Bensemble

Abstract

We introduce Bensemble, a PyTorch library for Bayesian ensembling with optional Rényi-divergence–guided model selection and posterior estimation, covering Laplace, and PBP.

1 Introduction

Rapid progress in deep learning has set new accuracy records, but it has also revealed a gap between what models *predict* and how *certain* they are. Real systems operate under distribution shift, limited data, and changing costs of error, so we need predictions that come with reliable uncertainty.

At its core, *Bayesian ensembling* treats model parameters as random rather than fixed, and forms decisions by averaging predictions across plausible models. This produces a *posterior predictive* distribution that better reflects epistemic uncertainty and tends to calibrate confidence to reality.

Thus, we present our Python library **Bensemble** that makes Bayesian ensembling easy to use in PyTorch. The library centers on building posterior predictives through a simple interface for training, drawing model samples (or approximations) and combining them. We also expose a tunable α-divergence (Rényi) objective to steer selection and aggregation.

2 A short refresher on the Bayesian approach

Let $\mathcal{D}$ be data and w denote model parameters.

$$\textbf{Posterior:} \quad p(w \mid \mathcal{D}) \;\propto\; p(\mathcal{D} \mid w)\, p(w). \tag{1}$$

$$\textbf{Evidence:} \quad p(\mathcal{D}) \;=\; \int p(\mathcal{D} \mid w)\, p(w)\, dw \quad . \tag{2}$$

$$\textbf{Predict:} \quad p(y \mid x, \mathcal{D}) \;=\; \int p(y \mid x, w)\, p(w \mid \mathcal{D})\, dw. \tag{3}$$

3 Package content.

Overview of modules and APIs: trainers, objectives, metrics (RMSE/LL/Brier) with common interfaces.

4 Implementation details.

Details on the code and structure. What and how to call and compute.

5 Demo.

Runnable example: single-network baseline → ensemble with methods in the package. We show RMSE, NLL, Brier.

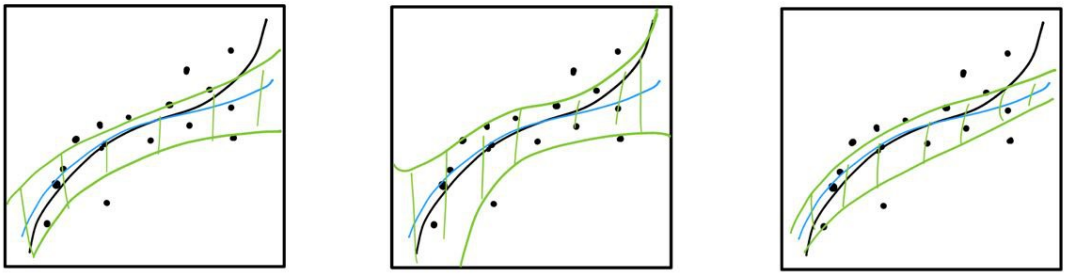

Figure 1: Preliminary variant of the graph we want to include. We want to compare predictions of various methods.

6 Conclusion.

Takeaways, when-to-use-what checklist, limitations (posterior misspecification, costs).