Quantification d'incertitudes et de robustesse via prédiction conforme

D. L. Razafindrakoto[1, 2]   A. Celisse[2]   J. Lacaille[1]

[1]Safran Aircraft Engines

[2]SAMM, Université Paris 1 Panthéon-Sorbonne

Introduction à la recherche en apprentissage statistique, 9 Janvier 2026

Parcours

Expérience en cours

- 2023-2026: Thèse CIFRE sur la “quantification d’incertitudes et de robustesse via prédiction conforme” chez Safran Aircraft Engines en collaboration avec le SAMM sous la direction de Jérôme Lacaille et Alain Celisse.
- 2023-2023: Stage de fin d’études sur la “quantification d’incertitudes” chez Safran Aircraft Engines sous l’encadrement de Jérôme Lacaille et Jean Coussirou.

Parcours

Expérience en cours

- 2023-2026: Thèse CIFRE sur la "quantification d'incertitudes et de robustesse via prédiction conforme" chez Safran Aircraft Engines en collaboration avec le SAMM sous la direction de Jérôme Lacaille et Alain Celisse.
- 2023-2023: Stage de fin d'études sur la "quantification d'incertitudes" chez Safran Aircraft Engines sous l'encadrement de Jérôme Lacaille et Jean Coussirou.

Formation

- 2020-2023: Diplôme d'ingénieur en Mathématiques Appliquées à SupGalilée (Université Sorbonne Paris Nord).
- 2022-2023: Master en Mathématiques des données à l'Université Sorbonne Paris Nord.
- 2022-2022: Mobilité ERASMUS à l'Université Technique de Munich (TUM).
- 2018-2021: Licence Mathématiques à l'Université Sorbonne Paris Nord.

Contexte métier

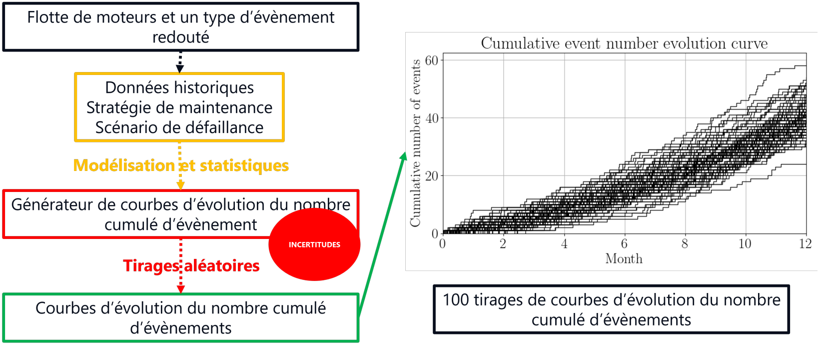


Figure: Comment les courbes sont générées?

Tube de confiance instant par instant

- Données: $\xi_1, \ldots, \xi_n$, ξ_{n+1} des variables aléatoires indépendantes indentiquement distribuées à valeur dans $\{0, \ldots, K\}^{T+1}$ où K et T sont des entiers plus grand que 1.

Tube de confiance instant par instant

- Données: $\xi_1, \ldots, \xi_n$, ξ_{n+1} des variables aléatoires indépendantes indentiquement distribuées à valeur dans $\{0, \ldots, K\}^{T+1}$ où K et T sont des entiers plus grand que 1.

Premier objectif

Pour tout niveau de contrôle de confiance $\alpha \in (0, 1)$, **calculer** un tube de confiance $\hat{B}_\alpha \subseteq \{0, \ldots, K\}^{T+1}$ tel que pour tout instant $t \in \{0, \ldots, T\}$,

$$\mathbb{P}\left[\xi_{n+1}^{(t)} \in \hat{B}_\alpha^{(t)}\right] \geq 1 - \alpha,$$

à partir des données observées ξ_1, ..., ξ_n.

Tube de confiance instant par instant

- Données: $\xi_1, \ldots, \xi_n$, ξ_{n+1} des variables aléatoires indépendantes indentiquement distribuées à valeur dans $\{0, \ldots, K\}^{T+1}$ où K et T sont des entiers plus grand que 1.

Premier objectif

Pour tout niveau de contrôle de confiance $\alpha \in (0, 1)$, **calculer** un tube de confiance $\hat{B}_\alpha \subseteq \{0, \ldots, K\}^{T+1}$ tel que pour tout instant $t \in \{0, \ldots, T\}$,

$$\mathbb{P}\left[\xi_{n+1}^{(t)} \in \hat{B}_\alpha^{(t)}\right] \geq 1 - \alpha,$$

à partir des données observées ξ_1, ..., ξ_n.

Solution

Prédiction conforme: Vovk, Gammerman, and Shafer (2005).

Méthodes et résultats

Le **seuil est $1/n_{train}$ plus bas** est que le seuil choisi pour que la **somme des probabilités en vert** dépasse $(1-\alpha)\left(1+\frac{1}{n_{train}}\right)$.

Les intervalles sont asymptotiquement optimaux mais peuvent être discontinus.

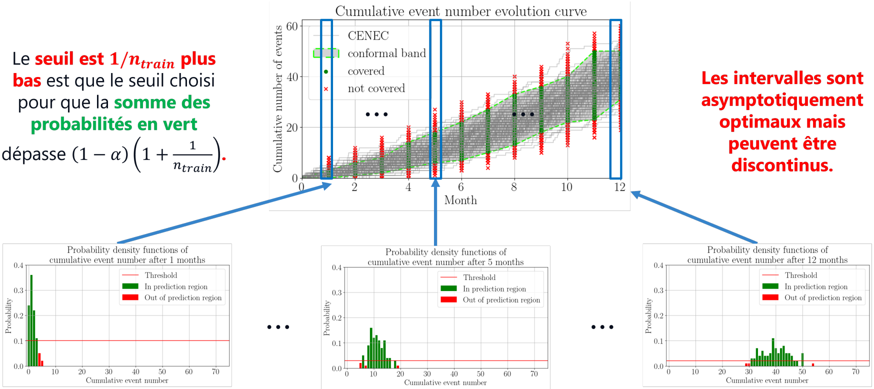
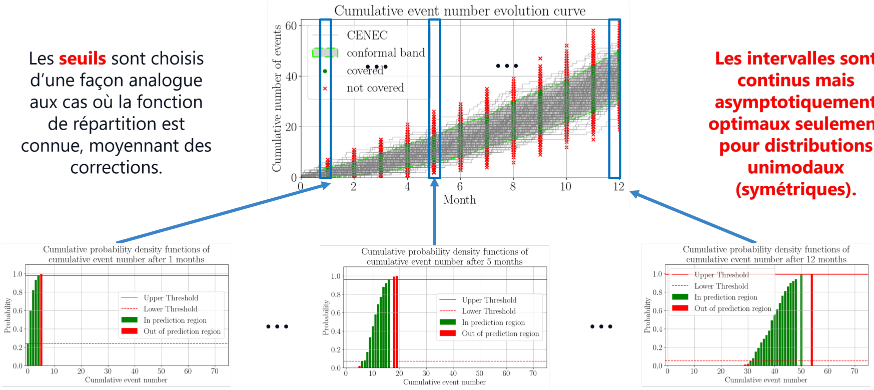


Figure: Tube de confiance avec un score de conformité basé sur la densité empirique.

Méthodes et résultats

Les **seuils** sont choisis d'une façon analogue aux cas où la fonction de répartition est connue, moyennant des corrections.



Les intervalles sont continus mais asymptotiquement optimaux seulement pour distributions unimodaux (symétriques).

Figure: Tube de confiance avec un score de conformité basé sur la fonction de répartition empirique.

Tube de confiance simultané

- Données: $\xi_1, \ldots, \xi_n$, ξ_{n+1} des variables aléatoires indépendantes indentiquement distribuées à valeur dans $\{0, \ldots, K\}^{T+1}$ où K et T sont des entiers plus grand que 1.

Tube de confiance simultané

- Données: $\xi_1, \ldots, \xi_n$, ξ_{n+1} des variables aléatoires indépendantes indentiquement distribuées à valeur dans $\{0, \ldots, K\}^{T+1}$ où K et T sont des entiers plus grand que 1.

Deuxième objectif

Pour tout niveau de contrôle de confiance $\alpha \in (0, 1)$, et pour tout niveau $\gamma \in (0, 1)$, **calculer** un tube de confiance $\hat{B}_{\alpha;\gamma} \subseteq \{0, \ldots, K\}^{T+1}$ tel que

$$\mathbb{P}\left[\frac{\left|\left\{ t \in \{0, \ldots, T\} : \xi_{n+1}^{(t)} \in \hat{B}_{\alpha;\gamma}^{(t)} \right\}\right|}{T+1} \geq 1 - \gamma \right] \geq 1 - \alpha,$$

à partir des données observées $\xi_1, \ldots, \xi_n$.

Tube de confiance simultané

- Données: $\xi_1, \ldots, \xi_n$, ξ_{n+1} des variables aléatoires indépendantes indentiquement distribuées à valeur dans $\{0, \ldots, K\}^{T+1}$ où K et T sont des entiers plus grand que 1.

Deuxième objectif

Pour tout niveau de contrôle de confiance $\alpha \in (0, 1)$, et pour tout niveau $\gamma \in (0, 1)$, **calculer** un tube de confiance $\hat{B}_{\alpha;\gamma} \subseteq \{0, \ldots, K\}^{T+1}$ tel que

$$\mathbb{P}\left[\frac{\left|\left\{ t \in \{0, \ldots, T\} : \xi_{n+1}^{(t)} \in \hat{B}_{\alpha;\gamma}^{(t)} \right\}\right|}{T+1} \geq 1 - \gamma \right] \geq 1 - \alpha,$$

à partir des données observées ξ_1, …, ξ_n.

Solution

Prédiction conforme: Papadopoulos (2008).

Méthode

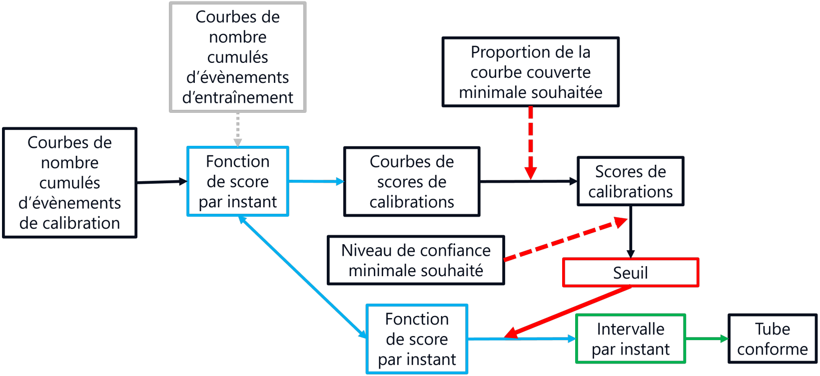


Figure: Schéma du calcul du tube de confiance simultané.

Contexte méthode

- Données: $(X_1, Y_1), \ldots, (X_n, Y_n), (X_{n+1}, Y_{n+1})$ variables aléatoires échangeables à valeur dans $\mathcal{X} \times \mathcal{Y}$ avec $\mathcal{X} \subseteq \mathbb{R}^p$ et $\mathcal{Y} \subseteq \mathbb{R}$.
- Tâche (régression): trouver un prédicteur $\hat{f} \in \mathcal{H}$ tel que $Y_{n+1} \approx \hat{f}(X_{n+1})$, où l'espace d'hypothèse $\mathcal{H}$ est un RKHS avec une fonction de noyau connue $k(\cdot, \cdot) : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$.

Contexte méthode

- Données: $(X_1, Y_1), \ldots, (X_n, Y_n), (X_{n+1}, Y_{n+1})$ variables aléatoires échangeables à valeur dans $\mathcal{X} \times \mathcal{Y}$ avec $\mathcal{X} \subseteq \mathbb{R}^p$ et $\mathcal{Y} \subseteq \mathbb{R}$.
- Tâche (régression): trouver un prédicteur $\hat{f} \in \mathcal{H}$ tel que $Y_{n+1} \approx \hat{f}(X_{n+1})$, où l'espace d'hypothèse $\mathcal{H}$ est un RKHS avec une fonction de noyau connue $k(\cdot, \cdot) : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$.

Objectif

Pour tout niveau de contrôle de confiance $\alpha \in (0, 1)$, **calculer** un intervalle de prédiction $\hat{C}_\alpha(X_{n+1}) \subset \mathcal{Y}$ tel que

$$\mathbb{P}\left[Y_{n+1} \in \hat{C}_\alpha(X_{n+1}) \right] \geq 1 - \alpha,$$

à partir des données observées $(X_1, Y_1), \ldots, (X_n, Y_n)$, de l'entrée test X_{n+1} et de l'espace d'hypothèse $\mathcal{H}$.

Contexte méthode

- Données: $(X_1, Y_1), \ldots, (X_n, Y_n), (X_{n+1}, Y_{n+1})$ variables aléatoires échangeables à valeur dans $\mathcal{X} \times \mathcal{Y}$ avec $\mathcal{X} \subseteq \mathbb{R}^p$ et $\mathcal{Y} \subseteq \mathbb{R}$.
- Tâche (régression): trouver un prédicteur $\hat{f} \in \mathcal{H}$ tel que $Y_{n+1} \approx \hat{f}(X_{n+1})$, où l'espace d'hypothèse $\mathcal{H}$ est un RKHS avec une fonction de noyau connue $k(\cdot, \cdot) : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$.

Objectif

Pour tout niveau de contrôle de confiance $\alpha \in (0, 1)$, **calculer** un intervalle de prédiction $\hat{C}_\alpha(X_{n+1}) \subset \mathcal{Y}$ tel que

$$\mathbb{P}\left[Y_{n+1} \in \hat{C}_\alpha(X_{n+1}) \right] \geq 1 - \alpha,$$

à partir des données observées $(X_1, Y_1), \ldots, (X_n, Y_n)$, de l'entrée test X_{n+1} et de l'espace d'hypothèse $\mathcal{H}$.

Solution

Prédiction conforme: Vovk, Gammerman, and Shafer (2005).

Prédiction "full" conforme

Scores de non-conformités

Pour toute valeur de sortie $y \in \mathcal{Y}$ et pour tout $i \in \{1, \ldots, n+1\}$, étant donné le jeu de données D^y contenant les données observées $(X_1, Y_1), \ldots, (X_n, Y_n)$, et la donnée de test augmentée (X_{n+1}, y),

$$
\begin{aligned}
S_{D^y}(X_i, Y_i) &:= s\left(Y_i, \hat{f}_{D^y}(X_i)\right), && \text{if } 1 \leq i \leq n, \\
S_{D^y}(X_i, y) &:= s\left(y, \hat{f}_{D^y}(X_i)\right), && \text{if } i = n+1,
\end{aligned}
$$

où $s(\cdot, \cdot) : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ est une function de non-conformité, et $\hat{f}_{D^y} \in \mathcal{H}$ est un prédicteur entraîné sur le jeu de donnée D^y.

Prédiction "full" conforme

Scores de non-conformités

Pour toute valeur de sortie $y \in \mathcal{Y}$ et pour tout $i \in \{1, \ldots, n+1\}$, étant donné le jeu de données D^y contenant les données observées $(X_1, Y_1), \ldots, (X_n, Y_n)$, et la donnée de test augmentée (X_{n+1}, y),

$$
\begin{aligned}
S_{D^y}(X_i, Y_i) &:= s\left(Y_i, \hat{f}_{D^y}(X_i)\right), & \text{if } 1 \leq i \leq n, \\
S_{D^y}(X_i, y) &:= s\left(y, \hat{f}_{D^y}(X_i)\right), & \text{if } i = n+1,
\end{aligned}
$$

où $s(\cdot,\cdot) : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ est une function de non-conformité, et $\hat{f}_{D^y} \in \mathcal{H}$ est un prédicteur entraîné sur le jeu de donnée D^y.

Intervalle de prédiction conforme (Vovk, Gammerman, and Shafer, 2005)

Pour tout niveau de contrôle de confiance $\alpha \in (0, 1)$,

$$
\hat{C}_{\alpha}^{\text{full}}(X_{n+1}) := \left\{ y \in \mathcal{Y} : \frac{1 + \sum_{i=1}^{n} \mathbb{1}\left\{S_{D^y}(X_i, Y_i) \geq S_{D^y}(X_{n+1}, y)\right\}}{n+1} > \alpha \right\}.
$$

Si les scores de non-conformités sont échangeables, et presque surement distincts alors

$$
1 - \alpha \leq \mathbb{P}\left[Y_{n+1} \in \hat{C}_{\alpha}^{\text{full}}(X_{n+1}) \right] \leq 1 - \alpha + \frac{1}{n+1}.
$$

Prédiction "full" conforme

Scores de non-conformités

Pour toute valeur de sortie $y \in \mathcal{Y}$ et pour tout $i \in \{1, \ldots, n+1\}$, étant donné le jeu de données D^y contenant les données observées $(X_1, Y_1), \ldots, (X_n, Y_n)$, et la donnée de test augmentée (X_{n+1}, y),

$$\begin{aligned} S_{D^y}(X_i, Y_i) &:= s\left(Y_i, \hat{f}_{D^y}(X_i)\right), & \text{if } 1 \leq i \leq n, \\ S_{D^y}(X_i, y) &:= s\left(y, \hat{f}_{D^y}(X_i)\right), & \text{if } i = n+1, \end{aligned}$$

où $s(\cdot, \cdot) : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ est une function de non-conformité, et $\hat{f}_{D^y} \in \mathcal{H}$ est un prédicteur entraîné sur le jeu de donnée D^y.

Intervalle de prédiction conforme (Vovk, Gammerman, and Shafer, 2005)

Pour tout niveau de contrôle de confiance $\alpha \in (0, 1)$,

$$\hat{C}_\alpha^{\text{full}}(X_{n+1}) := \left\{ y \in \mathcal{Y} : \frac{1 + \sum_{i=1}^n \mathbb{1}\left\{S_{D^y}(X_i, Y_i) \geq S_{D^y}(X_{n+1}, y)\right\}}{n+1} > \alpha \right\}.$$

Si les scores de non-conformités sont échangeables, et presque surement distincts alors

$$1 - \alpha \leq \mathbb{P}\left[Y_{n+1} \in \hat{C}_\alpha^{\text{full}}(X_{n+1}) \right] \leq 1 - \alpha + \frac{1}{n+1}.$$

- **Difficulté**: Trop couteux à calculer car il faudrait entraîner autant de prédicteurs que de valeurs de sortie.

Comment faire une approximation?

Scores de non-conformités approchés

Pour toute valeur de sortie $y \in \mathcal{Y}$ et tout $i \in \{1, \ldots, n+1\}$,

$$\tilde{S}^{\mathrm{lo}}_{D^y}(X_i, Y_i) \leq S_{D^y}(X_i, Y_i) \leq \tilde{S}^{\mathrm{up}}_{D^y}(X_i, Y_i) \qquad \text{if } 1 \leq i \leq n,$$

$$\tilde{S}^{\mathrm{lo}}_{D^y}(X_i, y) \leq S_{D^y}(X_i, y) \leq \tilde{S}^{\mathrm{up}}_{D^y}(X_i, y) \qquad \text{if } i = n+1.$$

Comment faire une approximation?

Scores de non-conformités approchés

Pour toute valeur de sortie $y \in \mathcal{Y}$ et tout $i \in \{1, \ldots, n+1\}$,

$$\begin{aligned} \tilde{S}^{\mathrm{lo}}_{D^y}(X_i, Y_i) \le S_{D^y}(X_i, Y_i) \le \tilde{S}^{\mathrm{up}}_{D^y}(X_i, Y_i) & \quad \text{if } 1 \le i \le n, \\ \tilde{S}^{\mathrm{lo}}_{D^y}(X_i, y) \le S_{D^y}(X_i, y) \le \tilde{S}^{\mathrm{up}}_{D^y}(X_i, y) & \quad \text{if } i = n+1. \end{aligned}$$

Intervalle de prédiction "Full" conforme approché (Ndiaye, 2022)

Pour tout niveau de contrôle de confiance $\alpha \in (0, 1)$,

$$\begin{aligned} \tilde{C}^{\mathrm{up}}_{\alpha}(X_{n+1}) &:= \left\{ y \in \mathcal{Y} : \frac{1 + \sum_{i=1}^{n} \mathbb{1}\left\{\tilde{S}^{\mathrm{up}}_{D^y}(X_i, Y_i) \ge \tilde{S}^{\mathrm{lo}}_{D^y}(X_{n+1}, y)\right\}}{n+1} > \alpha \right\} \supseteq \hat{C}^{\mathrm{full}}_{\alpha}(X_{n+1}) \\ &\supseteq \left\{ y \in \mathcal{Y} : \frac{1 + \sum_{i=1}^{n} \mathbb{1}\left\{\tilde{S}^{\mathrm{lo}}_{D^y}(X_i, Y_i) \ge \tilde{S}^{\mathrm{up}}_{D^y}(X_{n+1}, y)\right\}}{n+1} > \alpha \right\} =: \tilde{C}^{\mathrm{lo}}_{\alpha}(X_{n+1}). \end{aligned}$$

Comment faire une approximation?

Scores de non-conformités approchés

Pour toute valeur de sortie $y \in \mathcal{Y}$ et tout $i \in \{1, \ldots, n+1\}$,

$$\begin{aligned} \tilde{S}^{\mathrm{lo}}_{D^y}(X_i, Y_i) \leq S_{D^y}(X_i, Y_i) \leq \tilde{S}^{\mathrm{up}}_{D^y}(X_i, Y_i) & \qquad \text{if } 1 \leq i \leq n, \\ \tilde{S}^{\mathrm{lo}}_{D^y}(X_i, y) \leq S_{D^y}(X_i, y) \leq \tilde{S}^{\mathrm{up}}_{D^y}(X_i, y) & \qquad \text{if } i = n+1. \end{aligned}$$

Intervalle de prédiction "Full" conforme approché (Ndiaye, 2022)

Pour tout niveau de contrôle de confiance $\alpha \in (0, 1)$,

$$\begin{aligned} \tilde{C}^{\mathrm{up}}_{\alpha}(X_{n+1}) &:= \left\{ y \in \mathcal{Y} : \frac{1 + \sum_{i=1}^{n} \mathbb{1}\left\{ \tilde{S}^{\mathrm{up}}_{D^y}(X_i, Y_i) \geq \tilde{S}^{\mathrm{lo}}_{D^y}(X_{n+1}, y) \right\}}{n+1} > \alpha \right\} \supseteq \hat{C}^{\mathrm{full}}_{\alpha}(X_{n+1}) \\ &\supseteq \left\{ y \in \mathcal{Y} : \frac{1 + \sum_{i=1}^{n} \mathbb{1}\left\{ \tilde{S}^{\mathrm{lo}}_{D^y}(X_i, Y_i) \geq \tilde{S}^{\mathrm{up}}_{D^y}(X_{n+1}, y) \right\}}{n+1} > \alpha \right\} =: \tilde{C}^{\mathrm{lo}}_{\alpha}(X_{n+1}). \end{aligned}$$

Il en résulte que

$$\mathcal{L}\left(\tilde{C}^{\mathrm{up}}_{\alpha}(X_{n+1}) \setminus \hat{C}^{\mathrm{full}}_{\alpha}(X_{n+1}) \right) \leq \mathcal{L}\left(\tilde{C}^{\mathrm{up}}_{\alpha}(X_{n+1}) \setminus \tilde{C}^{\mathrm{lo}}_{\alpha}(X_{n+1}) \right).$$

Résultats

- Prédicteur: Pour un paramètre de régularisation $\lambda \in (0, +\infty)$, un jeu donnée D, et une fonction de perte $\ell(\cdot, \cdot) : \mathcal{Y} \times \mathcal{Y} \mapsto \mathbb{R}$,

$$\hat{f}_D \in \arg\min_{f \in \mathcal{H}} \frac{1}{|D|} \sum_{(x,y)} \ell(y, f(x)) + \lambda \|f\|^2_{\mathcal{H}} .$$

Résultats

- Prédicteur: Pour un paramètre de régularisation $\lambda \in (0, +\infty)$, un jeu donnée D, et une fonction de perte $\ell(\cdot,\cdot) : \mathcal{Y} \times \mathcal{Y} \mapsto \mathbb{R}$,

$$\hat{f}_D \in \arg\min_{f \in \mathcal{H}} \frac{1}{|D|} \sum_{(x,y)} \ell(y, f(x)) + \lambda \|f\|_{\mathcal{H}}^2 .$$

Approximation via stabilité algorithmique

Calcul d'un intervalle approchant $\tilde{C}_\alpha^{\mathrm{up},1}(X_{n+1})$ en remplaçant le prédicteur $\hat{f}_{D^y}$ pour chaque valeur de sortie y par un seul prédicteur $\hat{f}_{D^z}$, tel que pour tout $\alpha \in (0,1)$,

$$\mathcal{L}\left(\tilde{C}_\alpha^{\mathrm{up},1}(X_{n+1}) \setminus \hat{C}_\alpha^{\mathrm{full}}(X_{n+1})\right) \leq \mathcal{L}\left(\tilde{C}_\alpha^{\mathrm{up},1}(X_{n+1}) \setminus \tilde{C}_\alpha^{\mathrm{lo},1}(X_{n+1})\right) \leq O\left(\frac{1}{\lambda(n+1)}\right).$$

Résultats

- Prédicteur: Pour un paramètre de régularisation $\lambda \in (0, +\infty)$, un jeu donnée D, et une fonction de perte $\ell(\cdot, \cdot) : \mathcal{Y} \times \mathcal{Y} \mapsto \mathbb{R}$,

$$\hat{f}_D \in \arg\min_{f \in \mathcal{H}} \frac{1}{|D|} \sum_{(x,y)} \ell(y, f(x)) + \lambda \|f\|_{\mathcal{H}}^2 .$$

Approximation via stabilité algorithmique

Calcul d'un intervalle approchant $\tilde{C}_\alpha^{\mathrm{up},1}(X_{n+1})$ en remplaçant le prédicteur $\hat{f}_{D^y}$ pour chaque valeur de sortie y par un seul prédicteur $\hat{f}_{D^z}$, tel que pour tout $\alpha \in (0,1)$,

$$\mathcal{L}\left(\tilde{C}_\alpha^{\mathrm{up},1}(X_{n+1}) \setminus \hat{C}_\alpha^{\mathrm{full}}(X_{n+1})\right) \leq \mathcal{L}\left(\tilde{C}_\alpha^{\mathrm{up},1}(X_{n+1}) \setminus \tilde{C}_\alpha^{\mathrm{lo},1}(X_{n+1})\right) \leq O\left(\frac{1}{\lambda(n+1)}\right).$$

Approximation via fonction d'influence

Calcul d'un intervalle approchant $\tilde{C}_\alpha^{\mathrm{up},2}(X_{n+1})$ en remplaçant le prédicteur $\hat{f}_{D^y}$ pour chaque valeur de sortie y par une approximation $\tilde{f}_{D^y}$ ayant une expression connue, tel que pour tout $\alpha \in (0,1)$,

$$\mathcal{L}\left(\tilde{C}_\alpha^{\mathrm{up},2}(X_{n+1}) \setminus \hat{C}_\alpha^{\mathrm{full}}(X_{n+1})\right) \leq \mathcal{L}\left(\tilde{C}_\alpha^{\mathrm{up},2}(X_{n+1}) \setminus \tilde{C}_\alpha^{\mathrm{lo},2}(X_{n+1})\right) \leq O\left(\frac{1}{\lambda^3(n+1)^2}\right).$$

Conclusion

Récapitulatif

- Parcours.
- Calcul d'intervalle de prédiction "Full" conforme approché.
- Calcul de tubes de confiance.

Extension

- Calcul de région de prédiction pour la régression à sortie multivariée, et la classification multiclasse.
- Calcul de tubes de confiance sur la génération d'un modèle génératif de courbes.

Conclusion

Récapitulatif

- Parcours.
- Calcul d'intervalle de prédiction "Full" conforme approché.
- Calcul de tubes de confiance.

Extension

- Calcul de région de prédiction pour la régression à sortie multivariée, et la classification multiclasse.
- Calcul de tubes de confiance sur la génération d'un modèle génératif de courbes.

Merci beaucoup pour votre attention!

Références

[1] Eugene Ndiaye. "Stable conformal prediction sets". In: *International Conference on Machine Learning*. PMLR. 2022, pp. 16462–16479.

[2] Harris Papadopoulos. "Inductive conformal prediction: Theory and application to neural networks". In: *Tools in artificial intelligence*. Citeseer, 2008.

[3] Vladimir Vovk, Alexander Gammerman, and Glenn Shafer. *Algorithmic learning in a random world*. Vol. 29. Springer, 2005.