $\mathcal{X} \times \mathcal{Y}$, buscamos construir una función

$$\mathcal{C}_n : \mathcal{X} \to \mathcal{P}(\mathcal{Y}) = \{\text{subconjuntos de } \mathcal{Y}\}$$

tal que para un nuevo par iid $(X_{n+1}, Y_{n+1}) \sim P$ se cumpla

$$\boxed{\mathbb{P}\left(Y_{n+1} \in \mathcal{C}_n(X_{n+1})\right) \geq 0{,}9} \leftarrow \textit{Coverage}$$

A. N. Angelopoulos and S. Bates, "Conformal Prediction: A Gentle Introduction", *Foundations and Trends® in ML*, 2023

Introducción

- Enfoque: usar modelos de ML como caja negra y construir "intervalos de confianza" para sus predicciones
- A partir de un conjunto de entrenamiento $(X_i, Y_i) \sim P, i = 1, \ldots, n$ iid, con P distribución en $\mathcal{X} \times \mathcal{Y}$, buscamos construir una función

$$\mathcal{C}_n : \mathcal{X} \to \mathcal{P}(\mathcal{Y}) = \{\text{subconjuntos de } \mathcal{Y}\}$$

tal que para un nuevo par iid $(X_{n+1}, Y_{n+1}) \sim P$ se cumpla

$$\mathbb{P}\left(Y_{n+1} \in \mathcal{C}_n(X_{n+1})\right) \geq 1 - \alpha \text{ con } \alpha > 0 \text{ fijado de antemano}$$

A. N. Angelopoulos and S. Bates, "Conformal Prediction: A Gentle Introduction", *Foundations and Trends® in ML*, 2023

Introducción

- **P**: ¿Puedo construir la función $\mathcal{C}_n$ sin hipótesis sobre P y para n finito?
- **R**: Increíblemente, si. Defino:

$$\mathcal{C}_n(X_{n+1}) = \begin{cases} \mathcal{Y} & \text{con prob. } 1-\alpha \\ \emptyset & \text{con prob. } \alpha \end{cases}$$

- **P**: ¿Puedo hacer algo no trivial?

Introducción

- Asumamos (por ahora) que no hay features X_i y que $Y_i \in \mathbb{R}$
- Supongamos que quiero encontrar un $\mathcal{C}_n = (-\infty, \hat{q}_n]$, es decir, que

$$\mathbb{P}\left(Y_{n+1} \leq \hat{q}_n\right) \geq 1 - \alpha$$

- Intuición: tomar

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

- Problema: obtenemos $\mathbb{P}\left(Y_{n+1} \leq \hat{q}_n\right) \approx 1 - \alpha$, igualdad solo cuando $n \to \infty$
- Solución: ajustar cuantil usando estadísticos de orden
  - $Y_{(i)}$ es la i-ésima va más chica de $Y_1, \ldots, Y_{n+1}$
  - $Y_{(1)} \leq Y_{(2)} \leq \cdots \leq Y_{(n+1)}$, por ej. $Y_8 \leq Y_3 \leq \cdots \leq Y_1$
  - Y_i tiene rango R_i si $Y_i = Y_{(R_i)}$, por ej. Y_8 tiene rango 1, Y_3 rango 2, etc...

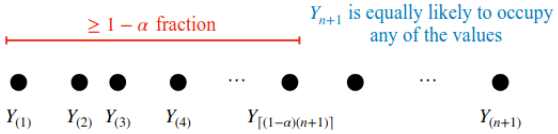
Prop: si $Y_1, Y_2, \ldots, Y_{n+1}$ son iid, entonces el rango de Y_{n+1} se distribuye uniformemente entre los valores $1, 2, \ldots, n+1$

Introducción

Prop: si $Y_1, Y_2, \ldots, Y_{n+1}$ son iid, entonces el rango de Y_{n+1} se distribuye uniformemente entre los valores $1, 2, \ldots, n+1$

Entonces:

$$\mathbb{P}\left(Y_{n+1} \text{ esté entre los } \lceil (1-\alpha)(n+1) \rceil \text{ valores más chicos de } Y_1, \ldots, Y_{n+1}\right) \geq 1 - \alpha$$



Lo que implica:

$$\mathbb{P}\left(Y_{n+1} \text{ esté entre los } \lceil (1-\alpha)(n+1) \rceil \text{ valores más chicos de } Y_1, \ldots, Y_n\right) \geq 1 - \alpha$$

R. J. Tibshirani, Notes on Conformal Prediction for *Advanced Topics in Statistical Learning*, Berkeley, spring 2023

Introducción

- Si queremos

$$\mathbb{P}\left(Y_{n+1} \leq \hat{q}_n\right) \geq 1-\alpha$$

basta tomar

$$\hat{q}_n = \lceil (1-\alpha)(n+1) \rceil \text{ valor más chico de } Y_1, \dots Y_n$$

- Equivalentemente

$$\hat{q}_n = \text{Quantile}\left(\frac{\lceil (1-\alpha)(n+1) \rceil}{n}; \frac{1}{n}\sum_{i=1}^{n} \delta_{Y_i}\right) \hat{q}_n = \text{Quantile}\left(\underbrace{\frac{\lceil (1-\alpha)(n+1) \rceil}{n}}_{\substack{\text{corrección por muestra} \\ \text{finita}}}; \frac{1}{n}\sum_{i=1}^{n} \delta_{Y_i}\right)$$

- **Obs**: Para que valga la propiedad sobre uniformidad y rangos no es necesario pedir iid, basta con que $Y_1, \dots, Y_{n+1}$ sean intercambiables

Def: $Y_1, \dots, Y_{n+1}$ son intercambiables si su distribución conjunta no cambia bajo permutaciones, i.e.

$$(Y_1, \dots, Y_{n+1}) \stackrel{d}{=} \left(Y_{\sigma(1)}, \dots, Y_{\sigma(n+1)}\right) \text{ para toda permutación } \sigma$$

Coverage

- Si las Y_i's son intercambiables, tenemos que

$$\mathbb{P}\left(Y_{n+1} \text{ esté entre los } \lceil (1-\alpha)(n+1) \rceil \text{ valores más chicos de } Y_1, \ldots, Y_n\right) \geq 1-\alpha$$

- Si además casi seguramente no hay empates entre las Y_i's, tenemos la igualdad:

$$\begin{aligned}\mathbb{P}\left(Y_{n+1} \text{ esté entre los } \lceil (1-\alpha)(n+1) \rceil \text{ valores más chicos de } Y_1, \ldots, Y_n\right) &= \sum_{i=1}^{\lceil (1-\alpha)(n+1) \rceil} \frac{1}{n+1} \\ &= \frac{\lceil (1-\alpha)(n+1) \rceil}{n+1} \\ &< \frac{(1-\alpha)(n+1)+1}{n+1}\end{aligned}$$

$$\Rightarrow \mathbb{P}\left(Y_{n+1} \text{ esté entre los } \lceil (1-\alpha)(n+1) \rceil \text{ valores más chicos de } Y_1, \ldots, Y_n\right) < (1-\alpha) + \frac{1}{n+1}$$

- Entonces:

$$\mathbb{P}\left(Y_{n+1} \leq \hat{q}_n\right) \in \left[1-\alpha, (1-\alpha) + \frac{1}{n+1}\right)$$

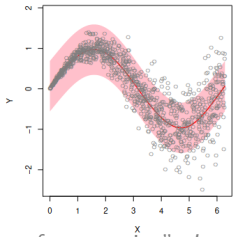
Regresión

- Ahora observo $(X_i, Y_i) \in \mathcal{X} \times \mathbb{R}, i = 1, \ldots, n$ iid y para (X_{n+1}, Y_{n+1}) quiero construir $\mathcal{C}_n$ tal que:

$$\mathbb{P}\left(Y_{n+1} \in \mathcal{C}_n(X_{n+1})\right) \geq 1 - \alpha$$

- Supongamos que tengo entrenado un predictor puntual de Y, i.e. dado x, $\hat{f}(x)$ predice el valor de y que esperamos ver (e.g lasso para regresión lineal)
- Vamos a "conformalizar" la predicción dada por $\hat{f}$, i.e. dar un $\mathcal{C}_n$ de la forma

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



J. Lei et al, "Distribution-free predictive inference for regression", *Journal of the American Statistical Association*, 2018.

Regresión

- *Split-conformal prediction*: divido mis n observaciones en dos conjuntos disjuntos D_T, D_C
  - D_T es el *proper training set*, $|D_T| = n_T$
  - D_C es el *calibration set*, $|D_C| = n_c$
- Entreno el predictor puntual con $(X_i, Y_i) \in D_T \Rightarrow$ obtengo $\hat{f}_{n_T}$
- Uso $(X_i, Y_i) \in D_C$ para encontrar $\hat{q}_{n_c}$
  - Para cada $(X_i, Y_i) \in D_C$ defino el residuo $R_i = |Y_i - \hat{f}_{n_T}(X_i)|$
  - Hallo cuantil $\hat{q}_{n_c} = \lceil (1-\alpha)(n_c+1) \rceil$ valor más chico de $R_i, i \in D_C$
- *Conformal set* $\mathcal{C}_n(x) = \left[\hat{f}_{n_T}(x) - \hat{q}_{n_c}, \hat{f}_{n_T}(x) + \hat{q}_{n_c}\right]$
- Razonando igual que antes:

$$\mathbb{P}\left(Y_{n+1} \in \mathcal{C}_n(X_{n+1}) \middle| (X_i, Y_i), i \in D_T\right) \in \left[1-\alpha, (1-\alpha) + \frac{1}{n_c+1}\right)$$

- ¿Por qué divido al conjunto de entrenamiento?

Regresión

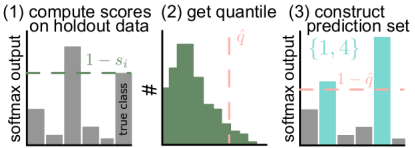
- Supongamos que hago el procedimiento anterior pero no divido las n observaciones iniciales
  - Entreno el predictor $\hat{f}_n$ con $(X_i, Y_i), i = 1, \ldots, n$
  - Para cada $i = 1, \ldots, n$ defino el residuo $R_i = |Y_i - \hat{f}_{n_T}(X_i)|$
  - Hallo cuantil $\hat{q}_n = \lceil (1-\alpha)(n_c+1) \rceil$ valor más chico de R_i
  - Hallo conformal set $\mathcal{C}_n(x) = \left[\hat{f}_n(x) - \hat{q}_n, \hat{f}_n(x) + \hat{q}_n\right]$
- Entonces

$$\begin{aligned} Y_{n+1} \in \mathcal{C}_n(X_{n+1}) &\Leftrightarrow R_{n+1} \leq \hat{q}_n \\ &\Leftrightarrow R_{n+1} \leq \lceil (1-\alpha)(n+1) \rceil \text{ valores más chicos de } R_1, \ldots, R_n \end{aligned}$$

- Problema: $\hat{f}_n$ fue entrenado con $(X_i, Y_i), i = 1, \ldots, n$, pero sin (X_{n+1}, Y_{n+1})
- $R_{n+1} = |Y_{n+1} - \hat{f}_n(X_{n+1})|$ **no** es independiente de (tampoco intercambiable con) $R_1, \ldots, R_n$
- Si dividimos y condicionamos al conjunto de entrenamiento, los residuos de calibración $R_i, i \in D_C$, y de test R_{n+1} **sí son** iid

Conformal prediction: clasificación

- Supongamos $(X_i, Y_i) \in \mathcal{X} \times \{1, \ldots, K\}, i = 1, \ldots, n$
- Divido mis n observaciones en dos conjuntos disjuntos D_T, D_C
  - Uso D_T para entrenar un clasificador probabilístico, i.e., $\hat{f}_{n_T}(x, k)$ estima $\mathbb{P}(Y = k|X = x)$ para $k = 1, \ldots, K$(e.g NN con softmax scores)
  - Defino $R_i = 1 - \hat{f}_{n_T}(X_i, Y_i), i \in D_C$ (e.g. $R_i = 1 - s(X_i, Y_i)$)
  - Hallo cuantil $\hat{q}_n = \lceil (1 - \alpha)(n_c + 1) \rceil$ valor más chico de R_i
  - Hallo conformal set $\mathcal{C}_n(X_{n+1}) = \left\{ y \in \{1, \ldots, K\} : 1 - \hat{f}_{n_T}(X_{n+1}, y) \leq \hat{q}_n \right\}$



A. N. Angelopoulos and S. Bates, "Conformal Prediction: A Gentle Introduction", *Foundations and Trends® in Machine Learning*, 2023

Generalización: score functions

- Para garantizar el *coverage* alcanza con una función de puntuación que sea "negatively-oriented" i.e. valores bajos $\leftrightarrow$ buen ajuste del modelo
- Para regresión habíamos tomado $R_i = |y - \hat{f}_{n_T}(x)|$, pero podríamos haber elegido cualquier $V(x, y) = V((x, y), \hat{f}_{n_T})$ (que sea negatively-oriented)
- En ese caso, definiendo $R_i = V(x, y), i \in D_C$ y

$$\mathcal{C}_n(x) = \{y : V(x, y) \leq \lceil (1 - \alpha)(n + 1) \rceil \text{ valores más chicos de } R_i, i \in D_C\}$$

tenemos la misma garantía de coverage que antes:

Teorema: si $(X_1, Y_1), (X_2, Y_2), \ldots, (X_{n+1}, Y_{n+1})$ son iid, entonces $\mathbb{P}(Y_{n+1} \in \mathcal{C}_n(X_{n+1})) \geq 1 - \alpha$

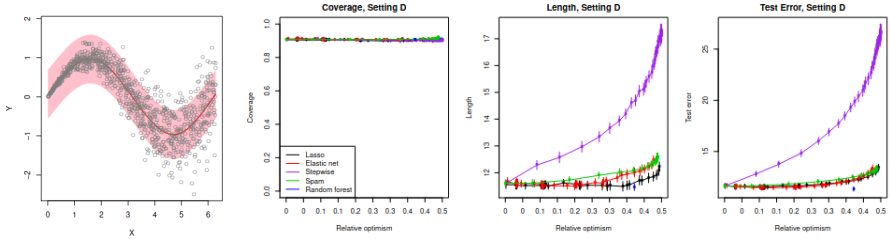
V. Vovk, A. Gammerman y C. Saunders, "Machine-learning applications of algorithmic randomness", *ICML*, 1999

Teorema: Si además la distribución conjunta de $\{R_i : i \in D_C\}$ es continua, entonces $\mathbb{P}(Y_{n+1} \in \mathcal{C}_n(X_{n+1})) < 1 - \alpha + \frac{1}{n_c+1}$

J. Lei et al, "Distribution-free predictive inference for regression", JASA, 2018

Impacto del estimador $\hat{f}$

- Los teoremas anteriores nos garantizan que, sin importar cómo se elija el estimador puntual $\hat{f}$, siempre vamos a tener coverage
- Sin embargo, cuanto mejor sea $\hat{f}$, menor será el tamaño del conjunto $\mathcal{C}_n$



- Intuición:

$$\text{Tamaño promedio} = \mathbb{E}_{(X_i,Y_i)\sim P, i\in D_C}\left[\int\int_{\mathcal{C}_n(x)} d\mu(y)dP_X(x)\right]$$

$$\text{Coverage} = \mathbb{E}_{(X_i,Y_i)\sim P, i\in D_C}\left[\int\int_{\mathcal{C}_n(x)} dP_{Y|X}(y)dP_X(x)\right]$$

Back to coverage

- Recordemos que para split-CP teníamos el coverage condicional

$$\mathbb{P}\left(Y_{n+1} \in \mathcal{C}_n(X_{n+1}) \middle| (X_i, Y_i), i \in D_T\right) \in \left[1-\alpha, (1-\alpha) + \frac{1}{n_c+1}\right)$$

- Marginalizando sobre el conjunto de entrenamiento tenemos

$$\mathbb{P}\left(Y_{n+1} \in \mathcal{C}_n(X_{n+1})\right) \in \left[1-\alpha, (1-\alpha) + \frac{1}{n_c+1}\right)$$

- ¿Y si condicionamos a todos nuestros datos?

$$\mathbb{P}\left(Y_{n+1} \in \mathcal{C}_n(X_{n+1}) \middle| (X_i, Y_i), i = 1, \ldots, n\right) \sim \text{Beta}(k_\alpha, n_c + 1 - k_\alpha)$$

con $k_\alpha = \lceil (1-\alpha)(n_c+1) \rceil$
Esa distribución tiene media $\frac{k_\alpha}{n_c+1} \approx 1-\alpha$ y varianza

$$\frac{k_\alpha(n_c+1-k_\alpha)}{(n_c+1)^2(n_c+2)} \approx \frac{\alpha(1-\alpha)}{n_c+2}$$

Back to coverage

- Recordemos que para split-CP teníamos el coverage condicional

$$\mathbb{P}\left(Y_{n+1} \in \mathcal{C}_n(X_{n+1}) \middle| (X_i, Y_i), i \in D_T\right) \in \left[1-\alpha, (1-\alpha) + \frac{1}{n_c+1}\right)$$

- Marginalizando sobre el conjunto de entrenamiento tenemos

$$\mathbb{P}\left(Y_{n+1} \in \mathcal{C}_n(X_{n+1})\right) \in \left[1-\alpha, (1-\alpha) + \frac{1}{n_c+1}\right)$$

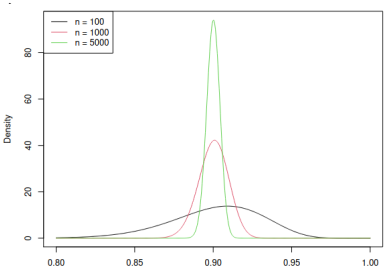
- ¿Y si condicionamos a todos nuestros datos?

$$\mathbb{P}\left(Y_{n+1} \in \mathcal{C}_n(X_{n+1}) \middle| (X_i, Y_i), i = 1, \ldots,\right.$$

con $k_\alpha = \lceil (1-\alpha)(n_c+1) \rceil$

Esa distribución tiene media $\frac{k_\alpha}{n_c+1} \approx 1-\alpha$ y varianza

$$\frac{k_\alpha(n_c+1-k_\alpha)}{(n_c+1)^2(n_c+2)} \approx \frac{\alpha(1-\alpha)}{n_c+2}$$



Demostración de lo anterior

- Quiero ver que

$$\mathbb{P}\left(Y_{n+1} \in \mathcal{C}_n(X_{n+1}) \middle| (X_i, Y_i), i = 1, \ldots, n\right) \sim \text{Beta}(k_\alpha, n_c + 1 - k_\alpha)$$

- Sean $U_1, \ldots, U_n$ iid. $U(0,1)$ y $U_{(1)} \leq U_{(2)} \leq \cdots \leq U_{(n)}$. Entonces $U_{(k)}$ tiene densidad

$$f_{(k)} = n\binom{n-1}{k-1} x^{k-1}(1-x)^{n-k}$$

  i.e. $U_{(k)} \sim \text{Beta}(k, n+1-k)$
- Sean ahora $R_1, \ldots, R_{n+1}$ iid con distribución F continua. Entonces

$$\mathbb{P}\left(R_{n+1} \leq R_{(k)} \middle| R_1, \ldots, R_n\right) \sim \text{Beta}(k, n_c + 1 - k)$$

  porque $U_i = F(R_i)$

X-conditional coverage

- ¿Qué pasa si condicionamos a X_{n+1}?

$$\mathbb{P}\left(Y_{n+1} \in \mathcal{C}_n(x) \middle| (X_i, Y_i), i \in D_T, X_{n+1} = x\right) \overset{?}{\geq} 1 - \alpha, \text{ para todo } x \in \mathcal{X}$$

- Lamentablemente, esto es pedir demasiado

Prop: si $\mathcal{C}_n : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ es una función cualquiera que cumple

$$\mathbb{P}\left(Y_{n+1} \in \mathcal{C}_n(x) \middle| X_{n+1} = x\right) \geq 1 - \alpha, \text{ para toda } P \text{ y para } P_X\text{-casi todo } x,$$

entonces:

$$\mathbb{E}\left[\mu(\mathcal{C}_n(n))\right] = \infty \text{ para toda } P \text{ y para } P_X\text{-casi todo } x \text{ que no sea un átomo}$$

R. F. Barber, E. J. Candès, A. Ramdas y R. J. Tibshirani, "The limits of distribution-free conditional predictive inference", *Information and Inference*, 2021

Full Conformal Prediction

- ¿Es necesario dividir mis datos en entrenamiento y calibración?
- No, y eso se llama Full Conformal Inference
- De nuevo, observamos $(X_i, Y_i), i = 1, \ldots, N$ y X_{n+1}, y queremos dar un conjunto de predicción que contenga a Y_{n+1}
- Idea: testear todos los posibles $y \in \mathcal{Y}$ (caro amigo)
- Para cada $y \in \mathcal{Y}$, entrenamos un modelo $\hat{f}_n^y$ con

$$(X_1, Y_1), \ldots, (X_n, y_n), (X_{n+1}, y)$$

(el entrenamiento tiene que ser invariante a permutaciones)

- Definimos los residuos

$$R_i^y = |Y_i - \hat{f}_n^y(X_i)|, i = 1, \ldots, n \text{ y } R_{n+1}^y = |y - \hat{f}_n^y(X_{n+1})|$$

y el umbral

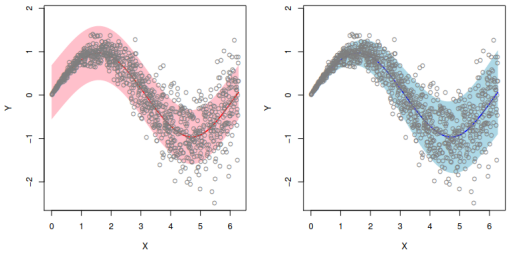
$$\hat{q}_n^y = \lceil (1-\alpha)(n+1) \rceil \text{ valor más chico de } R_1, \ldots, R_n$$

y el conjunto de predicciones es

$$\mathcal{C}(X_{n+1}) = \{y : R_{n+1}^y \leq \hat{q}_n^y\}$$

Adaptividad local

- Vimos que mejores estimadores puntuales llevan a conjuntos $\mathcal{C}_n(x)$ más chicos
- Dado un estimador puntual, ¿es posible achicar/agrandar el tamaño de $\mathcal{C}_n(x)$ si la predicción para x es fácil/difícil?



- **Studentized residuals**: hacemos split-CP, pero con D_T entrenamos un predictor puntual $\hat{f}_{n_T}$ y un predictor del spread $\hat{\sigma}_{n_T}$
- En D_C, normalizamos los residuos $R_i = \dfrac{|Y_i - \hat{f}_{n_T}(X_i)|}{\hat{\sigma}_{n_T}(X_i)}, i \in D_C$, elegimos el umbral $\hat{q}$ con los $\lceil (1-\alpha)(n+1) \rceil$ valores más chicos y devolvemos

$$\mathcal{C}_n(x) = \left[\hat{f}(x) - \hat{\sigma}_{n_T}(x)\hat{q}_n, \hat{f}(x) + \hat{\sigma}_{n_T}(x)\hat{q}_n \right]$$

Adaptividad local: regresión a cuantiles

- El enfoque anterior requiere primero entrenar $\hat{f}_{n_T}$ y luego $\hat{\sigma}_{n_T}$
- Además, si por ej. $\hat{f}_{n_T}$ se estima usando una NN, los residuos sobre D_T son $\approx 0 \rightarrow$ no son una buena aproximación de los residuos sobre D_C
- Recordar, distribución condicional y α-cuantil condicional:

$$F(y|X=x) = \mathbb{P}(Y \leq y | X = x), \quad c^{\alpha}(x) = \text{ínf}\{y \in \mathbb{R} : F(y|X=x) \geq \alpha\}$$

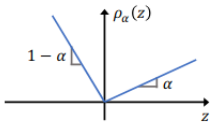
- Regresión usual: estimar $\mathbb{E}(Y_{n+1}|X_{n+1}=x)$ mediante

$$\hat{f}_{n_T}(x) = f(x;\hat{\theta}), \text{ con } \hat{\theta} = \underset{\theta}{\text{argmin}} \frac{1}{n_T} \sum_{i=1}^{n_T} [Y_i - f(X_i;\theta)]^2 + \mathcal{R}(\theta)$$

- **Quantile regression**: estimar directamente los cuantiles:

$$\hat{c}^{\alpha}_{n_T}(x) = f(x;\hat{\theta}), \text{ con } \hat{\theta} = \underset{\theta}{\text{argmin}} \frac{1}{n_T} \sum_{i=1}^{n_T} \rho_{\alpha}(Y_i, f(X_i;\theta)) + \mathcal{R}(\theta)$$

ρ_{α} es *pinball loss*



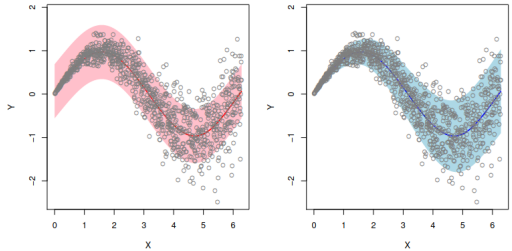
Adaptividad local: conformalized quantile regression

- **Conformalized quantile regression**: "conformalizar" los cuantiles
  - Estimar $\hat{c}_{n_T}^{\alpha/2}$ y $\hat{c}_{n_T}^{1-\alpha/2}$ usando quantile regression con el conjunto D_T
  - Tomar scores

$$R_i = \text{máx}\left\{\hat{c}_{n_T}^{\alpha/2}(X_i) - Y_i, Y_i - \hat{c}_{n_T}^{1-\alpha/2}(X_i)\right\} \; i \in D_C$$

  - Definir $\hat{q}_n = \lceil (1-\alpha)(n_c+1) \rceil$ valor más chico de $R_i, i \in D_C$
  - Devolver

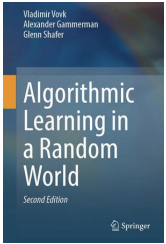
$$\mathcal{C}_n(x) = \left[\hat{c}_{n_T}^{\alpha/2}\hat{q}_n, \hat{c}_{n_T}^{1-\alpha/2} + \hat{q}_n\right]$$



Y. Romano, E. Patterson y E. Candès, "Conformalized Quantile Regression", NeurIPS, 2019.

Recursos adicionales

V. Vovk, A. Gammerman y G. Shafer, *Algorithmic Learning in a Random World*, Springer, 2005 (1º ed.) y 2022 (2º ed.)



V. Manokhin, *Awesome Conformal Prediction* en Github

README License

Awesome Conformal Prediction

Applied Conformal Prediction

NEW · 3 WEEKS · COHORT-BASED COURSE

Learn about, deep dive into and get hands-on experience with Conformal Prediction techniques for quantifying uncertainty in machine learning

course 'Applied Conformal Prediction is now opened for enrollment on Maven enroll into next cohort here, register interest for upcoming cohorts and information here

Events

1. Applied Conformal Prediction course starts in May 2024!
2. Kaggle competition - probabilistic forecasting I: Temperature

Books

1. Practical Guide to Applied Conformal Prediction: Learn and apply the best uncertainty frameworks to your industry applications by Valeriy Manokhin (2024)
2. Algorithmic Learning in a Random World by Vladimir Vovk and Alex Gammerman, also Glenn Shafer (2022). Second edition.
3. Conformal Prediction for Reliable Machine Learning by Vineeth Balasubramanian, Shen-Shyang Ho, Vladimir Vovk (2014) OLD BOOK, largely out of date, no code

Tutorials

1. A Conformal Prediction tutorial, an introductive review of the basics by Margaux Zaffran (2024)
2. Conformal Prediction Tutorial by Henrik Linusson (2021)
3. Predicting with Confidence - Henrik Boström (2016)
4. Henrik Linusson: Conformal Prediction by Henrik Linusson (2020)
5. Conformal Prediction: a Unified Review of Theory and New Challenges by Gianluca Zeni, Matteo Fontana1 and Simone Vantini (Politecnico di Milano, Italy, 2021)
6. Conformal Prediction: How to quantify uncertainty of machine learning models? ECAS-ENBIS course-ENBIS 2023 Annual conference by Margaux Zaffran (2023)
7. A Tutorial on Conformal Prediction by Glenn Shafer and Vladimir Vovk (2008)
8. Tutorial on Venn-ABERS prediction by Paolo Toccaceli (Royal Holloway, UK, 2019)

Papers

1. Introducing Conformal Prediction in Predictive Modeling. A Transparent and Flexible Alternative to Applicability Domain Determination by Ulf Norinder, Lars Carlsson, Scott Boyer, and Martin Eklund (2014)
2. Uncertainty Sets for Image Classifiers using Conformal Prediction by Anastasios N. Angelopoulos, Stephen Bates, Jitendra Malik, & Michael I. Jordan (Berkeley, 2021)
3. Conformal Prediction Under Covariate Shift by Ryan Tibshirani, Rina Foygel Barber, Emmanuel Candes, Aaditya Ramdas (Carnegie Mellon, Stanford, Chicago, 2019)