ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ

BÁO CÁO CUỐI KÌ
BỘ MÔN: CÁC VẤN ĐỀ HIỆN ĐẠI CÔNG NGHỆ THÔNG TIN

Chủ đề: Conformal Prediction

Giảng viên hướng dẫn: **PGS. TS. Nguyễn Hải Châu**

Nhóm 4: **Hoàng Văn Nguyên** (**21020370**)
Huỳnh Tiến Dũng (**21020007**)
Vũ Quốc Tuấn (**21020033**)

Lớp: **2324I_INT3507_3**

HÀ NỘI, 01/2024

Lời nói đầu

Trong thời đại ngày nay, khi sự phức tạp và không chắc chắn của dữ liệu ngày càng tăng, việc đưa ra dự đoán chính xác và đồng thời xác định mức độ tin cậy của chúng là một thách thức đối với nhiều lĩnh vực, từ khoa học dữ liệu đến ứng dụng thực tế. *Conformal prediction*, hay dự đoán hợp lý, nổi lên như một phương pháp mạnh mẽ để đối mặt với vấn đề này, cung cấp một cơ hội độc đáo để ước lượng mức độ không chắc chắn của mô hình dự đoán.

Trong bài báo cáo này, nhóm em xin giới thiệu về *conformal prediction*, tại sao cần áp dụng conformal prediction cho các mô hình machine learning, trình bày các giả thiết cũng như nhiều phương pháp khác nhau, và không thể thiếu phần thực nghiệm của nhóm.

Bằng cách hiểu rõ hơn về conformal prediction, chúng ta có thể mở ra cánh cửa cho việc nâng cao chất lượng của dự đoán và đồng thời cung cấp thông tin quan trọng về mức độ tin cậy, giúp hỗ trợ đưa ra quyết định trong nhiều tình huống quan trọng. Kính mời độc giả cùng nhóm em khám phá sâu hơn về con đường hứa hẹn của conformal prediction trong thế giới đầy thách thức của dữ liệu và mô hình ngày nay.

Trong quá trình phân tích, nhóm 4 chúng em đã cố gắng tìm hiểu về các tài liệu có liên quan đến chủ đề này. Tuy nhiên, bài báo cáo có thể sẽ vẫn còn tồn tại nhiều hạn chế và sai sót; vì vậy, nhóm chúng em hy vọng sẽ nhận được những góp ý và chia sẻ từ thầy cũng như các bạn đọc cho việc cải thiện những thiếu sót đó để cả nhóm có thể cải thiện và làm tốt hơn trong những lần sau.

Cuối cùng, chúng em xin gửi lời cảm ơn tới thầy **PGS.TS. Nguyễn Hải Châu** đã định hướng, hướng dẫn và giúp đỡ nhóm 4 hoàn thiện tốt bài báo cáo này ạ.

Chúng em xin chân thành cảm ơn!

Mục lục

Chương 1

Conformal là gì, tại sao cần áp dụng conformal prediction cho các mô hình machine learning

1.1 Conformal là gì

Conformal prediction (dự đoán hợp lý, hay còn gọi là conformal inference) là một phương pháp dùng để ước lượng tính không chắc chắn của các dự đoán được đưa ra bởi các mô hình machine learning hay thuật toán dự đoán bất kỳ. Về cơ bản, nó thực hiện điều này bằng cách chuyển đổi các dự đoán của thuật toán thành các tập dự đoán (prediction sets), có tính chất bao phủ (coverage) tốt với số mẫu hữu hạn. Hay nói cách khác Conformal prediction chuyển đổi các *classifier* và các *regressor* thành các *confidence predictor*.

Các confidence predictor cung cấp vùng dự đoán đa giá trị thay vì chỉ một giá trị; đối với bài toán phân lớp, bộ dự đoán đưa ra tập các nhãn, còn với bài toán hồi quy, thì là khoảng giá trị. Các dự đoán gắn liền với thước đo độ tin cậy: một vùng dự đoán với độ tin cậy γ chứa đầu ra đúng với xác suất γ, $\gamma \in [0, 1]$.

Khi nói đến "conformal" trong conformal prediction, thường ám chỉ việc thiết kế các phương pháp mô hình sao cho chúng tuân thủ các nguyên tắc conformal, tức là có khả năng cung cấp các ước lượng tin cậy đúng đắn, đưa ra tập các nhãn hoặc một khoảng với một độ tin cậy cho trước. Các phương pháp này có thể được áp dụng trong nhiều bối cảnh khác nhau, từ dự đoán hồi quy đến phân lớp và các tác vụ machine learning khác.

Mục tiêu cơ bản của conformal prediction như sau: Giả sử $(X_i, Y_i) \sim P$, $i = 1, \ldots, n$ là các cặp đặc trưng và dự đoán độc lập và phân phối giống nhau (i.i.d.), được lấy từ phân phối P trên $\mathcal{X} \times \mathcal{Y}$. Cho $\alpha \in (0, 1)$ là một mức lỗi giả định (nominal error level), và muốn tìm ra một *dải dự đoán* (prediction band):

$$\hat{C}_n : \mathcal{X} \to \{\text{tập con (hoặc khoảng) của } \mathcal{Y}\} \tag{1.1}$$

với tính chất là đối với mỗi cặp i.i.d. mới $(X_{n+1}, Y_{n+1}) \sim P$,

$$\mathbb{P}\big(Y_{n+1} \in \hat{C}_n(X_{n+1})\big) \geq 1 - \alpha \tag{1.2}$$

1.2 Tại sao cần áp dụng conformal prediction cho mô hình machine learning

Việc ước lượng tính không chắc chắn là cần thiết trong nhiều trường hợp:

- Khi muốn ước lượng tính không chắc chắn của cái gì trước khi quyết định làm gì.
- Khi muốn thiết kế các hệ thống có thể xử lý các tình huống bất ngờ.
- Khi đã tự động hóa một tác vụ bằng machine learning và cần báo thời điểm cần can thiệp.
- Muốn truyền đạt sự không chắc chắn về dự đoán của mô hình tới các bên liên quan.
- Xác định tính không chắc chắn để nhắc nhở con người can thiệp.

Các mô hình machine learning thông thường chỉ đưa ra một dự đoán và hoàn toàn tin tưởng vào chúng, chúng ta cần biết những dự đoán đó thực sự chắc chắn đến mức nào. Áp dụng conformal prediction cho các mô hình machine learning là một cách để đo lường và ước tính độ không chắc chắn trong dự đoán của mô hình. Độ không chắc chắn là một khía cạnh quan trọng của dự đoán, đặc biệt là khi muốn biết đến mức độ tin cậy của mô hình trong môi trường thực tế. Dưới đây là một số lý do mà áp dụng conformal prediction là nên và cần thiết:

- Quản lý độ không chắc chắn: Trong nhiều tình huống thực tế, dự đoán của mô hình không thể chắc chắn 100%. Conformal prediction giúp đo lường mức độ không chắc chắn này và cung cấp thông tin về khoảng tin cậy của dự đoán.
- Tính tin cậy cao: Conformal prediction tập trung vào việc cung cấp các kết quả dự đoán kèm theo thông tin về độ tin cậy. Điều này giúp người sử dụng đánh giá được mức độ chắc chắn của mô hình đối với từng dự đoán cụ thể. Trong nhiều ứng dụng quan trọng, độ tin cậy là một yếu tố quan trọng để đảm bảo tính chính xác và đáng tin cậy của các quyết định dự đoán.
- Phát hiện ngoại lệ: Conformal prediction cung cấp khả năng phát hiện ngoại lệ hoặc dữ liệu không chắc chắn. Khi mô hình gặp phải dữ liệu mà nó không quen thuộc, conformal prediction có thể cung cấp thông tin về mức độ không chắc chắn của dự đoán, giúp nhận biết được những trường hợp đặc biệt.
- Ứng dụng trong quyết định y khoa: Trong lĩnh vực y học, đặc biệt là trong việc đưa ra quyết định dựa trên dữ liệu đầu vào của bệnh nhân, việc biết được độ tin cậy của mô hình là quan trọng để hỗ trợ quyết định của bác sĩ và giảm rủi ro liên quan đến quyết định sai lầm.
- Quản lý rủi ro tài chính: Trong lĩnh vực tài chính, nơi mà quyết định dự đoán có thể ảnh hưởng đến việc đầu tư và giao dịch tài chính, việc có thông tin về độ tin cậy là rất quan trọng. Conformal prediction giúp đánh giá được rủi ro liên quan đến các dự đoán tài chính.
- Dễ giải thích: Cung cấp thông tin về độ tin cậy giúp làm cho mô hình trở nên dễ giải thích hơn. Người sử dụng có thể hiểu được tại sao một mô hình đưa ra một quyết định cụ thể và đánh giá được mức độ tin cậy của nó.
- Tính di động và thích ứng: Conformal prediction có thể được áp dụng cho mọi mô hình machine learning, bao gồm cả những mô hình phức tạp như deep learning. Điều này làm cho nó linh hoạt và có thể sử dụng trong nhiều bối cảnh khác nhau.

- Tránh overfitting: Một số mô hình machine learning có thể trở nên quá mức tinh chỉnh để dự đoán dữ liệu huấn luyện, và điều này có thể dẫn đến hiện tượng overfitting. Conformal prediction giúp giảm nguy cơ overfitting bằng cách cung cấp một phương tiện để đo lường độ không chắc chắn và hạn chế sự phụ thuộc quá mức vào dữ liệu huấn luyện.

- Quản lý các tác động của dữ liệu nhiễu: Dữ liệu nhiễu có thể ảnh hưởng đến tính độ chắc chắn của mô hình. Conformal prediction có thể giúp quản lý tác động của dữ liệu nhiễu bằng cách tạo ra khoảng tin cậy cho dự đoán, thay vì chỉ cung cấp một giá trị dự đoán duy nhất.

- Nâng cấp mô hình giúp cung cấp độ tin cậy có giá trị tĩnh cho từng mẫu riêng lẻ.

- Tự động hiệu chỉnh tốt, không phải tốn nhiều công sức.

- Đơn giản để thực thi, suy luận và kiểm định.

- Ứng dụng tốt cho các ứng dụng có sự rủi ro cao.

- Có những giá trị đặc trưng không phải lúc nào cũng chính xác cũng như nhãn (hoặc giá trị) trong tập dữ liệu làm mô hình không thể luôn đưa ra dự đoán chính xác, nên cần có phương pháp để tính độ không chắc chắn đấy cũng như đưa ra vùng dự đoán với độ tin cậy nhất định.

Tầm quan trọng của conformal prediction phụ thuộc vào ứng dụng mà machine learning đang được sử dụng. Dưới đây là một số trường hợp thực tế:

- Muốn biết tính chắc chắn, cận trên, cận dưới của sự liên kết ái lực giữa thuốc và thể thụ cảm trước khi tiến hành thử nghiệm trên thực tế giúp chọn lọc được các hợp chất để chế tạo thuốc, giúp tiết kiệm được chi phí, công sức chế tạo cũng như thí nghiệm vào thực tế.

- Việc ước lượng tính không chắc chắn có thể cải thiện khả năng phát hiện gian lận trong yêu cầu bồi thường bảo hiểm bằng cách cung cấp thêm thông tin cho nhân viên phụ trách đánh giá các yêu cầu bồi thường có khả năng gian lận. Điều này đặc biệt quan trọng khi mô hình machine learning được sử dụng để phát hiện gian lận có dự đoán không chắc chắn. Trong những trường hợp như vậy, nhân viên có thể sử dụng các ước tính không chắc chắn để ưu tiên xem xét khiếu nại và can thiệp nếu cần thiết.

- Ước lượng tính không chắc chắn có thể được sử dụng để cải thiện trải nghiệm người dùng trong ứng dụng ngân hàng. Mặc dù việc phân loại các giao dịch tài chính thành "tiền thuê nhà", "mua sắm", ... phần lớn có thể được tự động hóa thông qua machine learning, nhưng sẽ luôn có những giao dịch khó phân loại. Việc ước lượng tính không chắc chắn có thể xác định các giao dịch phức tạp và nhắc người dùng phân loại chúng.

- Dự báo nhu cầu bằng cách sử dụng machine learning có thể được cải thiện bằng cách sử dụng ước lượng tính không chắc chắn, điều này cung cấp thông tin bổ sung về độ tin cậy trong dự đoán. Điều này đặc biệt quan trọng trong những tình huống mà nhu cầu phải đáp ứng một ngưỡng nhất định để duy trì hoạt động sản xuất. Bằng cách hiểu được sự không chắc chắn của dự báo, tổ chức có thể đưa ra quyết định sáng suốt hơn về việc có nên tiếp tục sản xuất hay không.

- Khi tính toán tiền nong để đầu tư, nếu chỉ dự đoán không thì sẽ mang lại cho người dùng cảm giác không chắc chắn thì cần ước lượng được tính không chắc chắn đấy, thì người dùng mới yên tâm để sử dụng.
- Trong ngành vật lý vật liệu, khi muốn làm loại vật liệu mới, cần dự đoán tính chất của nó, dự đoán xong cũng rất cần biết tính không chắc chắn, nếu không có thì sẽ phải làm rất nhiều thí nghiệm dẫn đến hao tổn chi phí và công sức.

Tóm lại, áp dụng conformal prediction giúp nâng cao khả năng diễn giải và tin cậy của các mô hình machine learning, đặc biệt là khi đối mặt với môi trường thực tế có nhiều độ không chắc chắn.

Chương 2

Các giả thiết (assumption) của conformal prediction

2.1 Dữ liệu độc lập và phân phối giống nhau (Identical and independently distributed data)

Giả thiết tiêu chuẩn là giả thiết identical and independently distributed data (i.i.d.) thông thường được sử dụng trong thống kê: các ví dụ được lấy mẫu độc lập từ một phân phối xác suất không biết nào đó P trên $\mathbf{Z}$ ($\mathbf{Z}$ là một measurable space, được gọi là examples space). Tương đương, dãy vô hạn $z_1, z_2, \ldots$ được rút ra từ phân phối xác suất mũ P^∞ trong $\mathbf{Z}^\infty$.

Giả sử $(X_i, Y_i)_{i=1..n}$ và $X_{\text{test}}, Y_{\text{test}}$ là độc lập và phân phối giống nhau thì $\hat{q}$ được định nghĩa như sau:

$$\hat{q} = \inf\left\{q : \frac{\left|\{i : s(X_i, Y_i\} \leq q\right|}{n} \geq \frac{\lceil (n+1)(1-\alpha) \rceil}{n}\right\} \tag{2.1}$$

và dẫn đến tập dữ đoán là

$$\mathcal{C}(X) = \{y : s(X, y) \leq \hat{q}\} \tag{2.2}$$

thì

$$P\Big(Y_{\text{test}} \in \mathcal{C}(X_{\text{test}})\Big) \geq 1 - \alpha \tag{2.3}$$

2.2 Khả năng trao đổi (Exchangeability)

Exchangeability là một giả thiết (assumption) quan trọng, nếu dữ liệu được sử dụng cho calibration rất khác với dữ liệu muốn ước lượng tính không chắc chắn của dự đoán thì mức độ đảm bảo về độ phủ (coverage) sẽ giảm đi.

Để duy trì đảm bảo tính độ phủ, dữ liệu calibration phải "exchangeable" với dữ liệu mới. Ví dụ: nếu chọn ngẫu nhiên từ cùng một phân phối, chúng là exchangeable. Nếu chúng đến từ các bản phân phối khác nhau, chúng có thể không exchangeable. Dữ liệu dạng chuỗi thời gian (time series data) là không exchangeable vì thứ tự thời gian rất quan trọng.

Exchangeability khó nhận thấy hơn hơn giả thiết i.i.d.; exchangeability của $Y_1, \ldots, Y_{n+1}$ có nghĩa là phân phối đồng thời của chúng là không thay đổi theo các hoán vị:

$$(Y_1, \ldots, Y_{n+1}) \stackrel{d}{=} (Y_{\sigma(1)}, \ldots, Y_{\sigma(n+1)}), \quad \text{với mọi hoán vị } \sigma. \tag{2.4}$$

2.3 Sự lựa chọn của hàm f

Với không gian quan sát (observation space) $Z^{m+1} : X_m \times Y$, sự lựa chọn của hàm $f : Z^{m+1} \mapsto \mathbb{R}$:

- Nói chung, việc lựa chọn f là không liên quan đối với sự hợp lệ (validity).
- Đối với hồi quy, f nên là khả nghịch.

2.4 Nhiễu độc lập và đối xứng (Independent and symmetric noise)

Với $(X, Y) \sim P$, biến nhiễu $\epsilon = Y - \mu(X)$ độc lập với X và hàm mật độ của ϵ đối xứng quanh 0 và không tăng trên $[0, \infty)$.

Giả thiết này yếu hơn so với những giả thiết thường được đưa ra khác. Đặc biệt, thậm chí không yêu cầu ϵ phải có moment bậc nhất hữu hạn. Các điều kiện đối xứng và đơn điệu chỉ là để thuận tiện, và có thể được loại bỏ bằng cách xem xét các nhóm phân vị hoặc mức mật độ thích hợp của ϵ. Sự liên tục phân phối của ϵ cũng đảm bảo rằng với xác suất 1, các thặng dư được khớp sẽ dễ thấy, làm cho việc đảo nghịch hàm phân phối kinh nghiệm trở nên dễ dàng.

2.5 Độ ổn định lấy mẫu (Sampling stability)

Với n đủ lớn,

$$\mathbb{P}(\|\hat{\mu}_n - \tilde{\mu}\|_\infty \geq \eta_n) \leq \rho_n \tag{2.5}$$

trong đó, $\eta_n = o(1)$, $\rho_n = o(1)$ khi $n \to \infty$, và một số hàm $\tilde{\mu}$.

Không cần giả định rằng $\tilde{\mu}$ gần với hàm thực μ. Chỉ cần bộ ước lượng $\hat{\mu}_n$ tập trung xung quanh $\tilde{\mu}$. Điều này chỉ là một giả thiết về sự ổn định thay vì giả thiết về nhất quán. Ví dụ, điều này được đáp ứng trong hồi quy phi tham số dưới hiện tượng làm mịn quá mức (over-smoothing). Khi $\tilde{\mu} = \mu$, điều này trở thành một giả thiết về tính nhất quán sup-norm, và nó thỏa mãn. Thông thường, η_n có dạng $c(\log n/n)^{-r}$, và ρ_n có thứ bậc n^{-c}, với một số cố định $c > 0$ (sự chọn lựa của hằng số c là tùy ý và chỉ tác động đến số hạng không đổi trước η_n).

2.6 Độ nhạy nhiễu loạn (Perturb-one sensitivity)

Với n đủ lớn,

$$\mathbb{P}\left(\sup_{y \in \mathcal{Y}} \|\hat{\mu}_n - \hat{\mu}_{n,(X,y)}\|_\infty \geq \eta_n\right) \leq \rho_n \tag{2.6}$$

trong đó, $\eta_n = o(1)$, $\rho_n = o(1)$ khi $n \to \infty$.

Điều kiện perturb-one sensitivity yêu cầu rằng hàm đã được khớp không thay đổi nhiều nếu chỉ làm biến đổi giá trị y của mục dữ liệu cuối cùng.

2.7 Tính nhất quán của bộ ước lượng cơ sở (Consistency of base estimator)

Với n đủ lớn,

$$\mathbb{P}\left(\mathbb{E}_X\left[(\hat{\mu}_n(X) - \mu(X))^2 \,|\, \hat{\mu}_n\right] \geq \eta_n\right) \leq \rho_n \tag{2.7}$$

trong đó, $\eta_n = o(1)$, $\rho_n = o(1)$ khi $n \to \infty$.

Dễ dàng kiểm chứng giả thiết được suy ra từ điều kiện $\mathbb{E}\Delta_n^2(X) = o(1)$, bằng cách sử dụng bất đẳng thức Markov. Nhiều bộ ước lượng nhất quán thường có tính chất này.

2.8 Mô hình được hiệu chuẩn tốt (Well-Calibrated Model)

Mô hình phải được hiệu chỉnh một cách chính xác để đảm bảo rằng mức tin cậy được dự đoán tương ứng với tỷ lệ các trường hợp đúng. Điều này đảm bảo rằng mức tin cậy là đáng tin cậy và không quá lạc quan hoặc quá tiêu cực.

2.9 Sự hợp lý của mô hình thống kê (Validity of the Statistical Model)

Conformal prediction yêu cầu sự hợp lý của mô hình thống kê được sử dụng để xác định độ tin cậy. Điều này đảm bảo rằng mức tin cậy được xây dựng trên cơ sở của mô hình đó có ý nghĩa và có giả định thống kê hợp lý.

2.10 Tính không đồng nhất (Heteroscedasticity)

Conformal prediction giả định rằng sự không chắc chắn của mô hình có thể thay đổi tùy thuộc vào điểm dữ liệu cụ thể. Điều này được gọi là heteroscedasticity. Thay vì giả định một sự không chắc chắn đồng đều trên toàn bộ không gian dữ liệu, conformal prediction có thể xem xét sự không chắc chắn tại mỗi điểm dữ liệu cụ thể.

2.11 Số mẫu hữu hạn (Finite-Sample)

Các kết quả từ conformal prediction thường dựa trên giả định về kích thước tập dữ liệu hữu hạn. Điều này có thể áp dụng trong các tình huống khi có một số hữu hạn các mẫu dữ liệu để huấn luyện và suy luận của mô hình.

Chương 3

Một số phương pháp conformal prediction

Mục tiêu của các bài toán phân lớp tiêu chuẩn là phân loại một đối tượng thành một trong một số lớp riêng biệt. Thay vào đó, các bộ phân lớp hợp lý (conformal classifiers) tính toán và đưa ra p-value cho mỗi lớp có sẵn bằng cách thực hiện một bảng xếp hạng của đo lường độ không hợp lý (giá trị α) của đối tượng kiểm tra so với các mẫu từ tập huấn luyện. Tương tự như hypothesis testing tiêu chuẩn, p-value cùng với một ngưỡng (được gọi là significance level trong lĩnh vực Conformal prediction) được sử dụng để xác định liệu nhãn có nên thuộc vào tập dự đoán hay không. Ví dụ, đối với significance level là 0.1, tất cả các lớp có p-value là từ 0.1 trở lên sẽ được thêm vào tập dự đoán. Các thuật toán transductive tính độ không hợp lý bằng cách sử dụng tất cả dữ liệu huấn luyện có sẵn, trong khi các thuật toán inductive chỉ tính trên một tập con của tập huấn luyện.

Conformal prediction ban đầu được tạo ra cho bài toán phân lớp, nhưng sau đó đã được điều chỉnh để áp dụng cho cả hồi quy. Khác với bài toán phân lớp, nơi đầu ra là các p-value mà không cần một significance level cụ thể, bài toán hồi quy yêu cầu một significance level cố định vào thời điểm dự đoán để tạo ra khoảng dự đoán cho một đối tượng mới. Đối với conformal regression kinh điển, không có thuật toán transductive, do không thể giả định tất cả các nhãn có thể có cho một đối tượng, vì không gian nhãn là liên tục. Các thuật toán có sẵn đều được thiết kế cho thiết lập inductive, tính toán một quy tắc dự đoán một lần và áp dụng nó cho tất cả các dự đoán trong tương lai.

3.1 Inductive conformal prediction (ICP)

ICP (Inductive Conformal Prediction) yêu cầu phải chia dữ liệu. Giả sử có một tập dữ liệu $D = \{(x_i, y_i)\}_{i=1}^{n}$, trong đó $x_i \in \mathbb{R}^D$ là các đặc trưng và $y_i \in \mathbb{R}^Q$ là biến phản ứng (response variable). Như đã nói, bước đầu tiên là chia D thành ba tập tương đối độc lập: (i) tập huấn luyện (training set), (ii) tập calibration và (iii) tập kiểm định (validation set), sao cho $D = D_{train} \cup D_{cal} \cup D_{val}$ với $N = n_{train} + n_{cal} + n_{val}$. Tổng quan chung về ICP như sau:

1. Huấn luyện một mô hình machine learning trên tập huấn luyện $D_{train} = \{(x_i, y_i\}_{i=1}^{n_{train}}$
2. Định nghĩa một khái niệm heuristic về độ không chắc chắn được biểu diễn bằng hàm $s(x, y)$, thường được gọi là hàm điểm không hợp lý (non-conformity score function).

3. Đối với mỗi $(x, y) \in D_{cal} = \{(x_i, y_i)\}_{i=1}^{n_{cal}}$, áp dụng hàm s để thu được n_{cal} điểm không hợp lý $\{(s_i)\}_{i=1}^{n_{cal}}$.

4. Chọn một tỷ lệ lỗi miscoverage α và tính $\hat{q}$ như là $\frac{\lceil (n_{cal}+1)(1-\alpha) \rceil}{n_{cal}}$ quantile của điểm không hợp lý $\{(s_i)\}_{i=1}^{n_{cal}}$.

5. Với $\hat{q}$ được tính trong bước trước, tạo ra các khoảng tin cậy (confidence interval) hoặc tập hợp trong giai đoạn kiểm định, được ký hiệu là $C(x_{val})$ với $1-\alpha$ coverage.

3.2 Full Conformal Prediction

Có cách nào để coverage mà không phải chia dữ liệu? Đó là *Full conformal prediction*. Phương pháp này thường tốn kém và phức tạp hơn nhiều so với phiên bản chia dữ liệu của nó, nhưng vẫn là một ý tưởng hay và quan trọng — và ở một số trường hợp, nó thực sự có thể được tính toán một cách hiệu quả.

Trong full conformal prediction, xây dựng các phần thặng dư theo cách xử lý tất cả dữ liệu một cách đối xứng. Chọn $x \in \mathcal{X}$, muốn xác định liệu giá trị phản hồi cụ thể nào đó $y \in \mathbb{R}$ có nên thuộc vào tập dự đoán $\hat{C}_n(x)$ hay không. Gọi y là giá trị *thử nghiệm* (trial) hoặc *truy vấn* (query). Sau đó, huấn luyện mô hình dự đoán trên $(X_1, Y_1), \ldots, (X_n, Y_n), (x, y)$ — lưu ý đây là một tập huấn luyện được *mở rộng*, với $n+1$ điểm — được dùng để tạo ra một bộ dự đoán $\hat{f}_{n,(x,y)}$. Định nghĩa các phần thặng dư:

$$R_i^{(x,y)} = |Y_i - \hat{f}_{n,(x,y)}(X_i)|, \quad i = 1, \ldots, n \tag{3.1}$$

$$R_{n+1}^{(x,y)} = |y - \hat{f}_{n,(x,y)}(x)| \tag{3.2}$$

Tập hợp lý (conformal set):

$$\hat{C}_n(x) = \left\{ y : R_{n+1}^{(x,y)} \leq \lceil (1-\alpha)(n+1) \rceil \text{ smallest of } R_1^{(x,y)}, \ldots, R_n^{(x,y)} \right\} \tag{3.3}$$

Ta được:

$$\mathbb{P}\Big(Y_{n+1} \in \hat{C}_n(X_{n+1})\Big) \in \Big[1-\alpha,\, 1-\alpha+\frac{1}{n+1}\Big) \tag{3.4}$$

Sau khi thay điểm kiểm tra ngẫu nhiên (random test point), và rút gọn

$$R_i = R_i^{(X_{n+1}, Y_{n+1})}, \quad i = 1, \ldots, n+1 \tag{3.5}$$

Tất cả các phần thặng dư này đều có thể hoán đổi được (exchangeable) (chỉ đúng nếu thuật toán sử dụng để điều chỉnh bộ dự đoán $\hat{f}_{n,(x,y)}$ là một hàm đối xứng của dữ liệu huấn luyện mà nó nhận làm đầu vào, tức là không sử dụng thông tin về thứ tự mà các điểm huấn luyện đã được truyền vào). Do đó:

$$Y_{n+1} \in \hat{C}_n(X_{n+1}) \iff R_{n+1} \leq \lceil (1-\alpha)(n+1) \rceil \text{ smallest of } R_i,\ i = 1, \ldots, n \tag{3.6}$$

xảy ra với xác suất ít nhất là $1-\alpha$, và tối đa là $1-\alpha+1/(n+1)$ nếu những phần thặng dư hầu hết chắc chắn là khác nhau.

Tóm tắt lại phương pháp như sau:

- Bất kỳ hàm điểm hướng âm (negatively-oriented) và đối xứng phù hợp nào cũng có thể được sử dụng thay thế cho điểm thặng dư tuyệt đối. Định nghĩa như sau:

$$R_i^{(x,y)} = V\Big((X_i, Y_i);\ (X_1, Y_1), \ldots, (X_n, Y_n), (x, y)\Big) \quad i = 1, \ldots, n, \tag{3.7}$$

$$R_{n+1}^{(x,y)} = V\Big((x, y);\ (X_1, Y_1), \ldots, (X_n, Y_n), (x, y)\Big) \tag{3.8}$$

  với bất kỳ hàm V nào đối xứng trong $n+1$ đối số cuối cùng của nó. Ví dụ, huấn luyện một bộ dự đoán trên $n+1$ đối số cuối cùng—miễn là nó xử lý chúng theo cách đối xứng—và sau đó sử dụng nó để trả về một số điểm (score) cho đối số đầu tiên. Sau đó, tập dự đoán hợp lý trong (3.3) vẫn bảo đảm giống như trong (3.4).

- Có thể viết lại tập hợp lý (3.3) dưới dạng quantile tương đương và hàm phân phối tích lũy (CDF):

$$\hat{C}_n(x) = \left\{ y : R_{n+1}^{(x,y)} \leq \text{Quantile}\left(\frac{\lceil (1-\alpha)(n+1) \rceil}{n};\ \frac{1}{n} \sum_{i=1}^{n} \delta_{R_i^{(x,y)}} \right) \right\} \tag{3.9}$$

$$= \left\{ y : \frac{1}{n} \sum_{i=1}^{n} 1\left\{ R_i^{(x,y)} \leq R_{n+1}^{(x,y)} \right\} \leq \frac{\lceil (1-\alpha)(n+1) \rceil}{n} \right\}. \tag{3.10}$$

- Luôn có thể thêm vào ngẫu nhiên phụ trợ (auxiliary randomness) để đạt được tỷ lệ bao phủ (coverage) chính xác là $1-\alpha$ trong (3.4).

3.3 Mondrian inductive conformal prediction (MICP)

Trong các bộ dự đoán hợp lý Mondrian, các ví dụ calibration có sẵn được chia thành các category khác nhau và sau đó một bộ dự đoán kiểm định hợp lý được xây dựng cho từng category. Bộ dự đoán hợp lý Mondrian phổ biến nhất là bộ dự đoán hợp lý có điều kiện (conditional conformal predictor), trong đó các category đại diện cho các nhãn lớp có thể có, do đó cung cấp sự đảm bảo cho mỗi nhãn, tức là các lỗi sẽ được phân bổ đều trên các lớp. Không gian bài toán cũng có thể được chia đối với không gian đặc trưng; ví dụ: đối với mô hình cây, một cách phân chia rất tự nhiên là coi mỗi lá (đường đi) là một category riêng biệt, dẫn đến mỗi lá như vậy có giá trị độc lập. Gần đây, các bộ hồi quy hợp lý (conformal regressor) Mondrian đã được đề xuất. Hệ thống dự đoán hợp lý (conformal predictive system) Mondrian (split) tạo ra một phân phối dự đoán hợp lý theo kiểu Mondrian theo cách sau, với dãy huấn luyện Z_{tr} và đối tượng kiểm tra x:

1. Chia dãy huấn luyện Z_{tr} thành hai tập con không giao nhau: tập huấn luyện thích hợp (proper training set) Z_t và tập calibration $Z_c = (x_1, y_1), \ldots, (x_q, y_q)$.

2. Huấn luyện mô hình cơ bản h bằng cách sử dụng tập Z_t.

3. Chia tập Z_c thành k tập con không giao nhau $Z_{c1}, \ldots, Z_{ck}$, theo một nguyên tắc phân loại Mondrian κ với các category $\kappa_1, \ldots, \kappa_k$.

4. Đối với mỗi category κ_j, đặt $\hat{y}_{ji} = h(x_{ji})$ và đặt $\hat{\sigma}_{ji}$ là ước lượng chất lượng tương ứng, với mỗi $x_{ji} \in Z_{cj}$.

5. Thực hiện dự đoán điểm cho đối tượng kiểm tra $\hat{y} = h(x)$ và ước lượng chất lượng của nó là $\hat{\sigma}$.

6. Xác định đối tượng kiểm tra thuộc category κ_j nào và tạo ra một danh sách các điểm calibration bằng cách dùng:

$$C_{ji} = \hat{y} + \frac{\hat{\sigma}}{\hat{\sigma}_{ji}}(y_{ji} - \hat{y}_{ji}) \tag{3.11}$$

   với $i \in 1, \ldots, q$ và $q = |Z_{cj}|$

7. Sắp xếp $C_{ji}, \ldots, C_{jq}$ tăng dần, được $C_{j(1)}, \ldots, C_{j(q)}$

8. Đặt $C_{j(0)} = -\infty$ và $C_{j(q+1)} = \infty$

9. Cho $\tau \in U(0, 1)$

10. Trả về phân phối dự đoán:

$$Q(y) = \begin{cases} \dfrac{n+\tau}{q+1} & \text{nếu } y \in (C_{j(n)}, C_{j(n+1)}) \text{ với } n \in \{0, \ldots, q\} \\ \dfrac{n' - 1 + (n'' - n' + 2)\tau}{q+1} & \text{nếu } y = C_{j(n)} \text{ với } n \in \{1, \ldots, q\} \end{cases} \tag{3.12}$$

    Trong đó, $n' = \min\{m | C_{j(m)} = C_{j(n)}\}$ và $n'' = \max\{m | C_{j(m)} = C_{j(n)}\}$

Đối với phân lớp hợp lý Mondrian, các nhãn lớp có thể được sử dụng để định nghĩa các category, trong khi đối với hồi quy hợp lý Mondrian, các category được định nghĩa bằng cách sử dụng ước lượng chất lượng σ. Ở đây, việc tạo ra các category của nguyên tắc phân lớp Mondrian thông qua chia thành các dự đoán, sử dụng các bin có kích thước bằng nhau, tương tự như vấn đề phân lớp trong bối cảnh của dự đoán Venn. Để xử lý trường hợp đặc biệt khi số lượng dự đoán giống nhau lớn hơn kích thước bin, giả định rằng một số nhỏ ξ ngẫu nhiên được thêm vào mỗi dự đoán, điều này cho phép một lượng gần giống nhau các ví dụ sẽ thuộc vào mỗi bin (category). Số lượng bin (category) là do đó là một tham số của phương pháp, và nó nên được chọn sao cho kích thước của mỗi phần của tập calibration đủ lớn, cho phép các khoảng dự đoán với độ tin cậy được trích xuất.

3.4 Least Ambiguous set-valued Classifier (LAC)

Trong phương pháp LAC, điểm số hợp lý (conformity score) được định nghĩa là một trừ cho điểm số của nhãn thực sự. Đối với mỗi quan sát i trong tập calibration:

$$s_i(X_i, Y_i) = 1 - \hat{\mu}(X_i)_{Y_i} \tag{3.13}$$

Vì các điểm số hợp lý $s_1, \ldots, s_n$ được ước lượng cho tất cả các quan sát trong tập calibration, tính $(n+1) * (1-\alpha)/n$ quantile $\hat{q}$ như sau:

$$\hat{q} = \text{Quantile}\left(s_1, \ldots, s_n; \frac{\lceil (n+1)(1-\alpha) \rceil}{n}\right) \tag{3.14}$$

Cuối cùng, xây dựng một tập dự đoán bao gồm tất cả các nhãn có điểm số cao hơn quantile được ước lượng:

$$\hat{C}(X_{test}) = \{y : \hat{\mu}(X_{test})_y \geq 1 - \hat{q}\} \tag{3.15}$$

Phương pháp đơn giản này cho phép xây dựng các tập dự đoán có sự bảo đảm lý thuyết về độ phủ cận biên (marginal coverage). Tuy nhiên, phương pháp này thường dẫn đến các tập dự đoán nhỏ, nó có thể tạo ra các tập trống khi mô hình không chắc chắn, ví dụ như ở biên giữa hai lớp.

3.5 Top-K

Đặc điểm độc đáo của phương pháp Top-K là nó sẽ đưa ra cùng kích thước tập dự đoán cho tất cả các quan sát. Điểm số hợp lý (conformity score) là hạng của nhãn thực sự, với các điểm số được xếp hạng từ cao đến thấp. Các tập dự đoán được xây dựng bằng cách lấy $\hat{q}^{th}$ điểm số cao nhất. Quy trình được mô tả trong các phương trình sau:

$$s_i(X_i, Y_i) = j \quad \text{với} \quad Y_i = \pi_j \quad \text{và} \quad \hat{\mu}(X_i)_{\pi_1} > ... > \hat{\mu}(X_i)_{\pi_j} > ... > \hat{\mu}(X_i)_{\pi_n} \tag{3.16}$$

$$\hat{q} = \left\lceil \text{Quantile}\left(s_1, \ldots, s_n; \frac{\lceil (n+1)(1-\alpha) \rceil}{n} \right) \right\rceil \tag{3.17}$$

$$\hat{C}(X_{test}) = \{\pi_1, \ldots, \pi_{\hat{q}}\} \tag{3.18}$$

Như các phương pháp khác, quy trình này cho phép xây dựng các tập dự đoán với sự bảo đảm về độ phủ cận biên.

3.6 Adaptive Prediction Sets (APS)

Phương pháp APS vượt qua vấn đề gặp phải bởi phương pháp LAC thông qua việc xây dựng các tập dự đoán luôn không trống. Điểm số hợp lý (conformity score) được tính bằng cách cộng tổng điểm số được xếp hạng của mỗi nhãn, từ cao đến thấp cho đến khi đạt đến nhãn thực sự của quan sát:

$$s_i(X_i, Y_i) = \sum_{j=1}^{k} \hat{\mu}(X_i)_{\pi_j} \quad \text{với} \quad Y_i = \pi_k \tag{3.19}$$

Quantile $\hat{q}$ sau đó được tính theo cách tương tự như phương pháp LAC. Đối với việc xây dựng các tập dự đoán cho một quan sát mới, quy trình lấy tổng điểm số xếp hạng tương tự được áp dụng cho đến khi đạt đến quantile, như mô tả trong phương trình sau:

$$\hat{C}(X_{test} = \{\pi_1, \ldots, \pi_k\} \quad \text{với} \quad k = \inf\{k : \sum_{j=1}^{k} \hat{\mu}(X_{test})_{\pi_j} \geq \hat{q}\} \tag{3.20}$$

Mặc định, nhãn có tổng điểm cao hơn quantile sẽ được bao gồm trong tập dự đoán. Tuy nhiên, việc bao gồm nó cũng có thể được chọn ngẫu nhiên dựa trên sự khác biệt giữa điểm tích lũy và quantile để đảm bảo có độ phủ (coverage) hiệu quả vẫn gần với độ phủ mục tiêu là cận biên (marginal).

3.7 Regularized Adaptive Prediction Sets (RAPS)

Phương pháp RAPS là một cải tiến của phương pháp APS ở trên. Sự điều chuẩn (regularization) có thể vượt qua các tập dự đoán rất lớn được đưa ra bởi phương pháp APS. Điểm số hợp lý (conformity score) được tính bằng cách cộng tổng điểm số được điều chuẩn và xếp hạng của mỗi nhãn, từ cao đến thấp cho đến khi đạt đến nhãn thực sự của quan sát:

$$s_i(X_i, Y_i) = \sum_{j=1}^{k} \hat{\mu}(X_i)_{\pi_j} + \lambda(k - k_{reg})^+ \quad \text{với} \quad Y_i = \pi_k \tag{3.21}$$

Trong đó:

- π_i là nhãn được liên kết với điểm số được xếp hạng thứ i.
- $(z)^+$ là phần dương của z.
- k_{reg} là kích thước tập hợp tối ưu (nghĩa là nếu tất cả các tập dự đoán có k_{reg} phần tử, thì ta đạt được độ phủ mong muốn).
- λ là một tham số điều chuẩn (regularization parameter).

Các tối ưu hóa của k_{reg} và λ đòi hỏi một phần dữ liệu phụ (thông thường là 20% của dữ liệu calibration). Để chọn k_{reg}, cần chạy phương pháp Top-K trên phần dữ liệu mới này. Đối với việc chọn λ, cố gắng tìm giá trị của λ sao cho nó giảm thiểu kích thước của các tập dự đoán. Một Grid Search đơn giản được thực hiện trên các giá trị khác nhau của λ (thường sẽ chọn $\lambda \in \{0.001, 0.01, 0.1, 0.2, 0.5\}$).

Đối với việc xây dựng tập dự đoán cho một quan sát mới, quy trình sau được áp dụng:

$$\hat{C}(X_{test}) = \{\pi_1, \ldots, \pi_k\} \quad \text{với} \quad k = \inf\{k : \sum_{j=1}^{k} \hat{\mu}(X_{test})_{\pi_j} + \lambda(k - k_{reg})^+ \geq \hat{q}\} \tag{3.22}$$

Một cách hiểu đơn giản, mục tiêu của phương pháp là phạt các tập dự đoán có kích thước lớn hơn kích thước tối ưu của tập dự đoán. Mức độ điều chuẩn này được kiểm soát bởi tham số λ.

Mặc dù phương pháp RAPS có kích thước tập tương đối nhỏ, nhưng độ phủ (coverage) của nó có xu hướng cao hơn so với yêu cầu (đặc biệt là đối với các giá trị cao của α, có nghĩa là mức tin cậy thấp). Do đó, để đạt được độ phủ chính xác, có thể thực hiện một quá trình ngẫu nhiên liên quan đến nhãn cuối cùng trong tập dự đoán. Quá trình ngẫu nhiên này được thực hiện như sau:

- Đầu tiên: định nghĩa tham số V:

$$V_i = \frac{s_i(X_i, Y_i) - \hat{q}_{1-\alpha}}{\hat{\mu}(X_i)_{\pi_k} + \lambda \mathbb{1}(k > k_{reg})} \tag{3.23}$$

- So sánh từng V_i với $U \sim \text{Unif}(0, 1)$
- Nếu $V_i \leq U$, nhãn cuối cùng trong tập sẽ bị loại bỏ; ngược lại, giữ nguyên tập dự đoán.

3.8 Split conformal prediction

Xét một phân hoạch của $[n]$ thành một tập calibration L có kích thước w và một tập kiểm định (validation set) I có kích thước $m = n - w$, độc lập với các giá trị dữ liệu quan sát. $R = R(Z_L, Z_{n+1})$ được gọi là điểm hợp lý (conformity score), như một đơn vị đo của giá trị y như một phép thể hiện của Y_{n+1} cho giá trị quan sát của X_{n+1}.

$$R = |Y_{n+1} - \hat{\mu}_L(X_{n+1})| \tag{3.24}$$

với $\hat{\mu}_L$ là một ước lượng của $\mathbb{E}(Y|X)$ dựa trên $(Z_l)_{l\in L}$ và

$$R = \max\left\{\hat{q}_L^{\gamma}(X_{n+1}) - Y_{n+1}, Y_{n+1} - \hat{q}_L^{1-\gamma}(X_{n+1})\right\} \tag{3.25}$$

với $\hat{q}_L^{\gamma}$ là một ước lượng γ-quantile của $Y|X$. Tập kiểm định (validation set) $I = \{j_1, \ldots, j_m\}$ và đặt

$$R_i = R((Z_l)_{l\in L}, Z_{j_i}), \quad i \in [m] \tag{3.26}$$

Với $\alpha \in (0, 1)$, định nghĩa một quantile $R_\alpha = R_{\lceil (1-\alpha)(m+1)\rceil}$, trong đó $R_1 \leq ... \leq R_m$ là các thống kê có thứ tự bằng cách sắp xếp $R_1, \ldots, R_m$ theo thứ tự không giảm với các giá trị bằng nhau được giải quyết theo bất kỳ cách nào.

Thuật toán dưới đây mô tả làm thế nào để tìm tập dự đoán hợp lý tách (split conformal prediction set):

Algorithm 1 Split Conformal

Require: dữ liệu $(x_1, y_1), \ldots, (x_n, y_n), x_{n+1}$, tập kiểm định kích thước m, thống kê R, mức $\alpha \in (0, 1)$

1: Chia $[n]$ thành L kích thước w và I kích thước $m = n - w$

2: Tính $\{R_i\}_{i=1}^m$ và $R_\alpha = R_{(\lceil (1-\alpha)(m+1)\rceil)}$

return tập dự đoán hợp lý tách $C_\alpha(x_{n+1}) = \{y \in \mathbb{R} : R \leq R_\alpha\}$

Cụ thể, đối với R được định nghĩa như trong (3.24) và (3.25), thuật toán 1 trả về $C_\alpha(x_{n+1}) = [\hat{\mu}_L(x_{n+1}) - R_\alpha, \hat{\mu}_L(x_{n+1}) + R_\alpha]$ và $C_\alpha(x_{n+1}) = [\hat{q}_L^{\gamma}(x_{n+1}) - R_\alpha, \hat{q}_L^{1-\gamma}(x_{n+1}) + R_\alpha]$, tương ứng. Cái trước luôn là một khoảng, trong khi cái sau có thể là một khoảng hoặc một tập rỗng, $C_\alpha(x_{n+1}) = \emptyset$ nếu và chỉ nếu $R_\alpha < \frac{1}{2}[\hat{q}_L^{\gamma}(x_{n+1}) - \hat{q}_L^{1-\gamma}(x_{n+1})]$.

3.9 Multi Split Conformal Prediction

Split conformal prediction là một phương pháp tính toán hiệu quả để thực hiện suy luận dự đoán không phân bố. Tuy nhiên, nó liên quan đến sự phân chia ngẫu nhiên dữ liệu một lần, và kết quả có thể phụ thuộc nhiều vào việc chia đấy. Để giải quyết vấn đề này, *multi split conformal prediction*, một phương pháp đơn giản dựa trên bất đẳng thức Markov để tổng hợp các khoảng của split conformal prediction qua nhiều lần chia tách.

Phương pháp chia đa lần (multi split) bao gồm việc xây dựng nhiều lần các tập dự đoán của một lần chia đơn và tiếp tục tổng hợp kết quả bằng cách bao gồm các điểm mà được bao gồm trong các khoảng dự đoán của lần chia đơn với tần suất lớn hơn một ngưỡng.

Trước tiên, chọn số lần chia $B \in \mathbb{N}$. Sau đó, phân hoạch $[n]$ thành $L^{[b]}$ có kích thước $w^{[b]}$ và $I^{[b]}$ có kích thước $m^{[b]} = n - w^{[b]}$, độc lập với các giá trị dữ liệu quan sát và chọn một thống kê $R^{[b]}$ với $b = 1, \ldots, B$. Với $\beta \in (0,1)$, biến ngẫu nhiên Bernoulli $\phi_\beta^{[b]} = \mathbb{1}\{R^{[b]} > R_\beta^{[b]}\}$ có giá trị kì vọng $E(\phi_\beta^{[b]}) \leq \beta$. Đặt:

$$V_\beta = \sum_{b=1}^{B} \phi_\beta^{[b]} \tag{3.27}$$

là số lần thành công. Định lý sau đây cung cấp một cận trên cho $\text{pr}(V_\beta \geq k)$ là xác suất ít nhất k lần thành công trong n phép thử.

Theorem 1 *Đặt λ là một số nguyên không âm, với một số nguyên đã cho $1 \leq k \leq B$ và $\beta \in (0,1)$, điều sau đây đúng:*

$$\sum_{u=0}^{k-1} pr(V_\beta \in [k-u, k)) \geq \sum_{u=0}^{\lambda} pr(V_\beta \in [k, k+u)) \tag{3.28}$$

thì

$$pr(V_\beta \geq k) \leq \frac{B\beta}{k+\lambda} \tag{3.29}$$

Tham số λ có thể được coi là một tham số làm mịn (smoothing parameter). Định lý 1 có thể được sử dụng để tổng hợp kết quả của suy luận split conformal trên nhiều phân chia dữ liệu khác nhau. Đặt:

$$\Pi_\beta = 1 - \frac{V_\beta}{B} = \frac{1}{B} \sum_{b=1}^{B} \mathbb{1}\{Y_{n+1} \in C_\beta^{[b]}(X_{n+1})\} \tag{3.30}$$

là tỷ lệ các tập dự đoán $C_\beta^{[b]}(X_{n+1})$ chứa Y_{n+1}. Với $\alpha \in (0,1)$ và một ngưỡng $\tau = 1 - k/B$, tập dự đoán hợp lý chia đa lần (multi split conformal prediction set) được định nghĩa là:

$$C_\alpha^\tau(x_{n+1}) = \{y \in \mathbb{R} : \Pi_\beta^y > \tau\} \tag{3.31}$$

có độ che phủ (coverage) ít nhất là $1 - \alpha$ với $\phi_\alpha = \mathbb{1}\{V_\beta \geq k\} = \mathbb{1}\{\Pi_\beta \leq \tau\}$, trong đó, $\beta = \alpha(1-\tau)$ với không có điều kiện nào, hoặc là $\beta = \alpha(1 - \tau + \lambda/B)$ dưới điều kiện 3.28 của Định lý 1 đảm bảo $E(\phi_\alpha) \leq \alpha$.

Thuật toán 2 mô tả cách tính tập dự đoán hợp lý chia đa lần (multi split conformal prediction set).

Algorithm 2 Multi Split Conformal

Require: dữ liệu $(x_1, y_1), \ldots, (x_n, y_n), x_{n+1}$, số lần chia $B \in \mathbb{N}$, tập calibration kích thước $(m^{[b]})_{b=1}^B$, thống kê $(R^{[b]})_{b=1}^B$, ngưỡng $\tau \in [0, (B-1)/B]$, mức $\alpha \in (0,1)$, tham số làm mịn $\lambda \in \mathbb{N}_0$.

1: **for** $b \leftarrow 1$ đến B **do**
2: tính $C_\beta^{[b]}(x_{n+1})$ bằng thuật toán 1 với $m^{[b]}$, $R^{[b]}$ và mức $\beta = \alpha(1 - \tau + \lambda/B)$
3: **end for**

return tập dự đoán hợp lý chia đa lần $C_\alpha^\tau(x_{n+1}) = \{y \in \mathbb{R} : \Pi_\beta^y > \tau\}$

Nếu việc tính toán mỗi khoảng chia đơn $C_\beta^{[b]}$ mất thời gian $\leq T$, thì tổng thời gian để tính C_α^τ là $O(BlogB) + BT$. Lưu ý rằng tập dự đoán chia đa lần C_α^τ rất linh

hoạt vì nó cho phép sử dụng các phần chia có tỷ lệ m/n khác nhau và có thể là các thống kê S khác nhau qua các lần chia. Sự linh hoạt này có thể đặc biệt hữu ích trong conformal quantile regression vì nó cho phép xem xét các giá trị khác nhau cho quantile γ trong 3.25. Giá trị của γ và tỷ lệ m/n là quan trọng đối với hiệu suất của conformal quantile regression.

3.10 Cross-Conformal Prediction

Cross-Conformal Prediction (CCP) được định nghĩa như sau. Tập huấn luyện được chia thành K tập con không rỗng (*fold*) z_{S_k}, $k = 1, \ldots, K$, trong đó $K \in \{2, 3, \ldots\}$ là một tham số của thuật toán và $(S_1, \ldots, S_K)$ là một sự phân bố của $\{1, \ldots, l\}$. Với mỗi $k \in \{1, \ldots, K\}$ và mỗi nhãn tiềm năng $y \in \mathbf{Y}$ của x tìm điểm số hợp lý (conformity score) của các ví dụ trong z_{S_k} và của (x, y) bằng cách:

$$\alpha_{i,k} := A(z_{S_{-k}}, z_i), \quad i \in S_k, \qquad \alpha_k^y := A(z_{S_{-k}}, (x, y)) \tag{3.32}$$

với $S_{-k} := \cup_{j \neq k} S_j$ và A là một độ đo inductive conformity. Các p-value tương ứng được định nghĩa bởi:

$$p^y := \frac{\sum_{k=1}^{K} |\{i \in S_k \mid \alpha_{i,k} \leq \alpha_k^y\}| + 1}{l + 1}. \tag{3.33}$$

Định nghĩa của CCP tương tự như ICP (section 3.1), ngoại trừ việc sử dụng toàn bộ tập huấn luyện cho việc calibration. Điểm hợp lý (3.32) được tính bằng cách sử dụng hợp của tất cả các fold ngoại trừ fold hiện tại làm tập huấn luyện riêng. Calibration (3.33) được tính bằng cách kết hợp hạng (rank) của ví dụ kiểm tra (x, y) với một nhãn giả định trong tất cả các fold.

Nếu định nghĩa p-value riêng lẻ:

$$p_k^y := \frac{|\{i \in S_k \mid \alpha_{i,k} \leq \alpha_k^y\}| + 1}{|S_k| + 1} \tag{3.34}$$

với mỗi fold, có thể thấy rằng p^y về bản chất là trung bình của p_k^y. Đặc biệt, nếu mỗi fold có cùng kích thước $|S_1| = \cdots = |S_K|$, ta được một phép tính đơn giản:

$$p^y = \bar{p}^y + \frac{K - 1}{l + 1}(\bar{p}^y - 1) \approx \bar{p}^y \tag{3.35}$$

với $\bar{p}^y := \frac{1}{K} \sum_{k=1}^{K} p_k^y$ là trung bình cộng của p_k^y và $\approx$ với giả sử $K \ll l$.

Ưu điểm lớn nhất của CCP là tính ổn định của giá trị tin cậy của nó: độ lệch chuẩn của độ tin cậy trung bình ít hơn nhiều so với ICP và CCP cũng mang lại độ tin cậy cao hơn.

3.11 Jackknife

Phương pháp jackknife tiêu chuẩn dựa trên việc xây dựng một tập hợp các mô hình theo cơ chế bỏ đi một (leave-one-out). Việc ước lượng các khoảng dự đoán được thực hiện trong ba bước chính:

- Đối với mỗi mẫu $i = 1, \ldots, n$ trong tập huấn luyện, sử dụng hàm hồi quy $\hat{\mu}_{-i}$ trên toàn bộ tập huấn luyện với mẫu thứ i^{th} bị loại bỏ, dẫn đến n mô hình bị bỏ đi một (leave-one-out model).

- Điểm hợp lý bỏ đi một (leave-one-out conformity score) tương ứng được tính toán cho mỗi mẫu thứ i^{th} là $|Y_i - \hat{\mu}_{-i}(X_i)|$.
- Khớp hàm hồi quy $\hat{\mu}$ trên toàn bộ tập huấn luyện và tính khoảng dự đoán bằng cách sử dụng các điểm hợp lý bỏ đi một đã được tính toán ở trên.

$$\hat{\mu}(X_{n+1}) \pm ((1-\alpha) \text{ quantile của } |Y_1 - \hat{\mu}_{-1}(X_1)|, \ldots, |Y_n - \hat{\mu}_{-n}(X_n)|) \quad (3.36)$$

Khoảng tin cậy kết quả là như sau:

$$\hat{C}^{\text{jackknife}}_{n,\alpha}(X_{n+1}) = [\hat{q}^{-}_{n,\alpha}\{\hat{\mu}(X_{n+1}) - R^{LOO}_i\}, \hat{q}^{+}_{n,\alpha}\{\hat{\mu}(X_{n+1}) + R^{LOO}_i\}] \quad (3.37)$$

với

$$R^{LOO}_i = |Y_i - \hat{\mu}_{-i}(X_i)| \quad (3.38)$$

là điểm hợp lý bỏ đi một.

Phương pháp này tránh được vấn đề quá khớp (overfitting) nhưng có thể làm giảm chất lượng dự đoán khi $\hat{\mu}$ trở nên không ổn định (ví dụ điển hình là khi kích thước tập dữ liệu xấp xỉ số lượng đặc trưng).

3.12 Jackknife+

Khác với phương pháp jackknife tiêu chuẩn, ước lượng một khoảng dự đoán quanh dự đoán của mô hình được huấn luyện trên toàn bộ tập dữ liệu, phương pháp gọi là jackknife+ sử dụng mỗi dự đoán bỏ đi một (leave-one-out prediction) trên quan sát mới để xem xét sự biến thiên của hàm hồi quy. Khoảng tin cậy kết quả là như sau:

$$\hat{C}^{\text{jackknife+}}_{n,\alpha}(X_{n+1}) = [\hat{q}^{-}_{n,\alpha}\{\hat{\mu}_{-i}(X_{n+1}) - R^{LOO}_i\}, \hat{q}^{+}_{n,\alpha}\{\hat{\mu}_{-i}(X_{n+1}) + R^{LOO}_i\}] \quad (3.39)$$

Phương pháp này đảm bảo sự ổn định cao với một mức độ bao phủ (coverage level) là $1-2\alpha$ cho một mức độ bao phủ mục tiêu là $1-\alpha$, mà không có bất kỳ giả thiết nào trước đó về phân phối của dữ liệu (X, Y) hay về mô hình dự đoán.

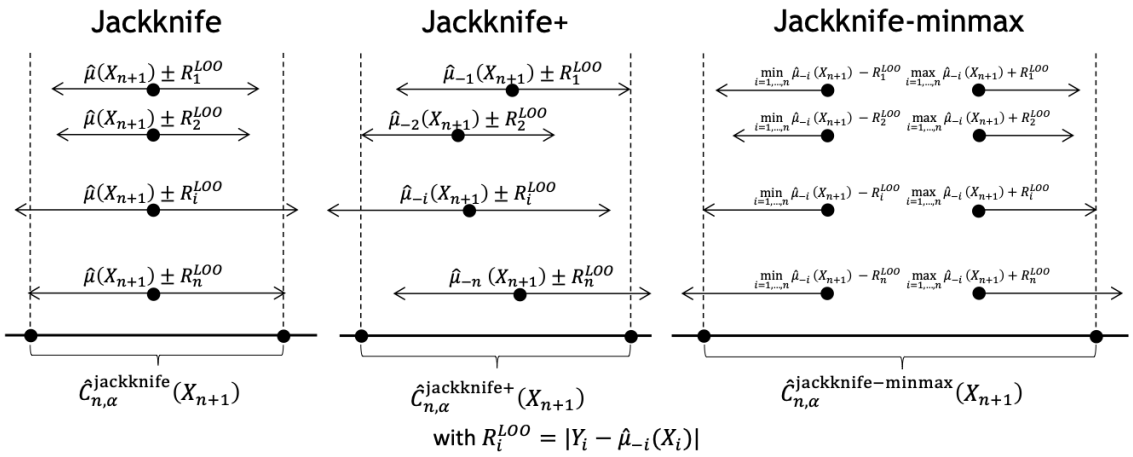
3.13 Jackknife-minmax

Phương pháp jackknife-minmax là một lựa chọn thận trọng hơn một chút vì nó sử dụng các giá trị tối thiểu và tối đa của các dự đoán bỏ đi một (leave-one-out prediction) để tính toán các khoảng dự đoán. Các khoảng dự đoán ước lượng là như sau:

$$\hat{C}^{\text{jackknife-mm}}_{n,\alpha}(X_{n+1}) = [\min \hat{\mu}_{-i}(X_{n+1}) - \hat{q}^{+}_{n,\alpha}\{R^{LOO}_I\}, \max \hat{\mu}_{-i}(X_{n+1}) + \hat{q}^{+}_{n,\alpha}\{R^{LOO}_I\}] \quad (3.40)$$

Phương pháp này đảm bảo mức độ bao phủ (coverage level) là $1-\alpha$ cho một mức độ bao phủ mục tiêu là $1-\alpha$.

Tuy nhiên, các phương pháp jackknife, jackknife+ và jackknife-minmax đều tốn nhiều tài nguyên tính toán vì chúng yêu cầu chạy một số lượng bằng số mẫu huấn luyện, điều này là không khả thi trong nhiều tình huống thường gặp trong khoa học dữ liệu.



Hình 3.1: Minh họa ba phương pháp jackknife và những điểm khác biệt chính

3.14 CV+

Để giảm thời gian tính toán, có thể áp dụng phương pháp kiểm định chéo (cross-validation) thay vì phương pháp bỏ đi một (leave-one-out), gọi là phương pháp CV+.

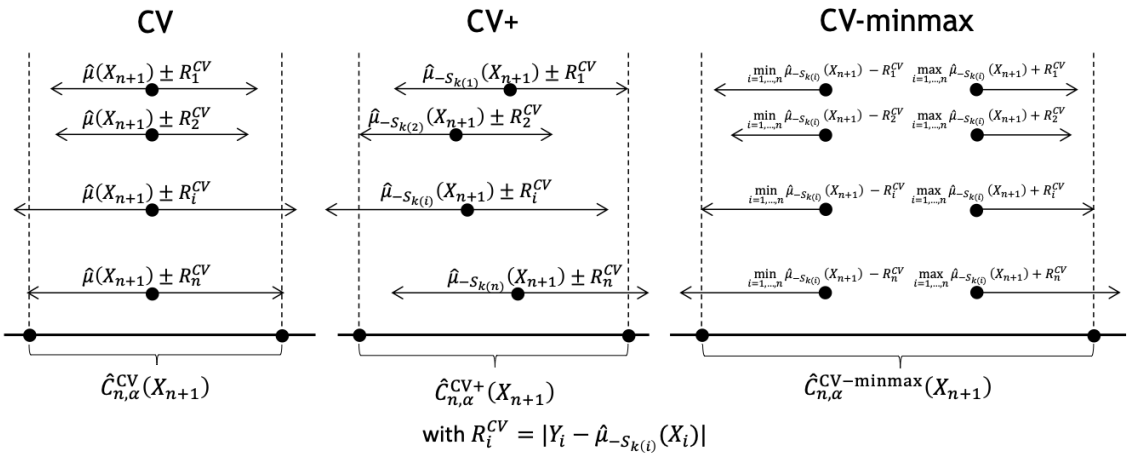
Tương tự như phương pháp jackknife+, việc ước lượng khoảng dự đoán với CV+ được thực hiện trong bốn bước chính:

- Chia tập huấn luyện thành K tập con $S_1, S_2, \dots, S_K$ không giao nhau có kích thước bằng nhau.
- K hàm hồi quy $\hat{\mu}_{-S_k}$ được khớp với tập huấn luyện (fold thứ k tương ứng bị loại bỏ).
- Điểm hợp lý bỏ đi fold (out-of-fold conformity score) tương ứng được tính cho mỗi quan sát thứ i là $|Y_i - \hat{\mu}_{-S_{k(i)}}(X_i)|$ với $k(i)$ là fold chứa i.
- Tương tự như jackknife+, các hàm hồi quy $\hat{\mu}_{-S_{k(i)}}(X_i)$ được sử dụng để ước lượng các khoảng dự đoán.

Giống như jackknife+, phương pháp này đảm bảo một mức độ bao phủ (coverage level) cao hơn là $1 - 2\alpha$ cho một mức độ bao phủ mục tiêu là $1 - \alpha$, mà không có bất kỳ giả thiết trước nào về phân phối của dữ liệu. Jackknife+ có thể được xem là một trường hợp đặc biệt của CV+ khi $K = n$. Trong thực tế, phương pháp này tạo ra các khoảng dự đoán rộng hơn một chút, nhưng đồng thời mang lại một sự cân nhắc hợp lý cho các tập dữ liệu lớn khi phương pháp Jackknife+ không khả thi.

3.15 CV và CV-minmax

Tương tự như CV+, phương pháp dựa vào các mô hình hồi quy bỏ đi fold (out-of-fold) được sử dụng để tính các khoảng dự đoán nhưng bằng cách dùng các phương trình được đưa ra trong section 3.11 Jackknife và section 3.13 Jackknife-minmax.



Hình 3.2: Minh họa ba phương pháp CV và những điểm khác biệt chính

3.16 Jackknife+-after-bootstrap

Để giảm thời gian tính toán và có dự đoán tốt hơn, có thể áp dụng một phương pháp bootstrap thay vì phương pháp bỏ đi một (leave-one-out), được gọi là phương pháp jackknife+-after-bootstrap. Theo cách hiểu thông thường, phương pháp này sử dụng phương pháp tập hợp để tính toán dự đoán được tổng hợp thứ i và thặng dư chỉ bằng cách lấy các tập con, trong đó quan sát thứ i không được sử dụng để huấn luyện với bộ ước lượng.

Tương tự như phương pháp CV+, việc ước lượng các khoảng dự đoán với jackknife+-after-bootstrap được thực hiện trong bốn bước chính:

- Tái lấy mẫu tập huấn luyện với việc thay thế (bootstrap) K lần, và do đó có được các bootstrap giao nhau $B_1, \ldots, B_K$ có kích thước bằng nhau.
- K hàm hồi quy $\hat{\mu}_{B_k}$ được khớp (fit) trên các bootstrap (B_k), và các dự đoán trên các tập bổ sung (B_k^c) được tính toán.
- Các dự đoán này được tổng hợp theo một hàm tổng hợp cụ thể agg, thường là mean hoặc median, và các điểm hợp lý $|Y_j - \mathrm{agg}(\hat{\mu}(B_{K(j)}(X_j)))|$ được tính cho mỗi X_j (với $K(j)$ là các bootstrap không chứa X_j).
- Các tập $\{\mathrm{agg}(\hat{\mu}_{\mathrm{K(j)}}(\mathrm{X_i})) + \mathrm{r_j}\}$ (trong đó j là chỉ số của tập huấn luyện) được sử dụng để ước lượng các khoảng dự đoán.

Như jackknife+, phương pháp này đảm bảo mức độ bao phủ (coverage level) cao hơn là $1 - 2\alpha$ cho một mức độ bao phủ mục tiêu là $1 - \alpha$, mà không có bất kỳ giả định nào trước đó về phân phối của dữ liệu. Trong thực tế, phương pháp này dẫn đến các khoảng dự đoán rộng hơn (khi sự không chắc chắn cao hơn) so với CV+ bởi vì sự phân tán dự đoán của các mô hình là cao hơn trong trường hợp này.

3.17 Conformalized quantile regression (CQR)

Phương pháp conformalized quantile cung cấp các khoảng dự đoán có chiều rộng tốt hơn với dữ liệu heteroscedastic. Sử dụng bộ hồi quy quantile với các giá trị quantile khác nhau để ước lượng ranh giới dự đoán, và thặng dư của các phương pháp này được sử dụng để tạo giá trị bảo đảm bao phủ.

$$\hat{C}_{n,\alpha}^{\mathrm{CQR}}(X_{n+1}) = [\hat{q}_{\alpha_{lo}}(X_{n+1}) - Q_{1-\alpha}(E_{low}, \mathcal{I}_2), \hat{q}_{\alpha_{hi}}(X_{n+1}) + Q_{1-\alpha}(E_{high}, \mathcal{I}_2)] \quad (3.41)$$

Trong đó, empirical quantile thứ $Q_{1-\alpha}(E, \mathcal{I}_2) := (1-\alpha)(1+1/|\mathcal{I}_2|)$ của $E_i : i \in \mathcal{I}_2$ và $\mathcal{I}_2$ là thặng dư của bộ ước lượng được khớp trên tập calibration. Lưu ý rằng trong phương pháp đối xứng, E_{low} và E_{high} là bằng nhau. Phương pháp này cung cấp một đảm bảo lý thuyết cho mức độ phủ mục tiêu là $1-\alpha$.

3.18 Ensemble batch prediction intervals (EnbPI)

Việc cài đặt của phương pháp tương tự như jackknife+-after-bootstrap. Các bộ ước lượng leave-one-out (LOO) được xấp xỉ nhờ một vài bootstrap. Tuy nhiên, khoảng tin cậy giống như phương pháp jackknife.

$$\hat{C}^{\text{EnbPI}}_{n,\alpha}(X_{n+1}) = [\hat{\mu}_{agg}(X_{n+1}) + \hat{q}_{n,\beta}\{R^{\text{LOO}}_i\}, \hat{\mu}_{agg}(X_{n+1}) + \hat{q}_{n,(1-\alpha+\beta)}\{R^{\text{LOO}}_i\}] \quad (3.42)$$

trong đó, $\hat{\mu}_{agg}(X_{n+1})$ là sự kết hợp của các dự đoán của bộ ước lượng LOO (trung bình hoặc trung vị), và $R^{\text{LOO}}_i = |Y_i - \hat{\mu}_{-i}(X_i)|$ là thặng dư của ước lượng LOO $\hat{\mu}_{-i}$ tại X_i.

Thặng dư không còn được xem xét ở dạng giá trị tuyệt đối mà ở dạng giá trị tương đối và độ rộng của các khoảng tin cậy được tối thiểu hóa, đến một khoảng cách nhất định giữa mức độ phân vị và sự tối ưu hóa tham số β.

Hơn nữa, các thặng dư được cập nhật trong quá trình dự đoán, mỗi khi có dữ liệu mới. Vì vậy, sự suy giảm giá trị của các dự đoán hoặc sự gia tăng mức độ nhiễu có thể được tính đến một cách linh hoạt.

Cuối cùng, độ đảm bảo bao phủ (coverage guarantee) không còn tuyệt đối mà tiệm cận theo hai giả thuyết:

1. Các lỗi là độc lập và phân phối giống nhau (i.i.d) ngắn hạn.
2. Chất lượng của ước lượng: tồn tại một chuỗi số thực $(\delta_T)_{T>0}$ hội tụ về không sao cho

$$\frac{1}{T}\sum_{1}^{T}(\hat{\mu}_{-t}(x_t) - \mu(x_t))^2 < \delta_T^2 \quad (3.43)$$

Mức độ bao phủ (coverage level) phụ thuộc vào kích thước của tập huấn luyện và $(\delta_T)_{T>0}$. Lưu ý: tập huấn luyện càng lớn, độ đảm bảo bao phủ càng tốt cho điểm theo sau tập huấn luyện. Tuy nhiên, nếu thặng dư được cập nhật dần dần, nhưng mô hình không được làm mới, thì tập huấn luyện càng lớn, quá trình cập nhật của thặng dư sẽ càng chậm. Do đó, cần có một sự thỏa hiệp cần phải được thực hiện về số lượng mẫu huấn luyện để phù hợp với mô hình và cập nhật các khoảng dự đoán.

Method	**Theoretical coverage**	**Typical coverage**	**Training cost**	**Evaluation cost**
Jackknife	No guarantee	$\simeq 1-\alpha$ or $< 1-\alpha$ if $\hat{\mu}$ unstable	n	n_{test}
Jackknife+	$\geq 1-2\alpha$	$\simeq 1-\alpha$	n	$n \times n_{test}$
Jackknife-minmax	$\geq 1-\alpha$	$> 1-\alpha$	n	$n \times n_{test}$
CV	No guarantee	$\simeq 1-\alpha$ or $< 1-\alpha$ if $\hat{\mu}$ unstable	K	n_{test}
CV+	$\geq 1-2\alpha$	$\gtrsim 1-\alpha$	K	$K \times n_{test}$
CV-minmax	$\geq 1-\alpha$	$> 1-\alpha$	K	$K \times n_{test}$
Jackknife-aB+	$\geq 1-2\alpha$	$\gtrsim 1-\alpha$	K	$K \times n_{test}$
CQR	$\geq 1-\alpha$	$\gtrsim 1-\alpha$	3	$3 \times n_{test}$
EnbPI	$\geq 1-\alpha$ (asymptotic)	$\gtrsim 1-\alpha$	K	$K \times n_{test}$

Bảng 3.1: Tóm tắt một số phương pháp được đề cập ở trên; n, n_{test} và K tương ứng là số mẫu huấn luyện, số mẫu kiểm tra và số fold cho kiểm định chéo

Tài liệu tham khảo

[1] Anastasios Angelopoulos, Stephen Bates, Jitendra Malik, and Michael I. Jordan. Uncertainty sets for image classifiers using conformal prediction. *CoRR*, abs/2009.14193, 2020.

[2] Anastasios N Angelopoulos and Stephen Bates. A gentle introduction to conformal prediction and distribution-free uncertainty quantification. *arXiv preprint arXiv:2107.07511*, 2021.

[3] Rina Foygel Barber, Emmanuel J. Candès, Aaditya Ramdas, and Ryan J. Tibshirani. Predictive inference with the jackknife+. *The Annals of Statistics*, 2019.

[4] Henrik Boström, Ulf Johansson, and Tuwe Löfström. Mondrian conformal predictive distributions. In Lars Carlsson, Zhiyuan Luo, Giovanni Cherubin, and Khuong An Nguyen, editors, *Proceedings of the Tenth Symposium on Conformal and Probabilistic Prediction and Applications*, volume 152 of *Proceedings of Machine Learning Research*, pages 24–38. PMLR, 08–10 Sep 2021.

[5] Matteo Fontana, Gianluca Zeni, and Simone Vantini. Conformal prediction: a unified review of theory and new challenges. *arXiv preprint arXiv:2005.07972*, 2020.

[6] Byol Kim, Chen Xu, and Rina Barber. Predictive inference is free with the jackknife+-after-bootstrap. *Advances in Neural Information Processing Systems*, 33:4138–4149, 2020.

[7] Jing Lei, Max G'Sell, Alessandro Rinaldo, Ryan J Tibshirani, and Larry Wasserman. Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113(523):1094–1111, 2018.

[8] Yaniv Romano, Evan Patterson, and Emmanuel Candes. Conformalized quantile regression. *Advances in neural information processing systems*, 32, 2019.

[9] Yaniv Romano, Matteo Sesia, and Emmanuel Candes. Classification with valid and adaptive coverage. *Advances in Neural Information Processing Systems*, 33:3581–3591, 2020.

[10] Mauricio Sadinle, Jing Lei, and Larry Wasserman. Least ambiguous set-valued classifiers with bounded error levels. *Journal of the American Statistical Association*, 114(525):223–234, June 2018.

[11] Glenn Shafer and Vladimir Vovk. A tutorial on conformal prediction. *Journal of Machine Learning Research*, 9(3), 2008.

[12] Aldo Solari and Vera Djordjilović. Multi split conformal prediction. *Statistics & Probability Letters*, 184:109395, 2022.

[13] Martim Sousa. Inductive conformal prediction: A straightforward introduction with examples in python *, 07 2022.

[14] Vladimir Vovk. Cross-conformal predictors. *arXiv preprint arXiv:1208.0806*, 2012.

[15] Vladimir Vovk, Alexander Gammerman, and Glenn Shafer. *Algorithmic Learning in a Random World*. Springer International Publishing, 2022.

[16] Chen Xu and Yao Xie. Conformal prediction interval for dynamic time-series. In *International Conference on Machine Learning*, pages 11559–11569. PMLR, 2021.