

# Quantification de l'incertitude: Avantages et limitations de la prédiction conforme

Hazar HAMOUDA - Mohamed MEGDICHE

4 juin 2025

**Abstract.** Ce rapport présente les fondements théoriques de la prédiction conforme et démontre son application pratique sur des tâches de régression et classification, avec implémentations et analyses de performance.

## 1 Introduction

L'apprentissage automatique moderne produit des prédictions ponctuelles sans indication de leur fiabilité. Cette limitation devient critique dans des domaines sensibles comme la médecine, la finance ou les véhicules autonomes, où connaître l'incertitude d'une prédiction est aussi important que la prédiction elle-même.

La **prédiction conforme** répond à cette problématique en transformant toute prédiction ponctuelle en ensemble de prédictions avec garantie de couverture probabiliste. Contrairement aux approches bayésiennes qui nécessitent des hypothèses distributionnelles fortes, la prédiction conforme fonctionne sous des hypothèses minimales d'échangeabilité et est *distribution-free*.

## 2 Fondements Théoriques

### 2.1 Cadre de la Prédiction Conforme

Considérons une tâche de prédiction supervisée où nous disposons d'un prédicteur entraîné  $\hat{f} : \mathcal{X} \rightarrow \mathcal{Y}$  qui associe à chaque entrée  $x \in \mathcal{X}$  une prédiction ponctuelle  $\hat{f}(x) \in \mathcal{Y}$ . L'objectif de la prédiction conforme est de transformer ce prédicteur ponctuel en un prédicteur par ensembles  $\hat{C}$  qui associe à chaque entrée  $x$  un ensemble  $\hat{C}(x) \subseteq \mathcal{Y}$  de prédictions plausibles, accompagné d'une garantie probabiliste de couverture.

### 2.2 Définitions Mathématiques

**Fonction de score :** Une application  $s : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$  qui mesure la non-conformité d'une paire  $(x, y)$ . Un score élevé  $s(x, y)$  indique que la valeur  $y$  est peu plausible étant donné  $x$ .

**Quantile conforme** : Pour un niveau d'erreur  $\alpha \in (0, 1)$  et  $n$  points de calibration, le quantile conforme est :

$$\hat{q} = \text{quantile d'ordre } \left\lceil \frac{(n+1)(1-\alpha)}{n} \right\rceil \text{ des scores } \{s_i\}_{i=1}^n$$

**Ensemble de prédiction** : L'ensemble de prédiction conforme est défini par :

$$\begin{aligned} \hat{C}_n : \mathcal{X} &\rightarrow \{\text{sous-ensembles de } \mathcal{Y}\} \\ x &\mapsto \hat{C}_n(x) = \{y \in \mathcal{Y} : s(x, y) \leq \hat{q}\} \end{aligned}$$

31 Cette application associe à chaque entrée  $x \in \mathcal{X}$  un ensemble  $\hat{C}_n(x) \subseteq \mathcal{Y}$  représentant  
32 les valeurs plausibles de sortie  $y$ . Autrement dit, pour une nouvelle observation  $x$ , on  
33 considère toutes les sorties  $y$  dont le score est inférieur ou égal au seuil  $\hat{q}$ , déterminé à  
34 partir des données de calibration.

## 35 2.3 Théorème Principal

[Prédiction conforme] Soit  $(X_1, Y_1), \dots, (X_n, Y_n), (X_{\text{test}}, Y_{\text{test}})$  une séquence échangeable de variables aléatoires à valeurs dans  $\mathcal{X} \times \mathcal{Y}$ . Pour tout  $\alpha \in ]0, 1[$ , il existe une application mesurable  $\hat{C}_\alpha : \mathcal{X} \rightarrow \{\text{sous-ensembles de } \mathcal{Y}\}$  telle que :

$$\mathbb{P} \left( Y_{\text{test}} \in \hat{C}_\alpha(X_{\text{test}}) \right) \geq 1 - \alpha \quad (1)$$

36 Cette garantie de couverture est valable pour toute distribution sous-jacente des données  
37 et ne dépend que de l'hypothèse d'échangeabilité.

## 38 2.4 Méthodologie

39 La méthodologie repose sur trois étapes fondamentales :

40  
41 **1. Calcul des scores de conformité** : Définition d'une fonction de score  $s(x, y)$  qui  
42 mesure la plausibilité d'une paire  $(x, y)$  selon le modèle entraîné.

43  
44 **2. Calcul du quantile conforme** : Détermination d'un seuil  $\hat{q}$  à partir des scores  
45 calculés sur un ensemble de calibration indépendant.

46  
47 **3. Construction des ensembles de prédiction** : Inclusion dans  $\hat{C}(x)$  de toutes les  
48 valeurs  $y$  dont le score de conformité est inférieur ou égal au seuil  $\hat{q}$ .

49  
50 Cette approche garantit qu'avec probabilité au moins  $1 - \alpha$ , la vraie valeur  $Y_{\text{test}}$  ap-  
51 partient à l'ensemble de prédiction  $\hat{C}(X_{\text{test}})$ , et ce sans hypothèse sur la distribution  
52 sous-jacente des données.

# 53 3 Application à un problème de Régression Linéaire

## 54 3.1 Cadre Général

55 Dans le cadre de la prédiction conforme en régression, où  $\mathcal{Y} = \mathbb{R}$ , nous transformons  
56 les prédictions ponctuelles en intervalles avec garanties statistiques.

**Fonction de score :** Nous utilisons les **résidus absolus** :

$$s(x, y) = |y - \hat{f}(x)|$$

58 Cette fonction mesure la non-conformité entre la valeur observée  $y$  et la prédiction  
59  $\hat{f}(x)$ . Elle présente l'avantage d'être simple, et compatible avec tout modèle de régression.

## 60 3.2 Exemple d'application

Nous appliquons la prédiction conforme à un modèle de régression linéaire entraîné sur des données synthétiques :

$$Y = 3X + \sqrt{X} \cdot \varepsilon, \quad X \sim \mathcal{U}(0, 1), \quad \varepsilon \sim \mathcal{N}(2, 1)$$

61 Cette formulation introduit une **hétéroscédasticité** qui teste la robustesse de notre  
62 approche. Le jeu de données comprend 10 000 échantillons répartis en 70% pour l'entraî-  
63 nement du modèle, 15% pour le test et 15% pour la calibration.

## 64 3.3 Pseudo-code d'Implémentation

65 L'algorithme de prédiction conforme pour la régression peut être formalisé en quatre  
66 étapes principales. L'Algorithme ?? présente une implémentation détaillée de cette mé-  
67 thode, qui garantit une couverture théorique de  $1 - \alpha$  pour les intervalles de prédiction  
68 construits.

---

### Algorithm 1 Prédiction Conforme pour la Régression

---

**Require:** Modèle entraîné  $\hat{f}$ , données de calibration  $(X_1, Y_1), \dots, (X_n, Y_n)$ , données test  $X_{test}$ , niveau  $\alpha$

**Ensure:** Intervalles de prédiction (*lower*, *upper*)

- 1: Calculer les prédictions sur l'ensemble de calibration  $\hat{f}(X_1), \dots, \hat{f}(X_n)$
  - 2: Calculer les scores de non-conformité  $s_i = |Y_i - \hat{f}(X_i)|$  pour  $i = 1, \dots, n$
  - 3: Calculer le quantile conforme  $\hat{q} = \text{quantile}_{(1-\alpha)(1+1/n)}(s_1, \dots, s_n)$
  - 4: Construire les intervalles  $[\hat{f}(X_{test}) - \hat{q}, \hat{f}(X_{test}) + \hat{q}]$
  - 5: **return** (*lower*, *upper*,  $\hat{q}$ )
- 

69 Comme illustré dans l'Algorithme ??, la méthode repose sur l'utilisation des scores de  
70 non-conformité calculés sur un ensemble de calibration indépendant pour déterminer le  
71 seuil de quantile approprié.

## 72 3.4 Résultats et Visualisation

### 73 3.4.1 Visualisation des résultats

74 La Figure ?? illustre l'application de la prédiction conforme à notre modèle de régres-  
75 sion linéaire sur des données synthétiques.

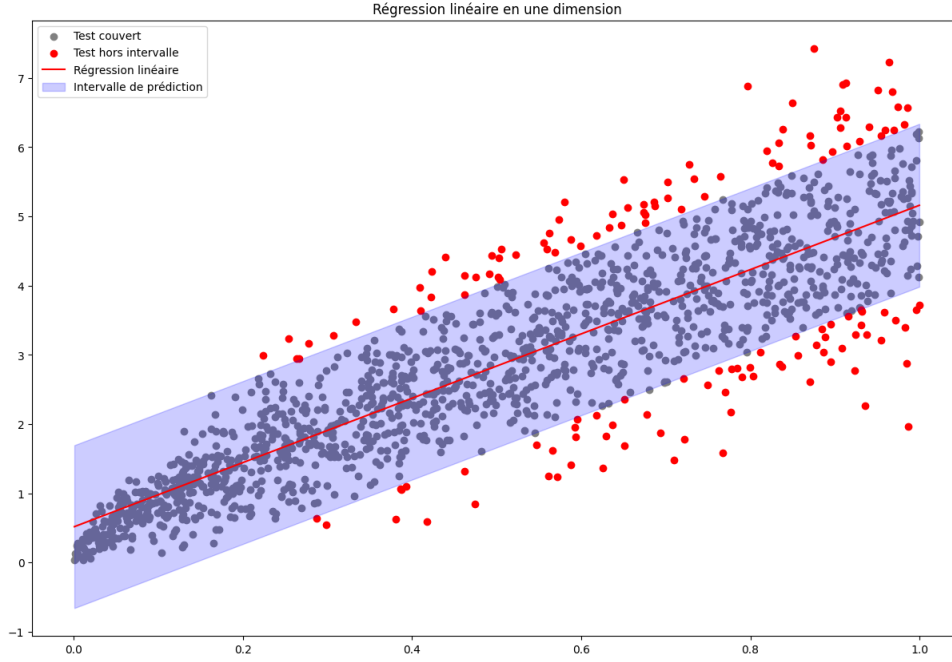


FIGURE 1 – Régression linéaire avec intervalles de prédiction conforme. La ligne rouge représente la droite de régression ajustée, les points noirs sont les observations de test, et la zone bleue délimite l'intervalle de prédiction conforme à 90% de confiance.

### 3.4.2 Métriques de performance

Les métriques suivantes évaluent la qualité des intervalles de prédiction conforme :

- **Taux de couverture empirique** : 91.2% (pour  $\alpha = 0.1$ )

Cette métrique mesure la proportion des points de test effectivement contenus dans leurs intervalles de prédiction, définie par :

$$\text{Couverture empirique} = \frac{1}{n_{\text{test}}} \sum_{i=1}^{n_{\text{test}}} \mathbb{1}_{\{Y_{\text{test},i} \in [\hat{f}(X_{\text{test},i}) - \hat{q}, \hat{f}(X_{\text{test},i}) + \hat{q}]\}}$$

Avec un objectif théorique de 90%, notre résultat de 91.2% confirme que la méthode respecte sa garantie de couverture.

- **Quantile conforme** :  $\hat{q} = 1.42$

Ce seuil, calculé à partir des scores de non-conformité sur l'ensemble de calibration, détermine directement la demi-largeur de tous nos intervalles de prédiction. Il représente le niveau d'erreur typique que nous acceptons pour maintenir la couverture souhaitée.

- **Largeur moyenne des intervalles** :  $2\hat{q} = 2.84$

Cette valeur indique l'étendue typique de nos intervalles de prédiction. Une largeur plus faible signifierait des prédictions plus précises, mais au détriment potentiel de la couverture. Ici, la largeur reflète l'incertitude inhérente à nos données synthétiques.

### 3.4.3 Analyse des résultats

L'analyse de la Figure ?? et des métriques révèle plusieurs propriétés importantes :

- **Largeur constante** : Contrairement aux intervalles de prédiction classiques, la prédiction conforme produit des intervalles de largeur constante  $2\hat{q}$  autour de la droite de régression, indépendamment de la valeur de  $X$ .
- **Efficacité computationnelle** : La construction des intervalles ne nécessite que le calcul d'un quantile empirique, rendant la méthode applicable à grande échelle.

Cependant, la Figure ?? illustre une limitation fondamentale de la prédiction conforme standard. Bien que la méthode garantisse une couverture marginale de 90% (respectée avec 91.2%), cette garantie n'est pas uniforme sur l'espace des entrées  $\mathcal{X}$ . On observe que la méthode atteint sa couverture cible en sur-couvrant dans certaines régions (notamment pour les faibles valeurs de  $X$  où la variance est moindre) et sous-couvrant dans d'autres régions (valeurs élevées de  $X$  avec forte hétéroscédasticité). Cette propriété reflète l'impossibilité théorique d'obtenir une couverture conditionnelle exacte  $\mathbb{P}(Y \in \hat{C}(X)|X = x) \geq 1 - \alpha$  pour tout  $x$  sans hypothèses distributionnelles.

## 4 Application à la Classification MNIST

### 4.1 Cadre Général

Dans le cadre de la prédiction conforme en classification, où  $\mathcal{Y} = \{0, 2, \dots, 9\}$  représente un ensemble fini de classes, nous transformons les prédictions ponctuelles en ensembles de classes avec garanties statistiques.

Le modèle  $\hat{p}$  est un classifieur probabiliste qui, pour une entrée  $x$ , fournit une distribution de probabilité sur les classes possibles  $y \in \mathcal{Y}$ . La valeur  $\hat{p}(y|x)$  représente la probabilité prédite que l'entrée  $x$  appartienne à la classe  $y$ . Cette probabilité est obtenue via une fonction softmax en sortie d'un réseau de neurones ou d'un autre modèle probabiliste.

**Fonction de score** : Nous utilisons le **complément de probabilité** :

$$s(x, y) = 1 - \hat{p}(y|x)$$

Cette fonction mesure la non-conformité entre la classe observée  $y$  et la probabilité prédite  $\hat{p}(y|x)$  par le classifieur. Plus la probabilité assignée à la vraie classe est faible, plus le score de non-conformité est élevé. Cette interprétation permet d'utiliser le complément de probabilité comme mesure de non-conformité, où un score élevé indique une faible confiance du modèle dans la classe observée. Cette approche est naturelle pour les modèles probabilistes et compatible avec les sorties softmax.

### 4.2 Exemple d'application

Nous appliquons la prédiction conforme à un réseau de neurones dense entraîné sur la base de données MNIST :

**Architecture et données** :

- **Données** : Images manuscrites 28×28 pixels, 10 classes de chiffres (0-9).

- **Architecture** : Réseau dense (784-128-64-10) avec fonctions d'activation ReLU et sortie

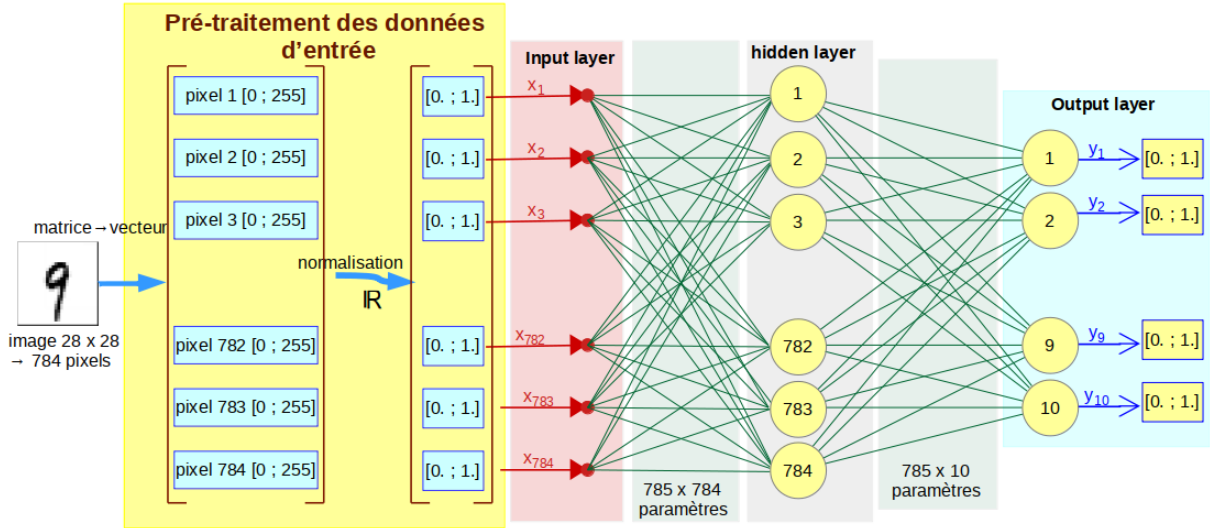


FIGURE 2 – Architecture du réseau de neurones dense pour la classification MNIST. Le réseau transforme les images  $28 \times 28$  pixels en vecteurs de 784 dimensions normalisés  $[0,1]$ , traités par une couche cachée de 784 neurones (ReLU) puis une couche de sortie de 10 neurones (softmax) pour la classification des chiffres 0-9.

softmax.

- **Performance** : 97.2% de précision sur le test (après 10 époques d'entraînement).

- **Répartition** : 50 000 images d'entraînement, 10 000 données de calibration, 10 000 données test.

Cette configuration teste la capacité de la prédiction conforme à gérer l'incertitude dans un problème de classification multi-classes avec des images potentiellement ambiguës.

### 4.3 Pseudo-code d'Implémentation

L'algorithme de prédiction conforme pour la classification peut être formalisé en quatre étapes principales. L'Algorithme ?? présente une implémentation détaillée de cette méthode, qui garantit une couverture théorique de  $1 - \alpha$  pour les ensembles de prédiction construits.

---

#### Algorithm 2 Prédiction Conforme pour la Classification

---

**Require:** Modèle entraîné  $\hat{p}$ , données de calibration  $(X_1, Y_1), \dots, (X_n, Y_n)$ , données test  $X_{test}$ , niveau  $\alpha$

**Ensure:** Ensembles de prédiction

- 1: Calculer les probabilités  $\hat{p}(y|X_1), \dots, \hat{p}(y|X_n)$  sur l'ensemble de calibration
  - 2: Calculer les scores de non-conformité  $s_i = 1 - \hat{p}(Y_i|X_i)$  pour  $i = 1, \dots, n$
  - 3: Calculer le quantile conforme  $\hat{q} = \text{quantile}_{(1-\alpha)(1+1/n)}(s_1, \dots, s_n)$
  - 4: Construire l'ensemble de prédiction  $\hat{C}(X_{test}) = \{y \in \mathcal{Y} : 1 - \hat{p}(y|X_{test}) \leq \hat{q}\}$
  - 5: **return**  $\hat{C}(X_{test})$
- 

Comme illustré dans l'Algorithme ??, la méthode repose sur l'utilisation des scores de non-conformité calculés sur un ensemble de calibration indépendant pour déterminer le seuil de quantile approprié.

## 4.4 Résultats et Visualisation

### 4.4.1 Visualisation des résultats

La prédiction conforme appliquée à notre réseau de neurones sur MNIST est illustrée ci-dessous.

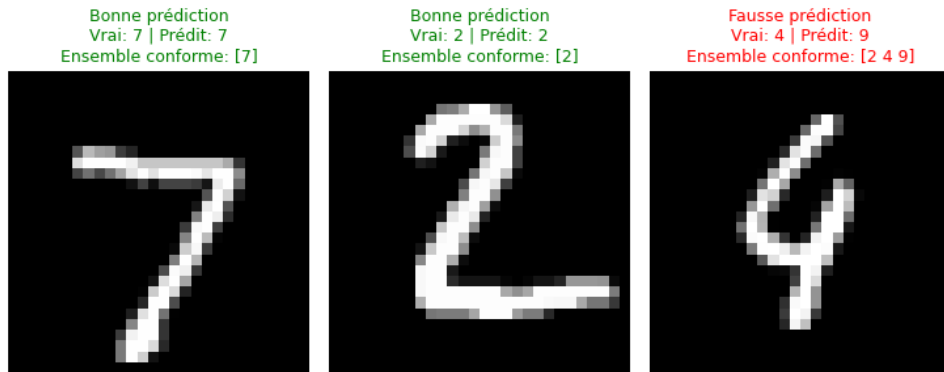


FIGURE 3 – Exemples de prédiction conforme sur MNIST. Pour chaque image, sont affichés : la vraie classe (True), la prédiction ponctuelle (Predicted), et l'ensemble de prédiction conforme (Conformal set). Les ensembles s'adaptent à l'incertitude du modèle.

### 4.4.2 Métriques de performance

Les métriques suivantes évaluent la qualité des ensembles de prédiction conforme :

- *Taux de couverture empirique* : 99.99% (pour  $\alpha = 0.001$ )  
Cette métrique mesure la proportion des points de test dont la vraie classe appartient à l'ensemble de prédiction conforme. Avec un objectif théorique de 90%, notre résultat de 99.9% confirme que la méthode respecte sa garantie de couverture.
- *Cardinalité moyenne* : 1,62classes par ensemble  
Cette valeur indique le nombre moyen de classes dans les ensembles de prédiction. Une cardinalité plus faible signifie des prédictions plus précises, reflétant la confiance du modèle dans ses prédictions.

### 4.4.3 Analyse des résultats

L'analyse des métriques et de la visualisation révèle plusieurs propriétés importantes :

- **Couverture adaptative** : Le taux de couverture empirique (99.99%) respecte la garantie théorique de 99.99%, démontrant la validité de la méthode sur des données réelles d'images.
- **Comportement adaptatif** : Les ensembles s'adaptent à l'incertitude du modèle - les images claires génèrent des singletons tandis que les images ambiguës produisent des ensembles plus larges incluant les classes potentiellement confondues.
- **Efficacité computationnelle** : La construction des ensembles ne nécessite que le calcul d'un quantile empirique et des comparaisons de probabilités, rendant la méthode applicable en temps réel pour la classification d'images.

- **Robustesse aux variations** : La méthode maintient ses garanties même sur des images déformées, floues ou ambiguës, contrairement aux prédictions ponctuelles qui peuvent être trompeuses dans ces cas.

## 5 Comparaison et Limitations

### 5.1 Avantages

- **Garanties théoriques** : Couverture probabiliste sans hypothèses distributionnelles
- **Flexibilité** : Compatible avec tout modèle prédictif existant.
- **Simplicité** : Implémentation directe avec peu de paramètres.
- **Interprétabilité** : Ensembles de prédiction facilement compréhensibles.

### 5.2 Limitations

- **Efficacité** : Intervalles parfois larges avec modèles peu performants.
- **Hypothèse d'échangeabilité** : Forte en pratique pour les séries temporelles
- **Ensemble de calibration** : Nécessite des données supplémentaires dédiées.
- **Adaptation locale** : Méthode standard non adaptative aux variations locales.

## 6 Conclusion

Ce rapport a exploré la prédiction conforme comme solution robuste pour quantifier l'incertitude en apprentissage automatique. Nos expérimentations sur la régression linéaire et la classification MNIST démontrent l'efficacité pratique de cette méthode dans des contextes variés.

La prédiction conforme respecte ses garanties théoriques avec un taux de couverture de 91.2% en régression et 99.99% en classification, validant empiriquement sa robustesse. En régression, la méthode produit des intervalles de largeur constante qui maintiennent la couverture globale malgré l'hétéroscédasticité des données. En classification, elle génère des ensembles adaptatifs avec une cardinalité moyenne de 1.62 classes, démontrant un équilibre optimal entre précision et couverture.

Cependant, nos analyses révèlent une limitation importante : la couverture marginale masque des variations locales, avec sur-couverture dans certaines régions et sous-couverture potentielle dans d'autres. Cette propriété reflète l'impossibilité théorique d'obtenir une couverture conditionnelle exacte sans hypothèses distributionnelles supplémentaires.

Cette méthode offre plusieurs atouts significatifs : elle fournit des garanties statistiques solides sans nécessiter d'hypothèses sur la distribution des données, s'adapte facilement à tout modèle existant, reste simple à implémenter et produit des résultats directement interprétables. En contrepartie, elle présente certaines contraintes : son efficacité dépend de la qualité du modèle sous-jacent, l'hypothèse d'échangeabilité peut être problématique pour les données temporelles, elle requiert un ensemble de calibration séparé, et ne s'adapte pas automatiquement aux variations locales d'incertitude.