Amortized Variational Bayesian Regression

Mert Ketenci, Victor A. Rodriguez, Noemie Elhadad, Adler Perotte

Columbia University, New York, NY

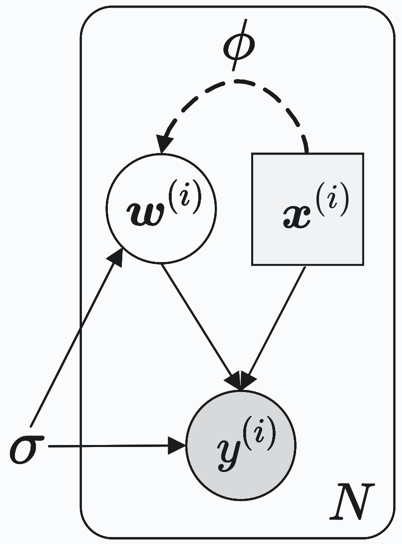
Background and Motivation

- Interpretability is as important as predictive performance in high-stakes decision-making domains, such as **healthcare**, where trusting the model and validating the predictions are crucial
- There often is a **trade-off** between performance and interpretability
- Two well-known branches: (**1**) Linear predictive models (**2**) Generalized additive models
- (**2**) is more **flexible** than (**1**), and hence, can model a broader range of data distribution. However, they **do not** have an **inherent** mechanism for estimating the **feature importance uncertainty and predictive uncertainty**
- Feature interactions are encoded **manually** in these models, and modeling every interaction may be computationally expensive

Method

- ### Graphical Model and Likelihood

- The graphical model and likelihood are as follows:



Model coefficient prior:

$$w^{(i)} \sim \mathcal{N}(0, I)$$

Regression task:

$$y^{(i)} \sim \mathcal{N}(\boldsymbol{x}^{\top(i)} \boldsymbol{w}^{(i)}, \sigma^2)$$

Classification task:

$$y^{(i)} \sim \mathcal{B}(S(\sigma \boldsymbol{x}^{\top(i)} \boldsymbol{w}^{(i)}))$$

Figure 1. The graphical model for AVBR. Solid and dahsed lines represent the predictive model $p_\sigma(y^{(i)}|\boldsymbol{x}^{(i)}, \boldsymbol{w}^{(i)})p(\boldsymbol{w}^{(i)})$ and variational approximation $q_\phi(\boldsymbol{w}^{(i)}|\boldsymbol{x}^{(i)})$, respectively (Kingma & Welling, 2014). The predictive model and variational approximation are learned jointly.

$$p_\sigma(Y, W|X) = \prod_{i=1}^{N} p_\sigma(y^{(i)}|\boldsymbol{x}^{(i)}, \boldsymbol{w}^{(i)})p(\boldsymbol{w}^{(i)})$$

- ### Remarks

1. We draw a unique weight vector for each (**x**,y) observation pair, **w**.

2. The model family implies that posterior p(**w**|**x**) = p(**w**).

3. Therefore, we rely on variational methods and define an approximate posterior distribution q(**w**|**x**).

- ### Inference and Learning

1. Kullback-Leibler divergence term encourages posterior to be marginally normal

2. Posterior distribution generates model weights conditioned on the covariates

3. The model relates **x** and y linearly through **w** (which is generated using **x**)

$$\tilde{\mathcal{L}}_M(\phi, \sigma; X_M, \boldsymbol{y}_M, \mathcal{E}_M) = \frac{1}{M}\sum_{i=1}^{M} \log p_\sigma(y^{(i)}|\boldsymbol{x}^{(i)}, g_\phi(\boldsymbol{x}^{(i)}, \boldsymbol{\epsilon}^{(i)})) - KL\left(q_\phi(\boldsymbol{w}^{(i)}|\boldsymbol{x}^{(i)})|p(\boldsymbol{w}^{(i)})\right)$$

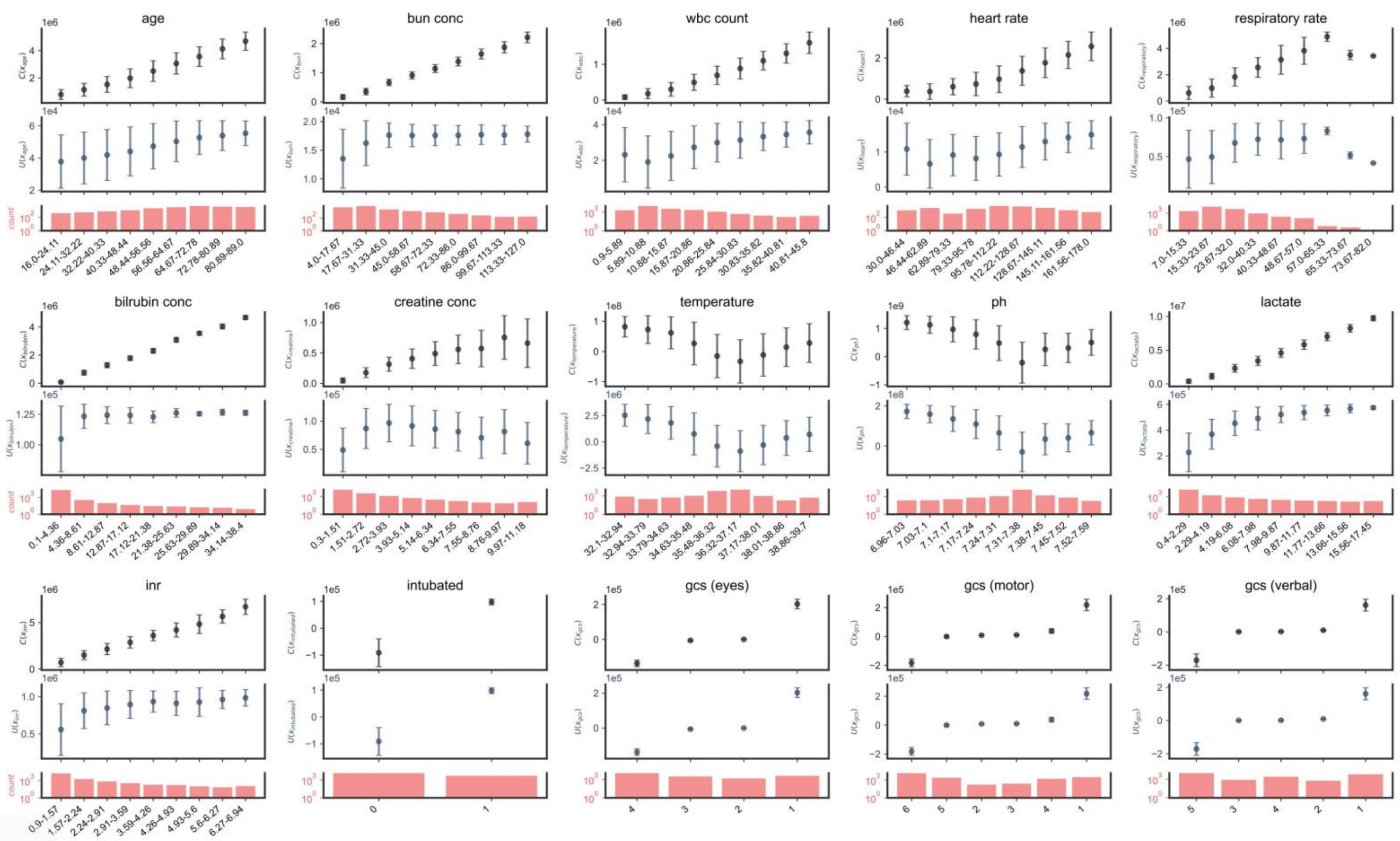
Results

- ### Quantitative

- **5** baseline models on **2 classifications** and **2 regression** datasets.
- **Area under the receiver operating characteristic curve** (ROC-AUC) and **root-mean-square error** (RMSE) for classification and regression metrics.
- Empirical results show that AVBR quantitatively performs **on par** with well-established machine learning algorithms.

- ### Qualitative

- We illustrate AVBR's interpretability on each dataset separately. Here we show the interpretability of the **Intensive Care Unit (ICU) dataset** (**higher** values indicate **higher** mortality likelihood):



Conclusion

We introduced AVBR - **an instance-wise linear model whose coefficients are given by a non-linear amortized inference network**. AVBR is **critical** for domains such as **healthcare**, where it is important to understand if there exist any **algorithmic biases** that would jeopardize **fairness** and **safety**. We demonstrated the quantitative and qualitative performance of AVBR on several benchmark datasets.