Uncertainty Disentanglement with Non-stationary Heteroscedastic Gaussian Processes for Active Learning

Zeel B Patel, Nipun Batra, Kevin Murphy

Overview

- Gaussian processes (GPs) are often used in active learning due to well-calibrated uncertainty prediction.
- The uncertainty can be split into aleatoric (irreducible) and epistemic (model) uncertainties.
- Only epistemic uncertainty is useful for model improvement in active learning.
- We propose a Non-stationary Heteroscedastic GP model which can disentangle epistemic and aleatoric uncertainties.

Model

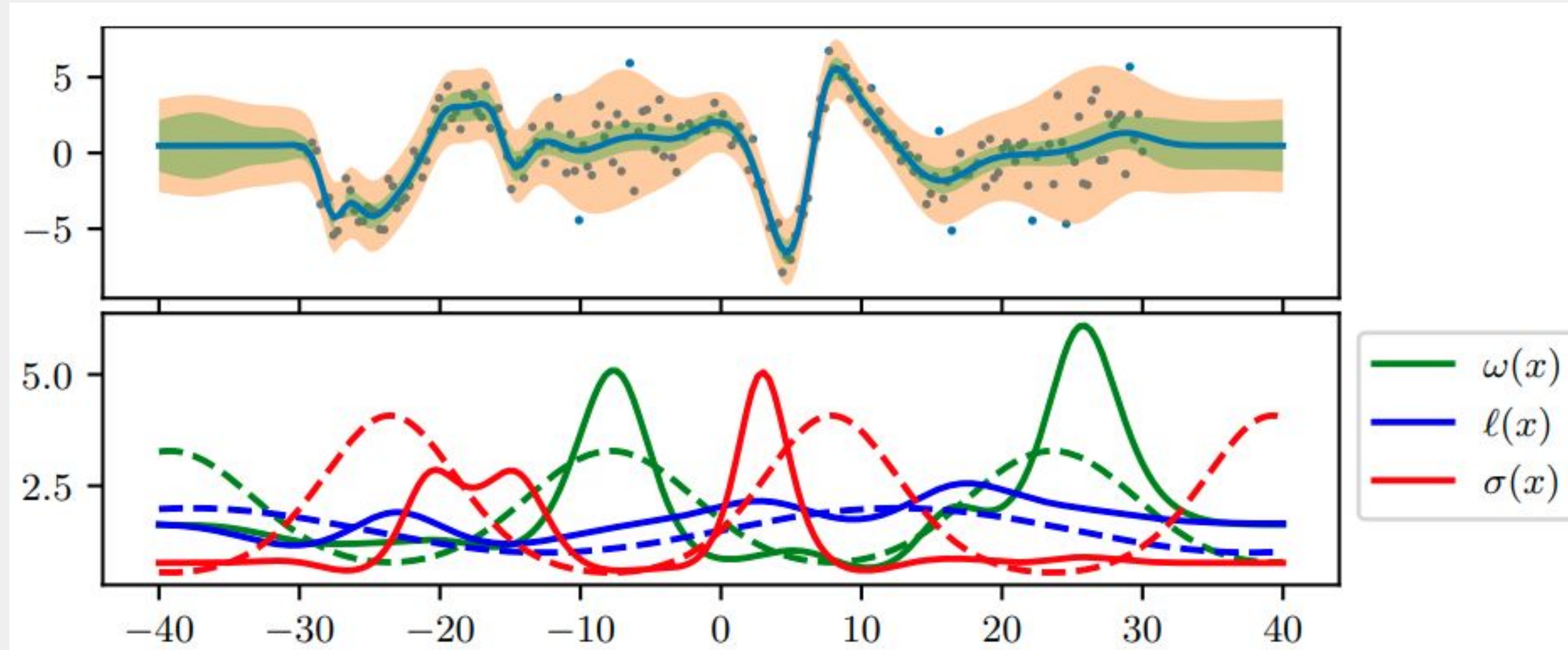
- Our model, kernel and likelihood noise are the following where all hyperparameters lengthscale ℓ, signal variance σ and noise variance ω are input depended.

$$\mathcal{K}_f(\mathbf{x},\mathbf{x}') = \sigma(\mathbf{x})\sigma(\mathbf{x}')\sqrt{\frac{2\ell(\mathbf{x})\ell(\mathbf{x}')}{\ell(\mathbf{x})^2+\ell(\mathbf{x}')^2}}\exp\left(-\frac{||\mathbf{x}-\mathbf{x}'||^2}{\ell(\mathbf{x})^2+\ell(\mathbf{x}')^2}\right)$$

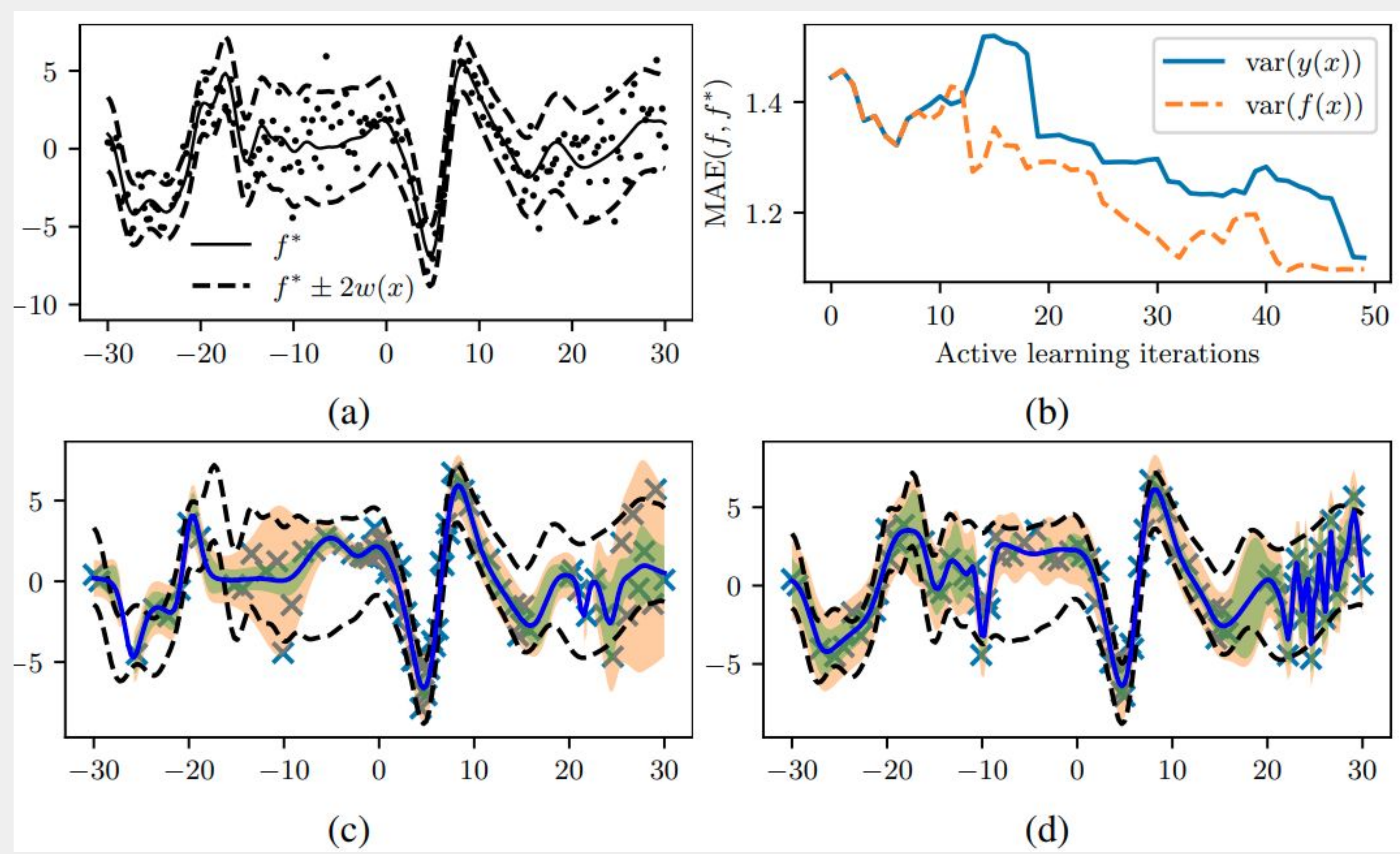
$$y(\mathbf{x}) = f(\mathbf{x}) + \varepsilon(\mathbf{x}), \quad \varepsilon(\mathbf{x}) \sim \mathcal{N}\left(0, \omega(\mathbf{x})^2\right)$$
$$f(\mathbf{x}) \sim GP(0, \mathcal{K}_f(\mathbf{x},\mathbf{x}'))$$

- We use a latent GP to learn the input depended hyperparameters with help of inducing points.

Model fit on a simulated dataset



Active learning



Fit after sampling 50 points with active learning using (c) overall uncertainty; (d) epistemic uncertainty. (b) Mean Squared Error (MAE) between predicted function f and ground truth f* is improving faster in (d) with epistemic uncertainty as compared to (c) overall uncertainty.

Datasets

- We use Jump1D, Motorcycle Helmet and NONSTAT-2D from previous literature.

Results & Insights

Model	Jump		Motorcycle		NONSTAT-2D	
	NLPD	RMSE	NLPD	RMSE	NLPD	RMSE
Stationary Homoskedastic GP	4.98	0.26	11.96	0.44	-50.72	0.09
(ℓ)-GP	5.01	0.26	11.92	0.44	-65.13	0.06
(ω)-GP	3.82	0.22	5.21	0.44	-50.81	0.09
(σ)-GP	0.92	0.30	11.56	0.44	-56.66	0.07
(ℓ,ω)-GP	5.01	0.26	5.68	0.45	-65.31	0.06
(ℓ,σ)-GP	-2.18	0.22	11.54	**0.44**	-49.28	0.07
(σ,ω)-GP	0.92	0.22	4.21	0.46	-54.35	0.10
(ℓ,σ,ω)-GP	**-2.20**	**0.22**	**4.09**	0.45	**-73.74**	**0.05**

- Metrics are the following: Negative Log Predictive Density (NLPD) Root Mean Squared Error (RMSE).
- Rows represent different methods in which we either used fixed or input-dependent length scale ℓ, signal variance σ and observation noise ω.
- (ℓ, σ, ω)-GP is the best or the second best across all datasets and metrics.