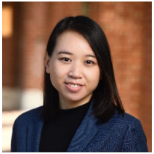
Efficient Approximate Predictive Inference Under Feedback Covariate Shift with Influence Functions

Drew Prinster, Suchi Saria, Anqi Liu

Johns Hopkins University

2023

Presentation Outline

Two Key Challenges in Conformal Prediction

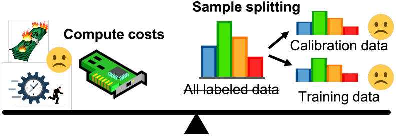
1. Resource constraints (compute & available data)

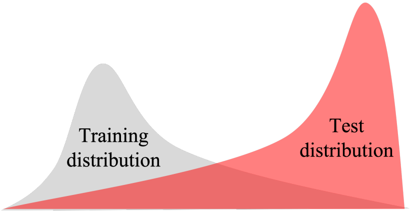
- *Computational budget*: e.g., extensive model retraining
- *Data-availability demands*: e.g., sample-splitting (which can harm model performance, especially in low data regime)

2. Data shifts

Real world data are often **not exchangeable**!

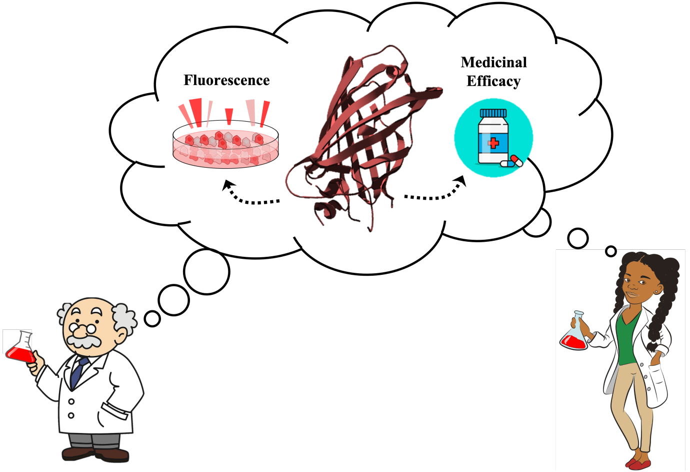
Common shifts between training & test data distributions can break standard conformal methods.





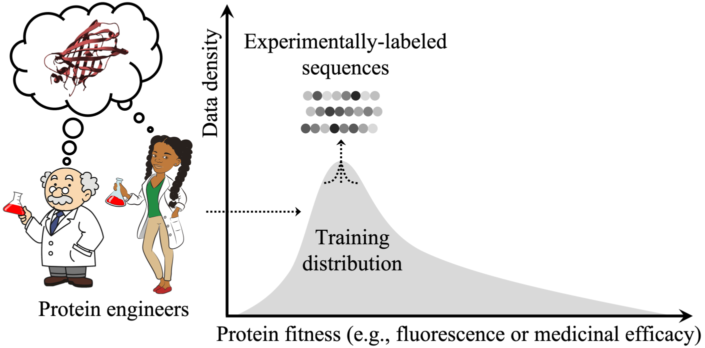
Our work (today and prior) is at the intersection of these challenges.

Biomolecular Design Setting

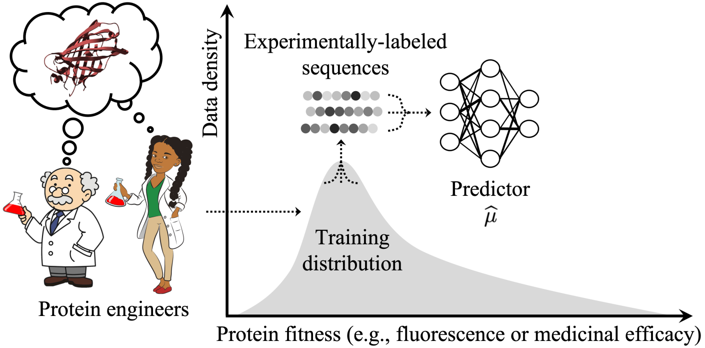


Detailed description in Fannjiang, Bates, Angelopoulos, Listgarten, and Jordan (2022)

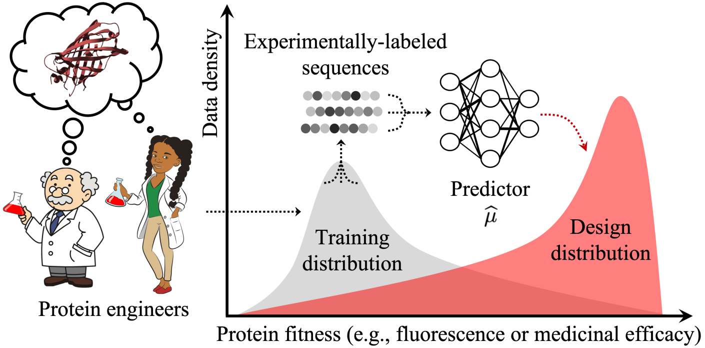
Biomolecular Design Setting



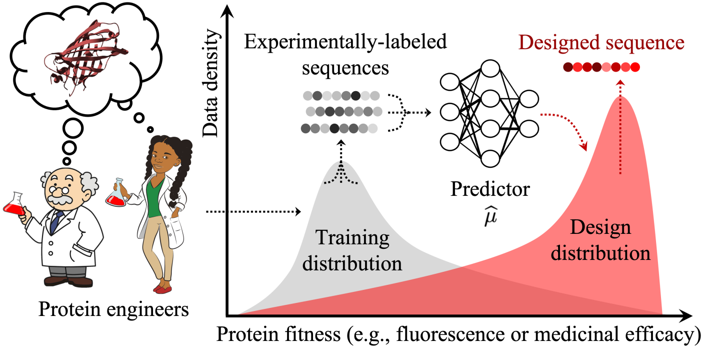
Biomolecular Design Setting



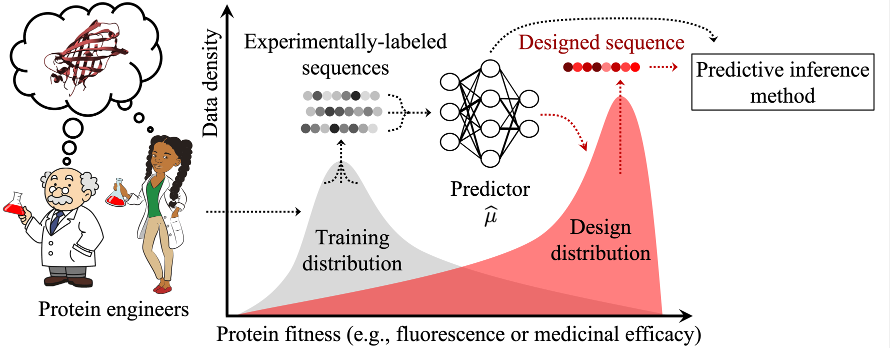
Biomolecular Design Setting



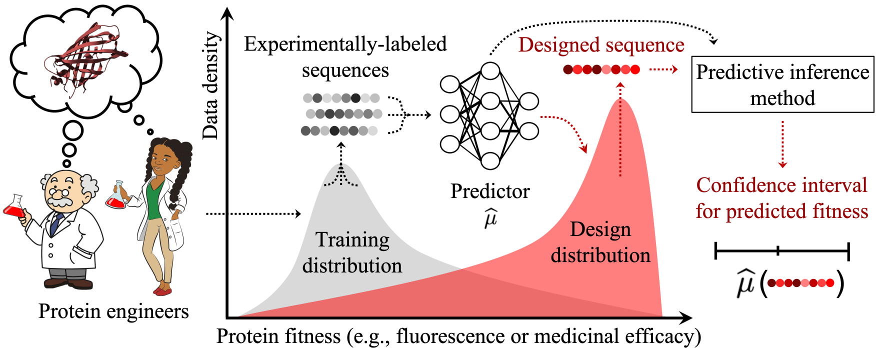
Biomolecular Design Setting



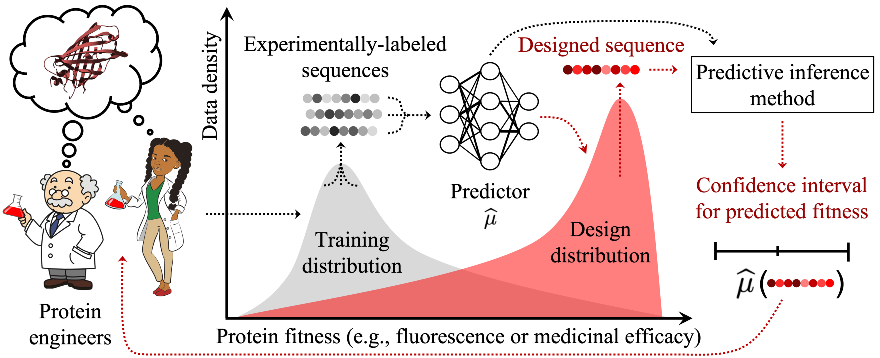
Biomolecular Design Setting



Biomolecular Design Setting



Biomolecular Design Setting

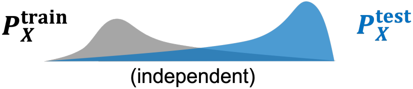


Background: Feedback Covariate Shift (FCS)

Standard conformal prediction "SCS":

$$Z_i = (X_i, Y_i) \overset{\text{i.i.d.}}{\sim} \boldsymbol{P_X^{\text{train}}} \times P_{Y|X}, \quad i = 1, \ldots, n$$

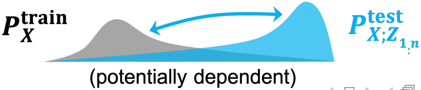
$$(\boldsymbol{X_{n+1}}, \boldsymbol{Y_{n+1}}) \sim \boldsymbol{P_X^{\text{test}}} \times P_{Y|X}$$



Feedback covariate shift "FCS"(One-shot biomolecular design is an instance; Fannjiang et al. (2022)):

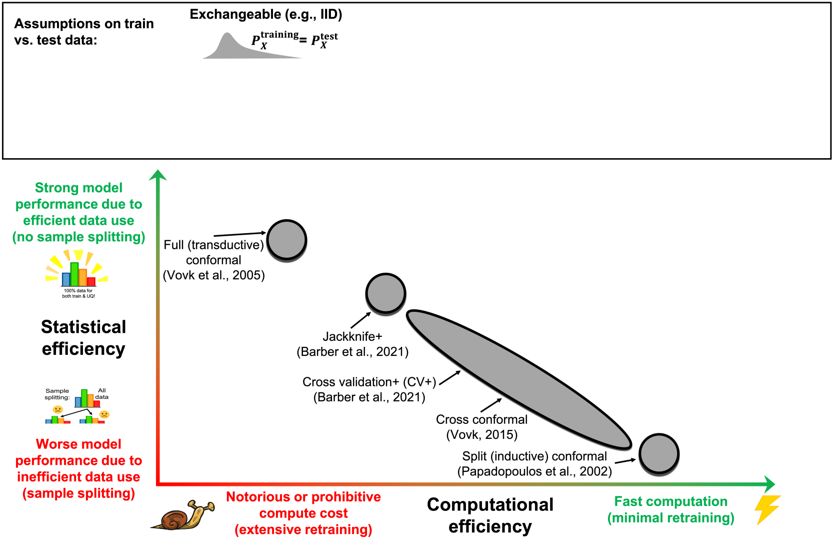
$$Z_i = (X_i, Y_i) \overset{\text{i.i.d.}}{\sim} \boldsymbol{P_X^{\text{train}}} \times P_{Y|X}, \quad i = 1, \ldots, n$$

$$(\boldsymbol{X_{n+1}}, \boldsymbol{Y_{n+1}}) \sim \boldsymbol{P_{X;Z_{1:n}}^{\text{test}}} \times P_{Y|X}$$

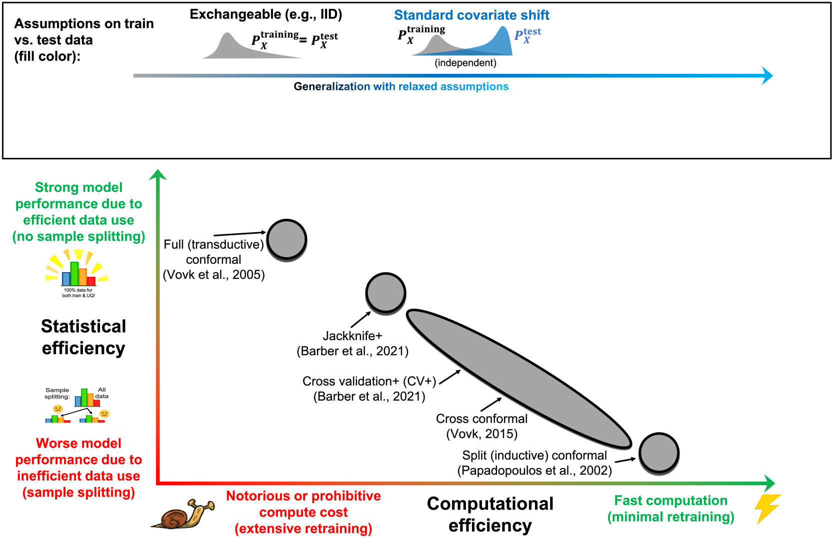


Presentation Outline

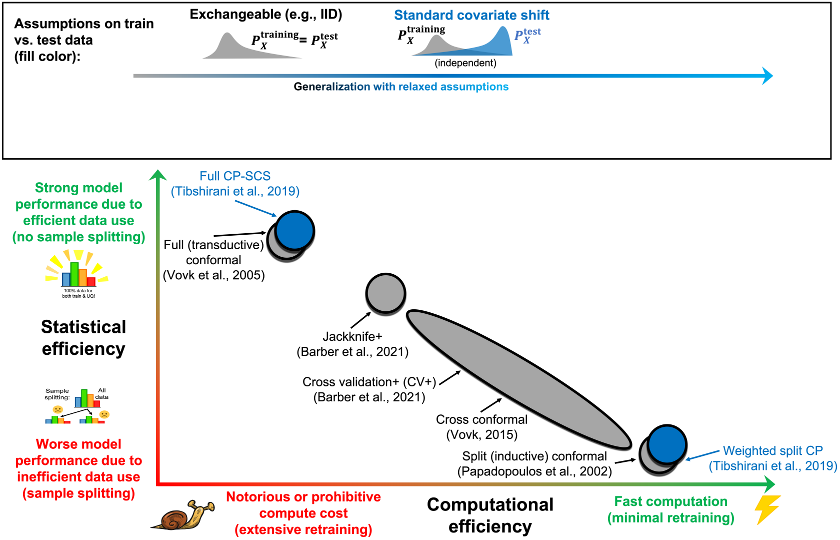
Related Work: Efficiency Tradeoffs



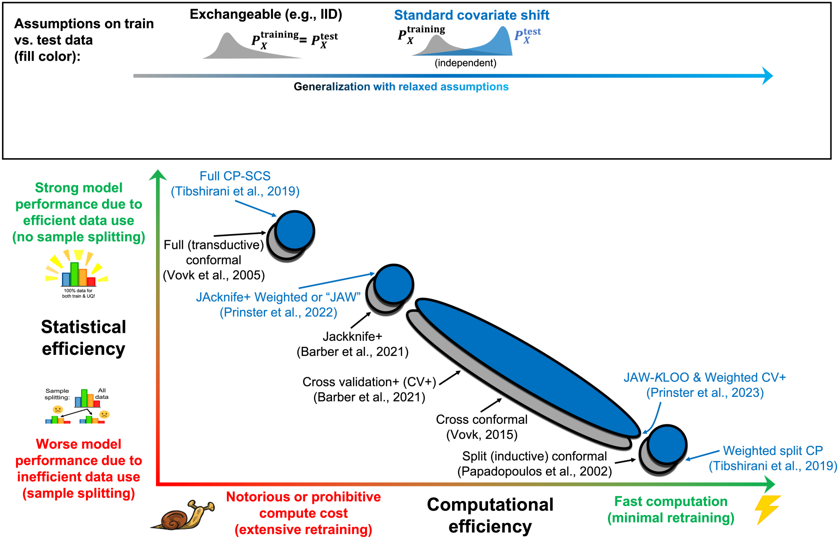
Related Work: Feedback Covariate Shift (FCS)



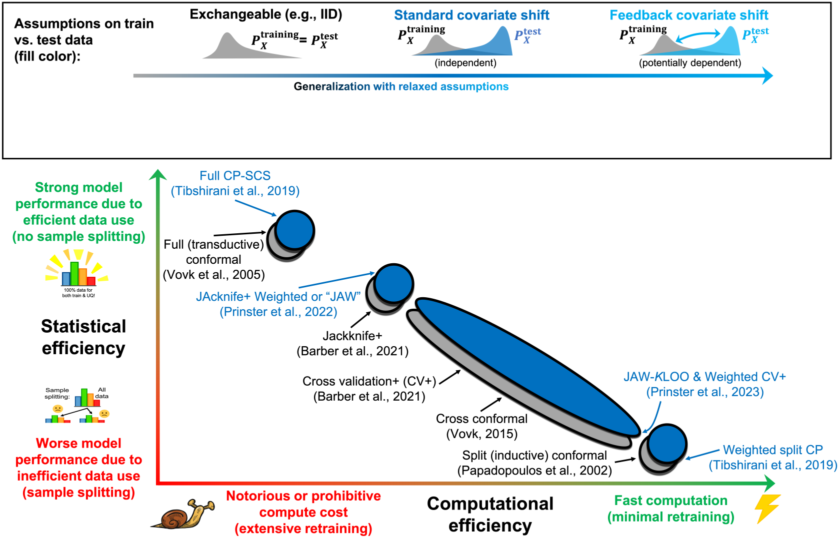
Related Work: Feedback Covariate Shift (FCS)



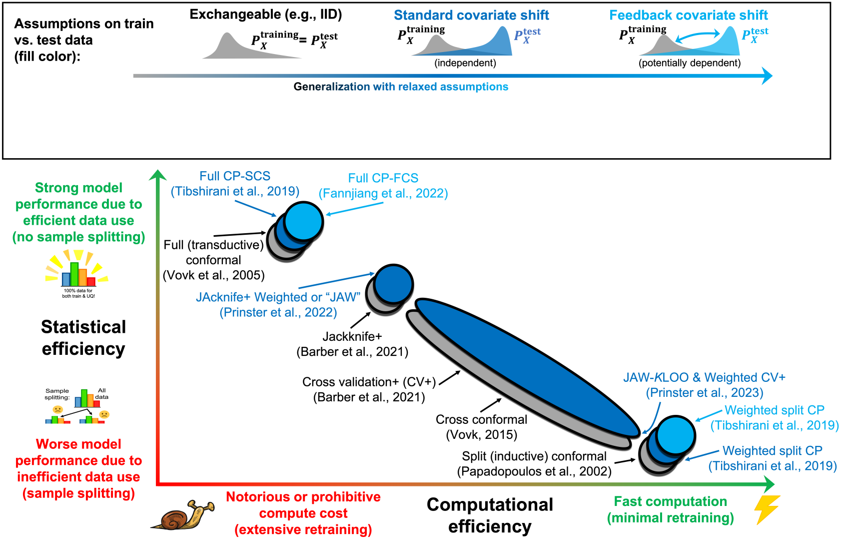
Related Work: Feedback Covariate Shift (FCS)



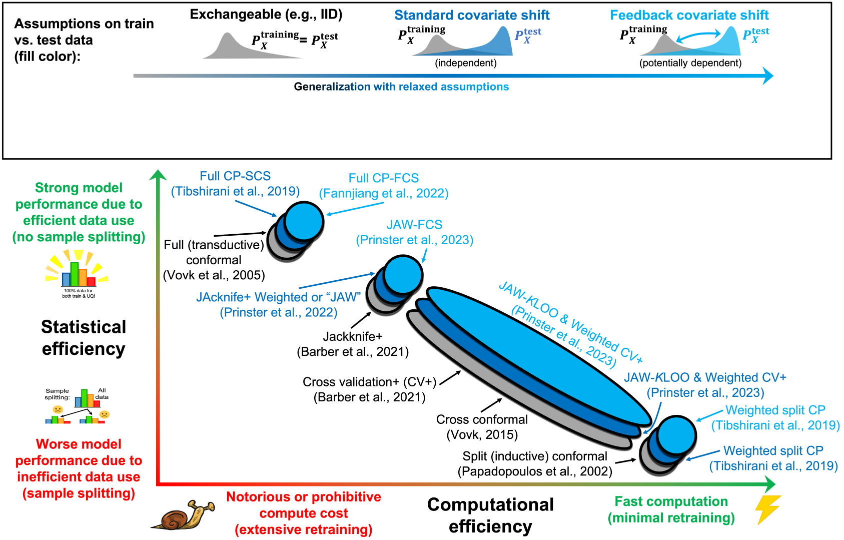
Related Work: Feedback Covariate Shift (FCS)



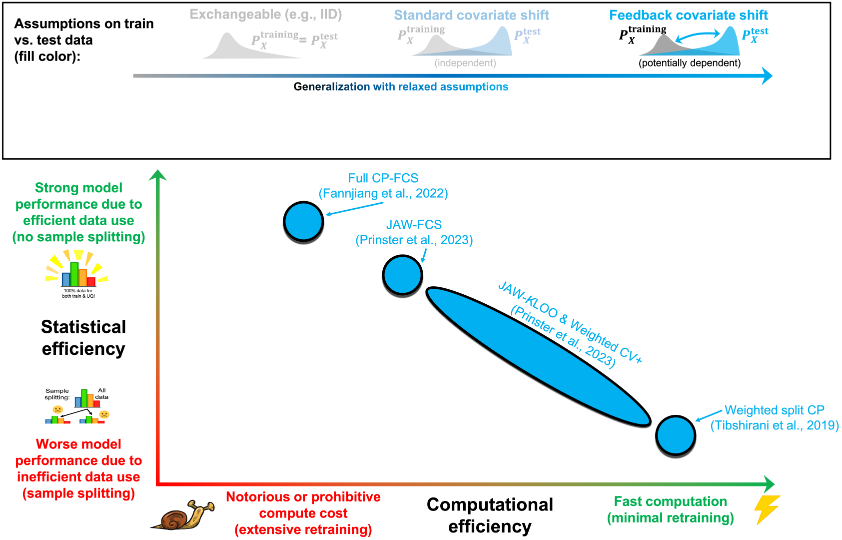
Related Work: Feedback Covariate Shift (FCS)



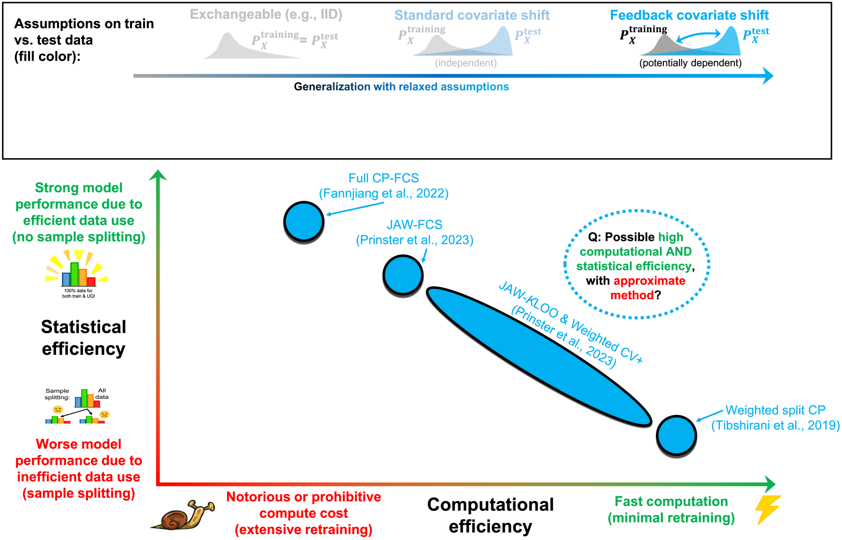
Related Work: Feedback Covariate Shift (FCS)



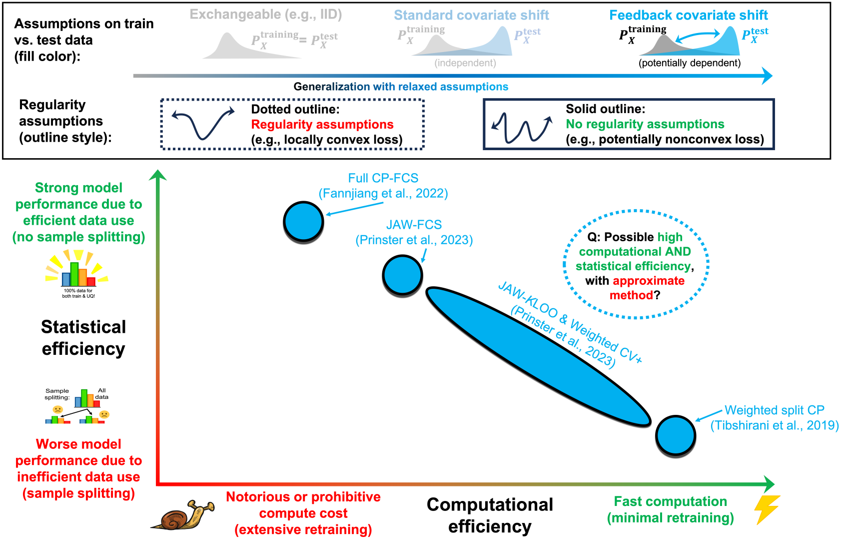
Related Work: Feedback Covariate Shift (FCS)



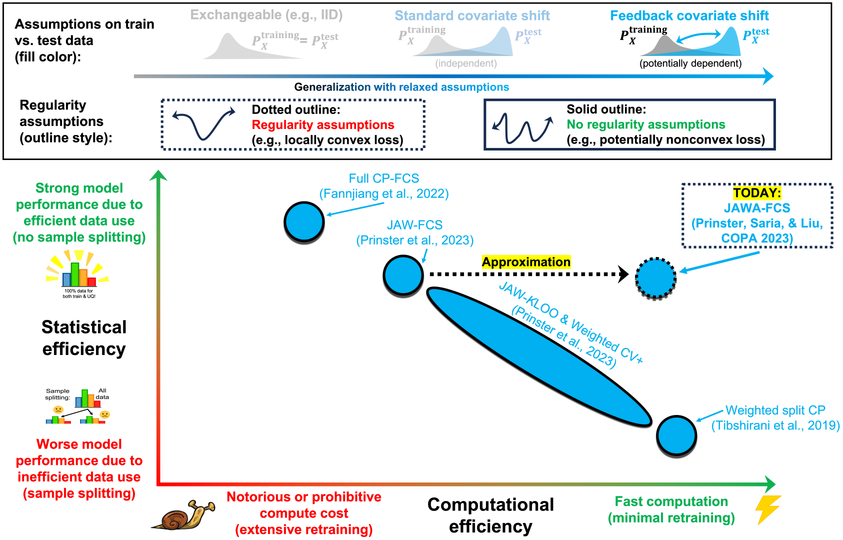
Related Work: Feedback Covariate Shift (FCS)



Related Work: Feedback Covariate Shift (FCS)



Proposed Work: Approximation of JAW-FCS



Presentation Outline

Jackknife+ predictive interval (Barber, Candes, Ramdas, & Tibshirani, 2021):

$$\widehat{C}_{n,\alpha}^{\text{Jackknife+}}(x) = \Bigg[Q_\alpha\Big(\sum_{j=1}^{n} \tfrac{1}{n+1} \delta_{\widehat{\mu}_{-j}(x) - |Y_j - \widehat{\mu}_{-j}(X_j)|} + \tfrac{1}{n+1}\delta_{-\infty} \Big), \\ Q_{1-\alpha}\Big(\sum_{j=1}^{n} \tfrac{1}{n+1} \delta_{\widehat{\mu}_{-j}(x) + |Y_j - \widehat{\mu}_{-j}(X_j)|} + \tfrac{1}{n+1}\delta_{\infty} \Big) \Bigg]$$

Some notation:

- $\delta_v :=$ point mass at value v
- $\widehat{\mu}_{-j} :=$ Leave-one-out (LOO) retrained model

$\Longrightarrow$ **Requires training n distinct predictors**

Background: JAcknife+ Weighted for FCS

(Fitted) likelihood ratio weights:

$$w(x;D) = \frac{\mathrm{d}P^{\text{test}}_{X;D}(x)}{\mathrm{d}P^{\text{train}}_{X}(x)}$$

e.g., $P^{\text{test}}_{X;D}(x) \propto \exp(\lambda \cdot \widehat{\mu}(x)))$

Normalized weights:

$$\widetilde{w}_{n+1,j}(x) = \frac{\overbrace{w(x;Z_{-j})\,w(X_j;Z_{-j})}^{\text{weights for LOO pt } j \text{ and test pt } n+1}}{\underbrace{\sum_{j'=1}^{n}\left[w(x;Z_{-j'})\,w(X_{j'};Z_{-j'})\right] + w(x;Z_{1:n})^2}_{\text{normalization constant}}}$$

JAckknife+ **W**eighted for **F**eedback **C**ovariate **S**hift or "**JAW-FCS**" (Prinster, Liu, & Saria, 2023):

$$\widehat{C}^{\text{JAW-FCS}}_{n,\alpha}(x) = \Bigg[Q_\alpha\Bigg(\sum_{j=1}^{n}\widetilde{w}_{n+1,j}(x)\,\delta_{\widehat{\mu}_{-j}(x)-|Y_j-\widehat{\mu}_{-j}(X_j)|} + \widetilde{w}_{(n+1)^2}(x)\,\delta_{-\infty}\Bigg),$$

$$Q_{1-\alpha}\Bigg(\sum_{j=1}^{n}\widetilde{w}_{n+1,j}(x)\,\delta_{\widehat{\mu}_{-j}(x)+|Y_j-\widehat{\mu}_{-j}(X_j)|} + \widetilde{w}_{(n+1)^2}(x)\,\delta_{\infty}\Bigg)\Bigg]$$

Note: Often $w(\cdot\,;Z_{-j})$ and $\widehat{\mu}_{-j}$ require the same LOO parameter est. $\hat{\theta}_{-i}$

Background: Influence Functions

Influence functions (Cook, 1977; Giordano, Jordan, & Broderick, 2019) approximate model parameter changes due to removing (or reweighting) a datapoint via a K-th order Taylor series.

$$\hat{\theta}_{-i}^{\text{IF-}K} := \hat{\theta} + \sum_{k=1}^{K} \frac{1}{k!} D_{-i}^{k} \hat{\theta}$$

$D_{-i}^{k}\hat{\theta} :=$ kth order derivative of parameters $\hat{\theta}$ w.r.t. removing point i

Main computational cost: Computing inverse Hessian

Prior works using IFs with jackknife+:

- Alaa and Van Der Schaar (2020) used higher order IFs to approximate the Jackknife+, but assume i.i.d. data
- Prinster, Liu, and Saria (2022) used higher orders to approximate the JAckknife+ Weighted for *Standard* Covariate Shift (JAW-SCS), but with different weights than in FCS

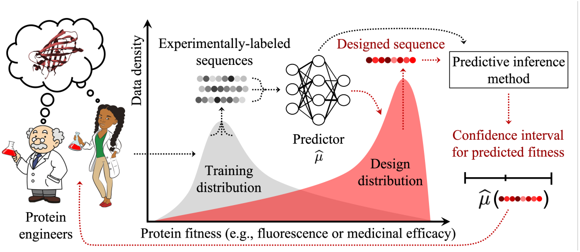
JAWA-FCS: **JA**cknife **W**eighted **A**pproximation for **F**eedback **C**ovariate **S**hift (K-th order Influence Funcation)

$$\widehat{C}^{\text{JAW-FCS}}_{n,\alpha}(x) = \Big[Q_\alpha\Big(\sum_{j=1}^{n} \widetilde{w}^{\text{IF}-K}_{n+1,j}(x)\,\delta_{\widehat{\mu}^{\text{IF}-K}_{-j}(x) - |Y_j - \widehat{\mu}^{\text{IF}-K}_{-j}(X_j)|} + \widetilde{w}^{\text{IF}-K}_{(n+1)^2}(x)\,\delta_{-\infty}\Big),$$
$$Q_{1-\alpha}\Big(\sum_{j=1}^{n} \widetilde{w}^{\text{IF}-K}_{n+1,j}(x)\,\delta_{\widehat{\mu}^{\text{IF}-K}_{-j}(x) + |Y_j - \widehat{\mu}^{\text{IF}-K}_{-j}(X_j)|} + \widetilde{w}^{\text{IF}-K}_{(n+1)^2}(x)\,\delta_{\infty}\Big)\Big]$$

Main idea: Approximating both the weights $w(\cdot\,; Z_{-j})$ and LOO predictions $\widehat{\mu}_{-j}$ using influence functions (IFs)

Presentation Outline

Experiments: Flourescent Protein Design Task

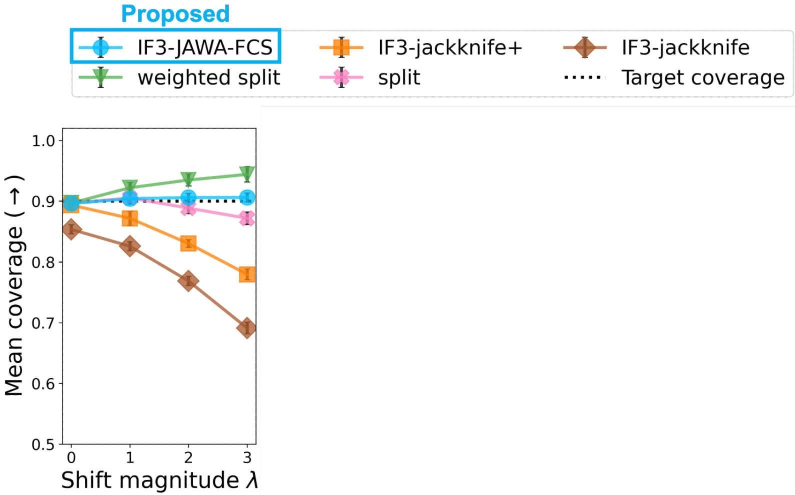


Experimental details:

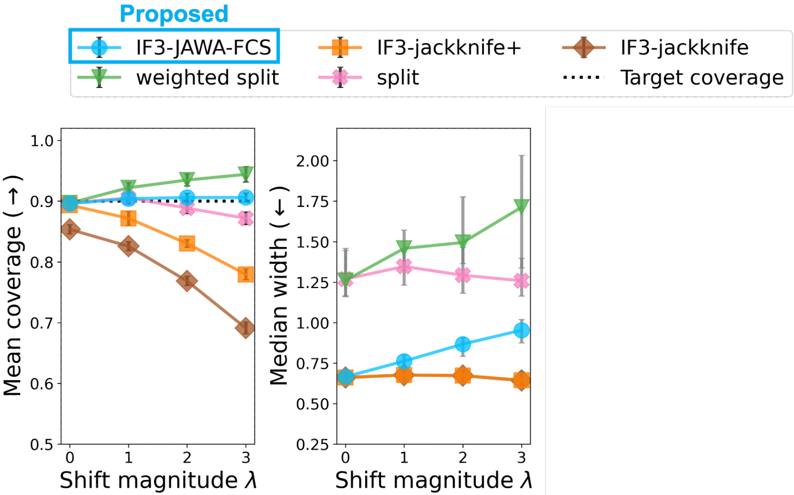
- $\hat{\mu}$: Small (25 hidden unit) neural network regressor with tanh activation function
- 0.5 L2 regularization parameter
- $n = 192$ training samples
- $K = 3$rd order influence function approximation
- $\alpha = 0.1$
- 20 experimental replicates

Runtime results: JAWA-FCS: <3 minutes JAW-FCS: 1 hour 24 minutes

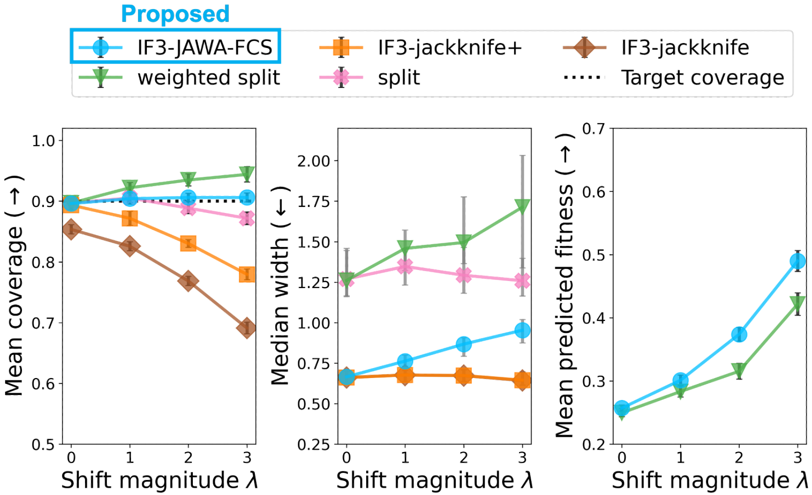
Flourescent Protein Design Results



Flourescent Protein Design Results



Flourescent Protein Design Results



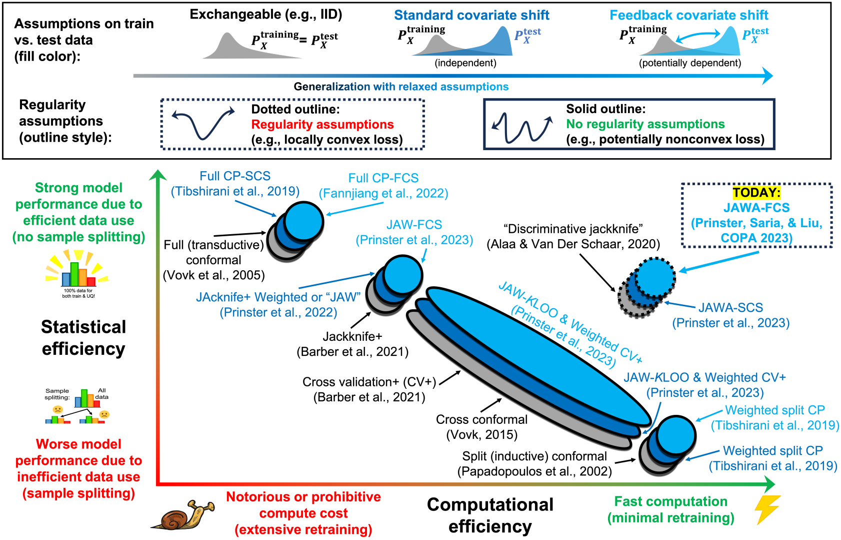
Presentation Outline

Limitations and Future Directions

Limitations and future directions:

- Experiments only with small neural net $\widehat{\mu} \Rightarrow$ See if results scale to larger $\widehat{\mu}$
- Empirical contribution only $\Rightarrow$ See how IF approximation error would impact guarantees.
- E.g., Giordano et al. (2019) give consistency conditions for LOO IF approximation (but do not consider guarantees for prediction estimates or coverage):
  - $\hat{\theta}$ is local minimum of objective function
  - Existence and boundedness of higher-order derivatives
  - Objective is strongly convex in neighborhood of $\hat{\theta}$

Today’s Contribution in Context (Visually)



Bibliography I

Alaa, A., & Van Der Schaar, M. (2020). Discriminative jackknife: Quantifying uncertainty in deep learning via higher-order influence functions. In *International conference on machine learning* (pp. 165–174).

Barber, R. F., Candes, E. J., Ramdas, A., & Tibshirani, R. J. (2021). Predictive inference with the jackknife+. *The Annals of Statistics*, *49*(1), 486–507.

Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, *19*(1), 15–18.

Dua, D., & Graff, C. (2017). *UCI machine learning repository.* Retrieved from `http://archive.ics.uci.edu/ml`

Bibliography II

Fannjiang, C., Bates, S., Angelopoulos, A. N., Listgarten, J., & Jordan, M. I. (2022). Conformal prediction under feedback covariate shift for biomolecular design. *Proceedings of the National Academy of Sciences*, *119*(43), e2204569119.

Giordano, R., Jordan, M. I., & Broderick, T. (2019). A higher-order swiss army infinitesimal jackknife. *arXiv preprint arXiv:1907.12116*.

Papadopoulos, H., Proedrou, K., Vovk, V., & Gammerman, A. (2002). Inductive confidence machines for regression. In *Machine learning: Ecml 2002: 13th european conference on machine learning helsinki, finland, august 19–23, 2002 proceedings 13* (pp. 345–356).

Poelwijk, F. J., Socolich, M., & Ranganathan, R. (2019). Learning the pattern of epistasis linking genotype and phenotype in a protein. *Nature communications*, *10*(1), 1–11.

Bibliography III

Prinster, D., Liu, A., & Saria, S. (2022). Jaws: Auditing predictive uncertainty under covariate shift. *Advances in Neural Information Processing Systems*.

Prinster, D., Liu, A., & Saria, S. (2023). Jaws-x: Addressing efficiency bottlenecks of conformal prediction under standard and feedback covariate shift. *International Conference on Machine Learning*.

Tibshirani, R. J., Foygel Barber, R., Candes, E., & Ramdas, A. (2019). Conformal prediction under covariate shift. *Advances in neural information processing systems*, *32*.

Vovk, V. (2015). Cross-conformal predictors. *Annals of Mathematics and Artificial Intelligence*, *74*(1), 9–28.

Vovk, V., Gammerman, A., & Shafer, G. (2005). *Algorithmic learning in a random world*. Springer Science & Business Media.

Acknowledgements

Thank you!!