Federated CP New Setting

Min

May 29, 2024

1 Idea1

For K agents each with $Z_k = \{(X_i^k, Y_i^k)\}_{i=1}^{n_k}$ sampled from distribution P^k, and point predictor f_1 based score S_i^k. If all $P^k = P$ and a new test point X, Y from P^1. For any trial data y and score S follow the procedure in "Conformal prediction with local weights: randomization enables robust guarantees"[2]:

- Find some kernel function $H(\cdot,\cdot)$, sample $\tilde{X}$ based on $H(X,\cdot)$.
- Calculate empirical function $\tilde{F} = \sum_{i,k} w_i^k \delta_{S_i^k} + w\delta_S$ with weight

$$w_i^k = \frac{H(X_i^k, \tilde{X})}{\sum_{i',k'} H(X_{i'}^{k'}, \tilde{X}) + H(X, \tilde{X})},\ w = \frac{H(X, \tilde{X})}{\sum_{i',k'} H(X_{i'}^{k'}, \tilde{X}) + H(X, \tilde{X})}.$$

- Conformal set is $C_\alpha(X) = \{S \le Q(1-\alpha, \tilde{F})\}$.

However for all k $P^k = P$ is not practical, potential covariate shift exists

1.1 Experiment1

Assume agent $k = 1, \cdots, K$ each has n samples $X_1^k, \cdots, X_n^k$ follow $N(\mu_k, 9)$. Sythesize $Y_i^k = (X_i^k)^2 + \epsilon$, where $\epsilon \sim N(0, (ep * |X|)^2)$, ep be some parameter. Under this problem only have covariate shift.

- X has different distribution for each agents

- $EY|X$ is same for all agents
- $Y - EY|X$ has same distribution for all agents

Covariate shift has little influence on this method as the method is localized. More on Fig 1. $ep = h = 0.5$

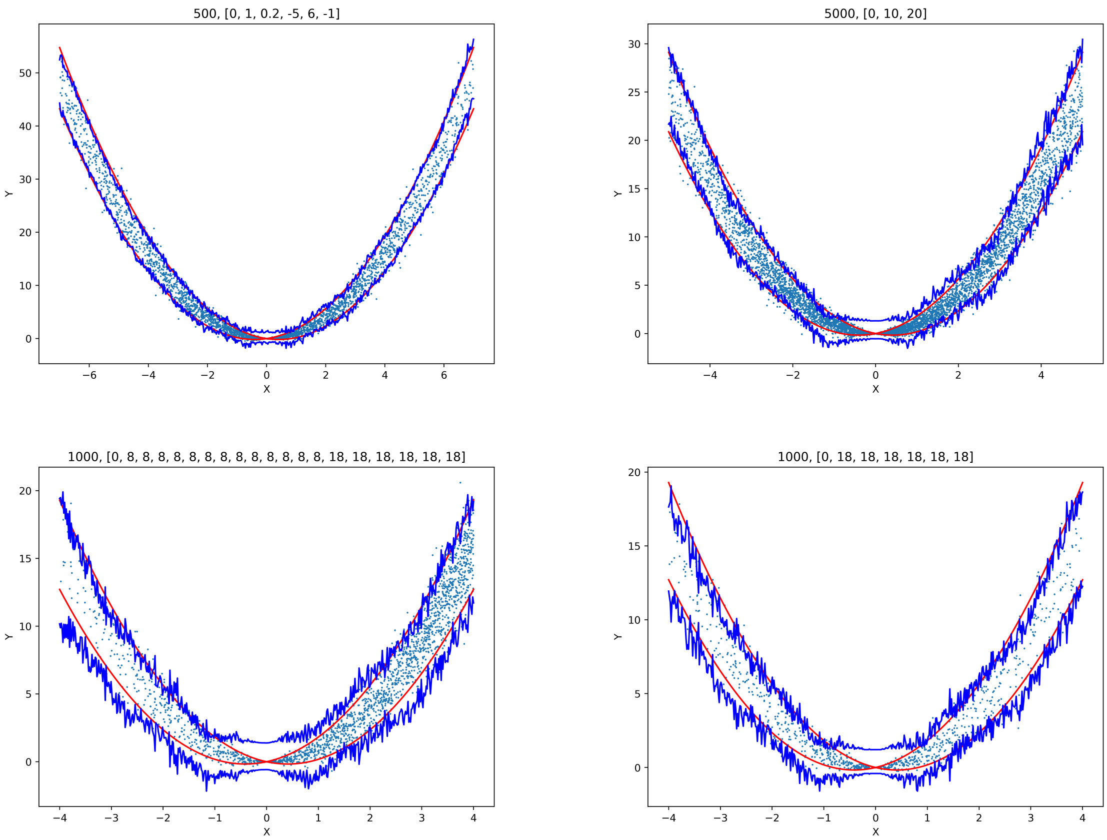

Figure 1: Fig1

Generate 1000 samples with location 0, and 1000 samples with location 0 and 20. The left shows the performance on only within samples with 0 location and 1000 samples is better than the right one using 2000 data samples. This comes from covariate shift, the scores from samples with location 20 influences origin behavior, from Fig 2

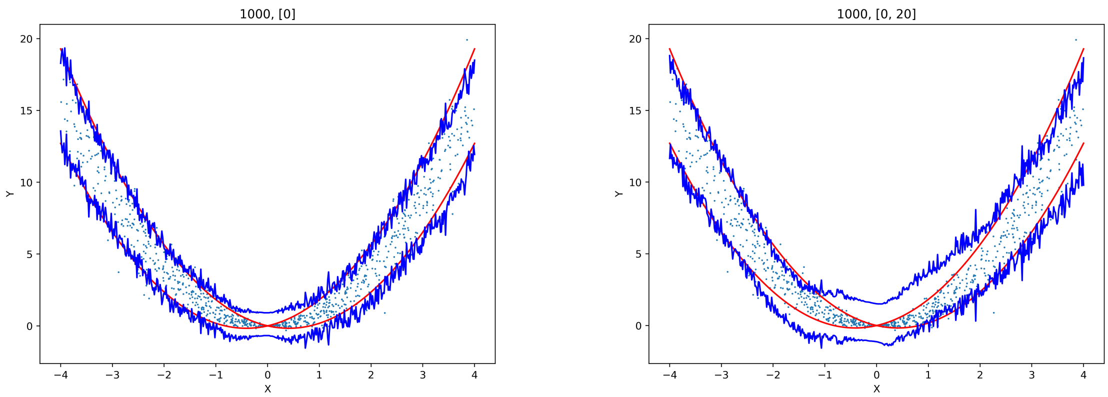

Figure 2: Fig2

1.2 Experiment2

Assume agent $k = 1, \cdots, K$ each has n samples $X_1^k, \cdots, X_n^k$ follow $N(\mu_k, 9)$. Sythesize $Y_i^k = (X_i^k)^2 + \epsilon$, where $\epsilon \sim N(0, (ep * |X| + ep * \theta_k)^2)$, ep be some parameter and θ_k generated for each agent randomly between 1 and 10.

- X has different distribution for each agents
- $EY|X$ is same for all agents
- $Y - EY|X$ has different distribution for different agents

The heterogeneity greatly influences the performance as this method treat data from different agents with same operation with respect to covariates. See in Fig 3. $ep = h = 0.5$.

To see how extra agent information influences performance, assume 4 agents with location 0, 5, 5, 10 and $\theta = (9.8, 2.1, 1.6, 6.3)$. Use only first agent data to estimate(left) and compare with the performance using all agents(right) in Fig 4.

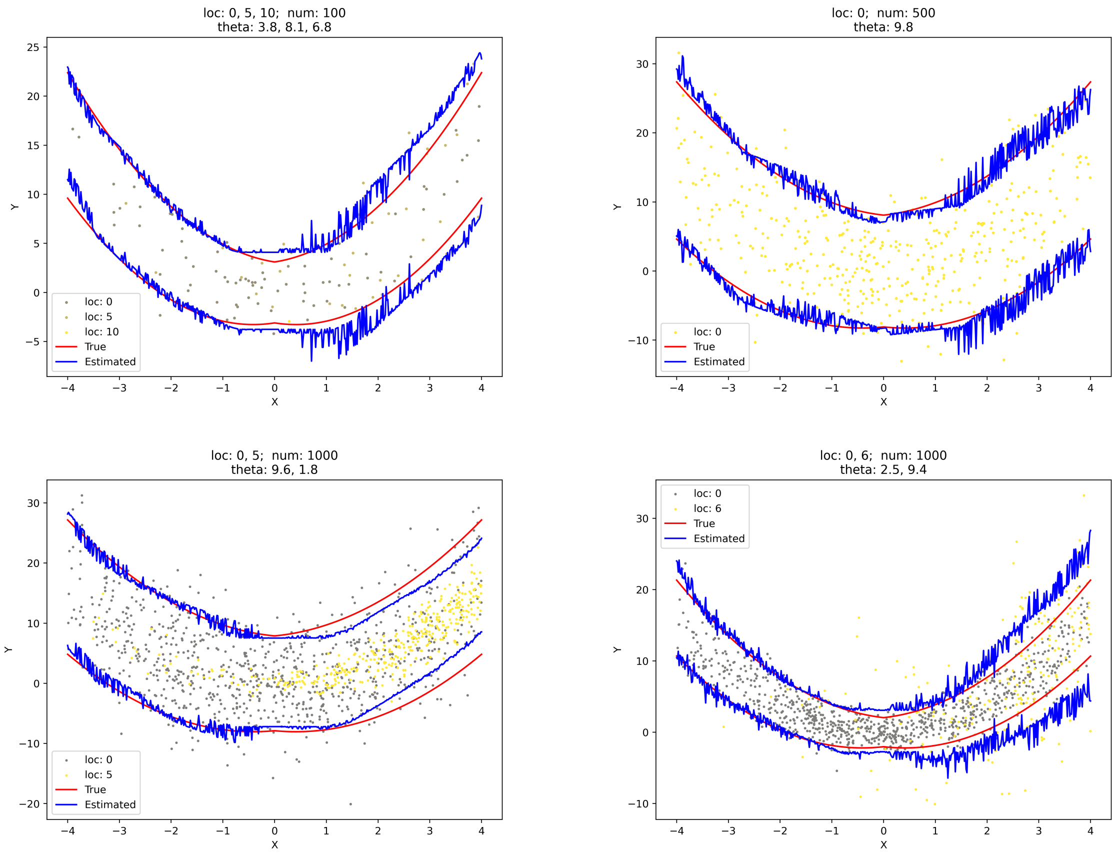

Figure 3: Fig3

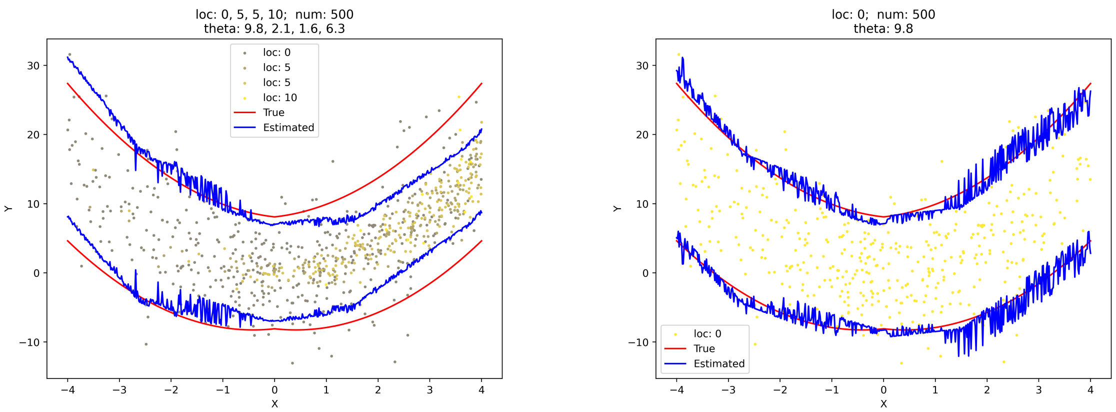

Figure 4: Fig4

2 Idea2

This idea is based on Guan[1].

For K agents each with $Z_k = \{Z_i^k = (X_i^k, Y_i^k)\}_{i=1}^{n_k}$ sampled from distribution P^k, and point predictor f based score S_i^k.

Given a new sample $X_{n_1+1}^1, Y_{n_1+1}^1$ from P^1, any trial data y and $Z_{n_1+1}^1 = (X_{n_1+1}^1, y)$. Now given some weight $p_{i,j}^{1,k} = \phi(Z_i^1, Z_j^k; Z_1, Z_k)$, where

$$\phi(z, z'; Z, Z') : \mathcal{Z} \times \mathcal{Z} \times \mathcal{Z}^n \times \mathcal{Z}^{n'} \to \mathbb{R},$$

is exchangeable within Z and Z' respectively. Construct

$$\hat{F}_i^1 = C_i \left[\sum_{k=1}^{K} \sum_{j=1}^{n_k} p_{i,j}^{1,k} \delta_{S_j^k} + p_{i,n_1+1}^{1,1} \delta_{S_{n_1+1}^1} \right],\ i = 1, \cdots, n_1 + 1.$$

Write $z_k = \{z_i^k = (x_i^k, y_i^k)\}_{i=1}^{n_k}$, corresponding score s_i^k, and $E_Z = \{Z_1 = z_1, \cdots, Z_K = z_K\}$. Write

$$\hat{F}_i^{1\prime} = C_i \left[\sum_{k=1}^{K} \sum_{j=1}^{n_k} p_{i,j}^{1,k} \delta_{s_j^k} + p_{i,n_1+1}^{1,1} \delta_{s_{n_1+1}^1} \right],\ i = 1, \cdots, n_1 + 1.$$

Now calculate

$$\begin{aligned} &\mathbb{P}\left(S_{n_1+1}^1 \le Q(\alpha', \hat{F}_{n_1+1}^1) \middle| E_Z \right) \\ Z_{n_1+1}^1 \text{ is r.v } =& \sum_{i=1}^{n_1+1} \mathbb{P}\left(Z_{n_1+1}^1 = z_i^1 \middle| E_Z \right) \mathbb{1}\left(S_{n_1+1}^1 \le Q(\alpha', \hat{F}_{n_1+1}^1) \middle| E_Z, S_{n_1+1}^1 = z_i^1 \right) \\ \text{Exchangeability of } Z_i^1 =& \frac{1}{n_1+1} \sum_{i=1}^{n_1+1} \mathbb{1}\left(S_{n_1+1}^1 \le Q(\alpha', \hat{F}_{n_1+1}^1) \middle| E_Z, Z_{n_1+1}^1 = z_i^1 \right). \end{aligned}$$

Note the core idea lies here is that $\hat{F}_i^1$ is concentrated around point X_i^1. When $Z_{n_1+1}^1 = Z_i^1$ and condition on E_Z, $\hat{F}_{n_1+1}^1 = \hat{F}_i^{1\prime}$, thus

$$\mathbb{P}\left(S_{n_1+1}^1 \le Q(\alpha', \hat{F}_{n_1+1}^1) \middle| E_Z \right) = \frac{1}{n_1+1} \sum_{i=1}^{n_1+1} \mathbb{1}\left(s_i^1 \le Q(\alpha', \hat{F}_i^{1\prime}) \middle| E_Z \right).$$

Find minimal α' makes $\frac{1}{n_1+1} \sum_{i=1}^{n_1+1} \mathbb{1}\left(s_i^1 \le Q(\alpha', \hat{F}_i^{1\prime}) \middle| E_Z \right) \ge 1 - \alpha$, then we have

$$\mathbb{P}\left(S_{n_1+1}^1 \le Q(\alpha', \hat{F}_{n_1+1}^1) \middle| E_Z \right) \ge 1 - \alpha.$$

However the previous process is very difficult to calculate, next is a practical implement. Simplify problem to n samples $X_1, \cdots, X_n; Y_1, \cdots, Y_n$ and score $S_1, \cdots, S_n$ with new test point X_{n+1}. Assume scores are ordered such that $S_1 \leq \cdots \leq S_n$ and S_{l_i} is the biggest S_j smaller than S_i. For example if $S_i > S_{i-1}$, $l_i = i - 1$, define $l_1 = 0$. Given kernel function calculated weight $p_{i,j}$ based on X_i, X_j. $p_{i,j} = p_{j,i}$ and $\sum_{j=1}^{n+1} p_{i,j} = 1$.

- Define θ_i be cumulative probability centered on X_i till l_i, which is $\theta_i = \sum_{j=1}^{l_i} p_{i,j}$.
- Define $\tilde{\theta}_i$ be cumulative probability centered on X_{n+1} till l_i, which is $\tilde{\theta}_i = \sum_{j=1}^{l_i} p_{n+1,j}$.
- Find three disjoint index sets

$$\begin{aligned} A_1 &= \left\{ i : \tilde{\theta}_i > \theta_i + p_{n+1,i} \right\} \\ A_2 &= \left\{ i : \theta_i \geq \tilde{\theta}_i \right\} \\ A_3 &= \left\{ i : \theta_i < \tilde{\theta}_i \leq \theta_i + p_{n+1,i} \right\} \end{aligned}$$

- Define following sets

$$\begin{aligned} B_1 &= \{\theta_i + p_{n+1,i} : i \in A_1\} \\ B_2 &= \{\theta_i : i \in A_2\} \\ B_3 &= \{l_i : i \in A_3\} \end{aligned}$$

- For $k = 1, \cdots, n+1$ do following separately:
  - Count n_1, n_2 be the number of items in B_1, B_2 that smaller than $\tilde{\theta}_k$, n_3 be the number of items in B_3 that smaller than $l(k)$.
  - Define $T(k) = (n_1 + n_2 + n_3)/(n+1)$.
- Find $k^* = \arg\max \{k : S(k) < \alpha\}$. Take $S^* = S_{\min(n,k^*)}$.
- Construct conformal set $C_\alpha(X_{n+1}) = \{y : S_{n+1} \leq S^*\}$.

References

[1] Leying Guan. Localized conformal prediction: A generalized inference framework for conformal prediction. *Biometrika*, 110(1):33–50, 2023.

[2] Rohan Hore and Rina Foygel Barber. Conformal prediction with local weights: randomization enables local guarantees. *arXiv preprint arXiv:2310.07850*, 2023.