STAT 403 5/2/25

Lecture 13

Conformal prediction

Lecture Notes VI – Modern resampling methods. Conformal prediction

Marina Meila
mmp@stat.washington.edu

Department of Statistics
University of Washington

May 2025

Re-Sampling

Bootstrap ✓

Jackknife ✓ - LOO → distibution of $y_i - \hat{y}_i$

Leave-One-Out (∝ LOO CV) → averages "errors"

Validation Likelihood (KDE)

...

⇒ score

Bag of little bootstraps ✓

Conformal prediction. Jackknife+ ←

Conformal prediction

NEW X

AFTER TRAINING

- **Conformal prediction:** CI for a single prediction $\hat{y} = f(x)$

Given data set $\mathcal{D} = \{(x_i, y_i), i = 1 : N\}$
Training algorithm $\mathcal{A}$, so that $\mathcal{A}(\mathcal{D}) = f$ the predictor

Want CI for $\hat{y} = f(x)$ where x is a new data point
so that the CI is NOT dependent on $\mathcal{A}$ being statistically correct (e.g. $\mathcal{A}$ overfits, ...)

- `jackknife+` is a simple algorithm for CP
- More advanced algorithms exist. This is an active area of research in statistics.

!!!! Do NOT use CP to "crossvalidate" your algorithm!

Conformal prediction

- **Conformal prediction:** CI for a single prediction $\hat{y} = f(x)$

$1-\alpha$ = confidence level

Given data set $\mathcal{D} = \{(x_i, y_i), i = 1 : N\}$
Training algorithm $\mathcal{A}$, so that $\mathcal{A}(\mathcal{D}) = f$ the predictor
Want CI for $\hat{y} = f(x)$ where x is a new data point
so that the CI is NOT dependent on $\mathcal{A}$ being statistically correct (e.g. $\mathcal{A}$ overfits, ...)

- `jackknife+` is a simple algorithm for CP
- More advanced algorithms exist. This is an active area of research in statistics.

$$\mathcal{D} \longrightarrow \boxed{A} \longrightarrow f$$

$$\mathcal{D}_{-i} \longrightarrow \boxed{A} \longrightarrow f_{-i} \qquad i=1:n$$

!!!! Do NOT use CP to "crossvalidate" your algorithm!

$\mathcal{D}_{-i} = \mathcal{D} \setminus \{(x_i, y_i)\}$ Leave Out data point i

jackknife+

jackknife+ Algorithm

In data set $\mathcal{D} = \{(x_i, y_i), i = 1:n\}$
Training algorithm $\mathcal{A}$, so that $\mathcal{A}(\mathcal{D}) = f$ the predictor
Confidence level $1-\alpha$

Want CI for $\hat{y} = f(x)$ where x is a new data point

1. Precompute $f_{-i} \leftarrow \mathcal{A}(\mathcal{D}_{-i})$ for $i = 1:n$
2. Compute "leave one out" residuals $R_i = |y_i - f_{-i}(x_i)|$, for $i = 1:n$
3. For every new x: compute $f(x)$, then get $1-\alpha$ Prediction Interval $[a, b]$ for $f(x)$ by
   - 3.1 Compute lower bounds $a_i = f_{-i}(x) - R_i$, for $i = 1:n$
   - 3.2 Sort $a_{1:n}$
   - 3.3 Set $a \leftarrow \lfloor \frac{\alpha}{2} n \rfloor$ quantile of $a_{1:n}$
   - 3.4 Compute upper bounds $b_i = f_{-i}(x) + R_i$, for $i = 1:n$
   - 3.5 Sort $b_{1:n}$
   - 3.6 Set $b \leftarrow \lceil \left(1 - \frac{\alpha}{2}\right) n \rceil$ quantile of $b_{1:n}$
   - 3.7 Output $CI^{\alpha} = \underline{[a, b]}$
4. **Theorem** $P[\underline{y(x) \in CI^{\alpha}}] \geq 1-\alpha$

jackknife+

training, CV, ...

TRAIN + PREP

$$\mathcal{D} \longrightarrow \boxed{A} \longrightarrow f$$

$$\mathcal{D}_{-i} \quad i=1:n \quad \boxed{f_{-i}} \longrightarrow f_{-i}(x_i) = \hat{y}_{-i} \qquad \underline{R_i} = |\hat{y}_{-i} - y_i|$$

x_i

PREDICTING

using $f,\ f_{-1,\dots,-n}$

new x:

$$\mathcal{D} \longrightarrow \boxed{A} \longrightarrow f \xrightarrow{x} \left(f(x) = \hat{y}\right) \leftarrow \text{prediction}$$

$$\mathcal{D}_{-i} \quad i=1:n \quad f_{-i} \longrightarrow f_{-i}(x) \qquad \{f_{-i}(x) + R_i = b_i,\ i=1:n\}$$

x

$f(x) = \hat{y}$

y

$\lceil n \tfrac{1-\alpha}{2} \rceil$ quantile of $\{b_i, i=1:n\}$

b

$\alpha = 20\%$

$1 - \frac{\alpha}{2} = 90\%$

$\left\lceil n \frac{1-\alpha}{2} \right\rceil$

$CI_\alpha = [a, b]$

jackknife+ $\quad CI_\alpha = [a, b]$

~~PREDICTING~~ using f

new x:

$$\mathcal{D} \rightarrow \boxed{A} \rightarrow f \xleftarrow{} x, \quad f \rightarrow f(x) = \hat{y} \leftarrow \text{prediction}$$

$$\mathcal{D}_{-i} \rightarrow \boxed{A} \rightarrow f_{-i} \rightarrow f_{-i}(x), \quad i = 1:n, \quad x \rightarrow f_{-i}$$

"ceiling" $\lceil z \rceil = \min \{m \in \mathbb{Z}, m \geq z\}$

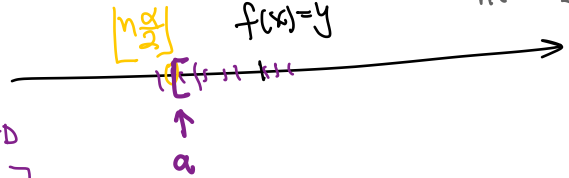
floor $\lfloor z \rfloor = \max \{m \in \mathbb{Z}, m \leq z\}$

$\alpha = 20\%$

$\frac{\alpha}{2} = 10\%$

$f_{-i}(x) - R_i = a_i$

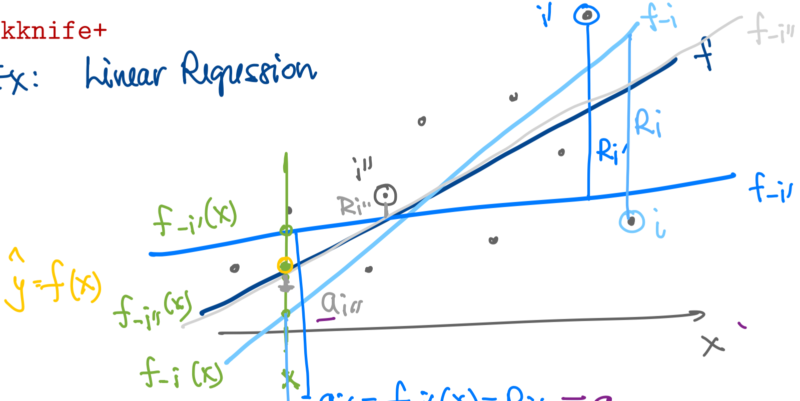
$a = \lfloor n \frac{\alpha}{2} \rfloor$ quantile $\rightarrow \{a_i, i = 1:n\}$



Rem: $f(x)$ NOT USED to obtain $[a, b]$

jackknife+

Ex: Linear Regression



$a_{i'} = f_{-i'}(x) - R_{i'} = a$

$a_i = f_{-i}(x) - R_i$

$\leftarrow$ pick 10% quantile

$\{a_{1:n}\} \Rightarrow a$

$\alpha = 20\%$

$n\frac{\alpha}{2} = 1$

jackknife+

Ex: Linear Regression

$b = \lceil n \frac{1-\alpha}{2} \rceil$ quantile

of $\{ b_i, i = 1:n \}$