Deep Evidential Regression on PAINN Graph Neural Network

Alexandra Polymenopoulou (s212558), Melina Siskou (s213158), Paolo Federico (s212975)

Model Uncertainty

"Understanding when a model cannot be trusted"

Ability to infer model uncertainty is crucial for wide-scale adoptions of models, since it helps to interpret confidence when dealing with out-of-distribution (OOD) test samples, and recognizes when the model is likely to fail.

Aim of the project

Demonstrate that the model can assign high uncertainty to OOD samples and low uncertainty to in-distribution ones while performing accurate predictions.

Evidential Deep Learning

From Maximum Likelihood Estimation to Distribution of Distributions

Instead of assuming that μ and σ are known things that our network could predict:

$$(y_1, \ldots, y_N) \sim \mathcal{N}(\mu, \sigma^2)$$

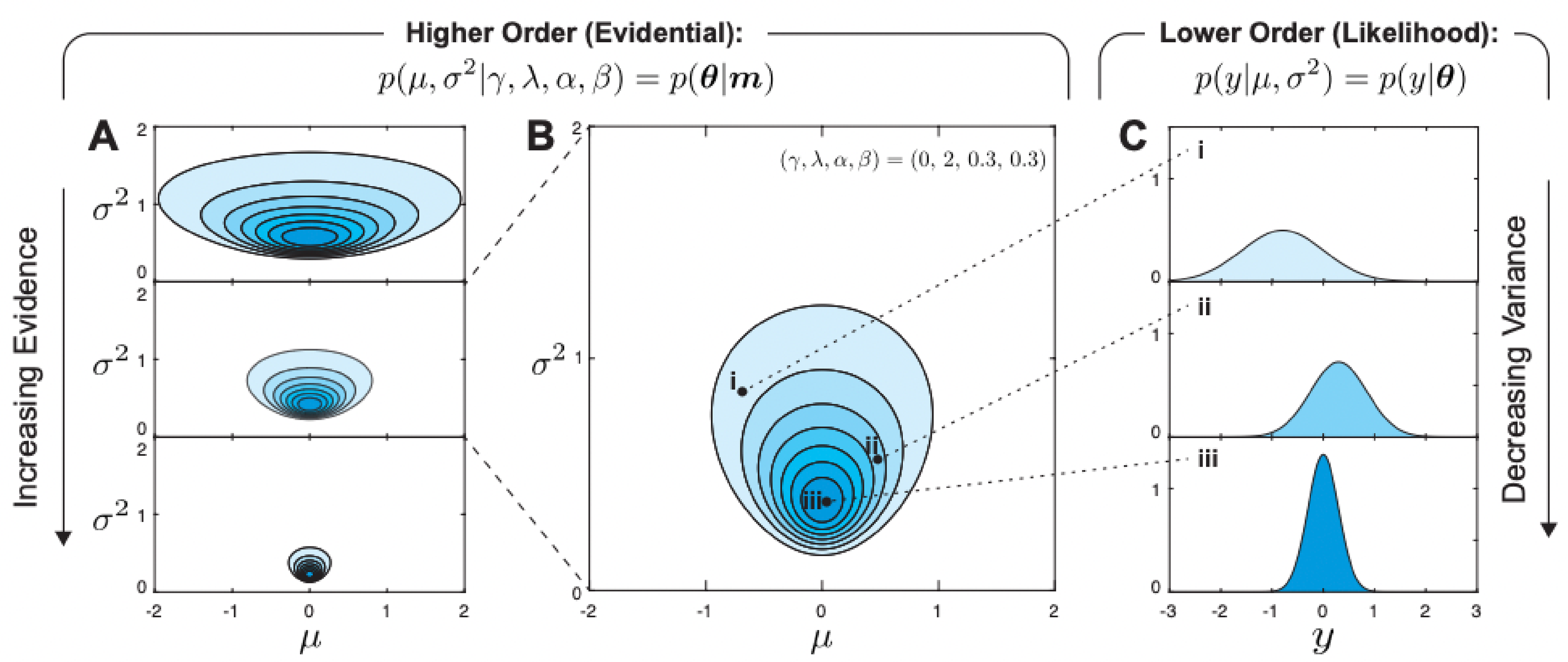
We consider the problem where the observed targets y are drawn from a Gaussian distribution as in Maximum Likelihood Estimation, but now with unknown mean and variance. By placing evidential priors over the original likelihood function, the NN is trained to infer the hyperparameters of the evidential distribution.

$$\mu \sim \mathcal{N}(\gamma, \sigma^2 \upsilon^{-1}) \qquad \sigma^2 \sim \Gamma^{-1}(\alpha, \beta).$$

where $\Gamma(\cdot)$ is the gamma function, $\boldsymbol{m} = (\gamma, \upsilon, \alpha, \beta)$, and $\gamma \in \mathbb{R}$, $\upsilon > 0$, $\alpha > 1$, $\beta > 0$.

Evidential learning tries to enable direct estimation of epistemic and aleatoric uncertainty by trying to learn these higher order distributions over the individual likelihood parameters.

$$\underbrace{\mathbb{E}[\mu] = \gamma}_{\text{prediction}}, \qquad \underbrace{\mathbb{E}[\sigma^2] = \tfrac{\beta}{\alpha-1}}_{\text{aleatoric}}, \qquad \underbrace{\operatorname{Var}[\mu] = \tfrac{\beta}{\upsilon(\alpha-1)}}_{\text{epistemic}}.$$



Learning the Evidential Distribution

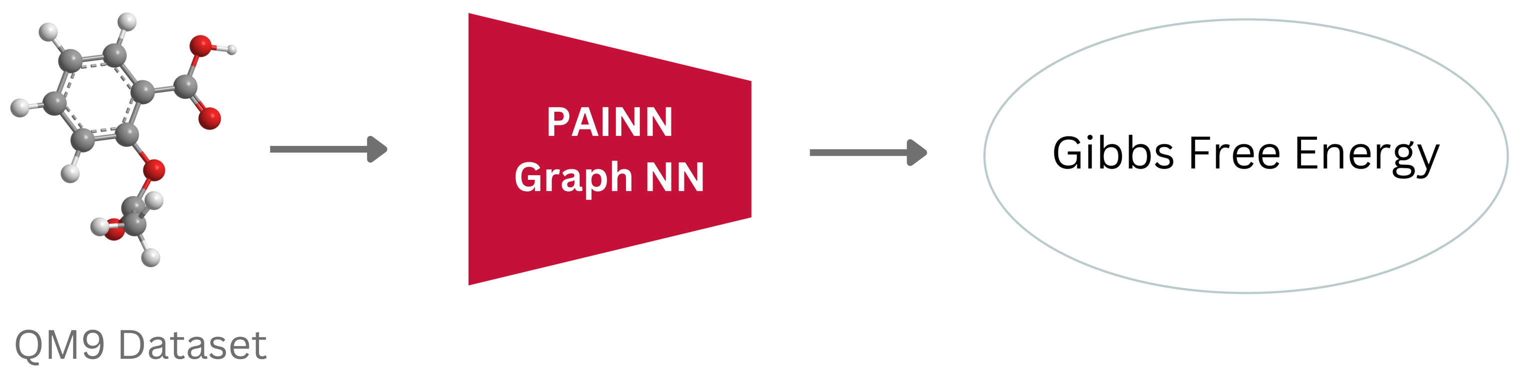
1. Maximizing the model fit by maximizing model evidence in support of our observations.
2. Minimizing evidence on errors by inflating uncertainty when the prediction is wrong.

$$\mathcal{L}_i(\boldsymbol{w}) = \mathcal{L}_i^{\mathrm{NLL}}(\boldsymbol{w}) + \lambda \mathcal{L}_i^{\mathrm{R}}(\boldsymbol{w}).$$

Here, λ trades off uncertainty inflation with model fit. Setting λ = 0 yields an over-confident estimate while setting λ too high results in over-inflation.

PAINN Model

Polarized Atom Interaction Neural Network



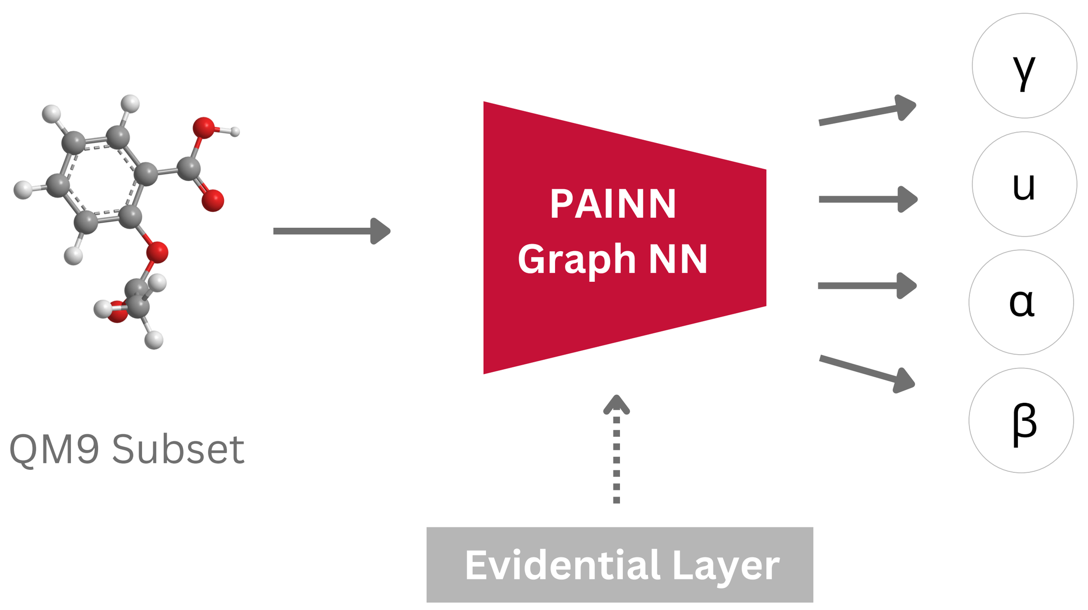
1. Embedding / Gaussian Expansion of nodes and edges
2. Message Passing: For each node, gathers the neighboring embeddings, aggregates them, and passes them through an update function
3. NN layers: Dense (Shifted Soft Plus) → Dense → 1 output neuron
4. Graph Output: Sums all the atom predictions into a single molecule prediction
5. Loss function: Root Mean Squared Error (RMSE)

Integrating an Evidential Layer

Instead of predicting a molecule property, the model now predicts the parameters γ, u, α, β of the evidential distribution.

1. Embedding / Gaussian Expansion of nodes and edges
2. Message Passing
3. Dense (Shited Soft Plus)
4. Graph Output Sum
5. Evidential layer: Dense(Softmax) → 4 output distribution param.
6. Evidential regression supported loss function

Conversion from TensorFlow to PyTorch to fit PAINN implementation



Training Parameters

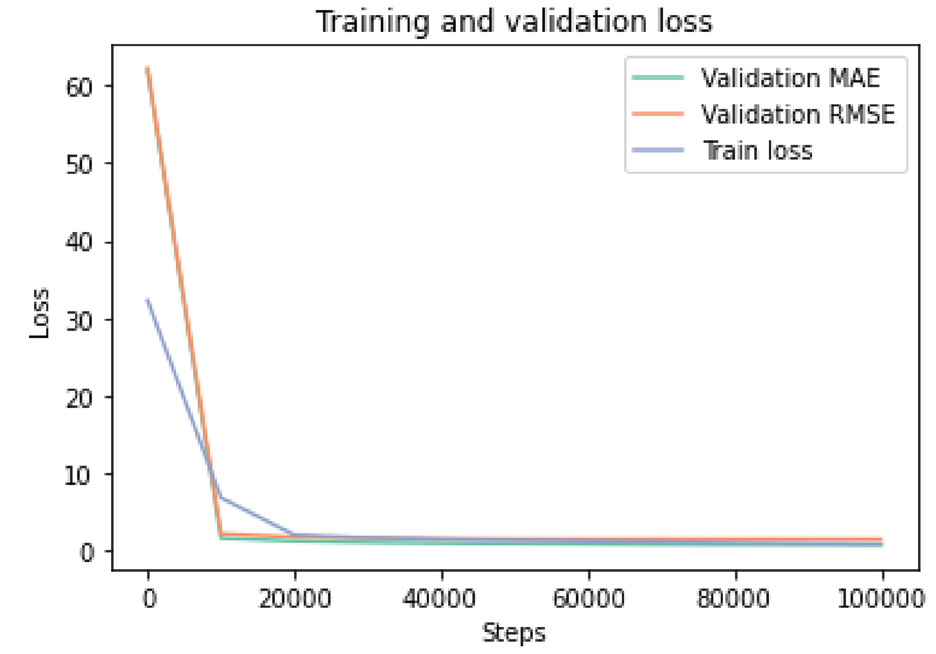
- Experimenting with λ values in loss calculation. Best results using λ = 1e-6
- Experimenting with learning rates. Best results using 1e-3
- Applying a weight decay to the optimizer

Problem Setup

Training with in-distribution data

Subset ~2/3 of the initial training QM9 dataset by filtering molecules with:

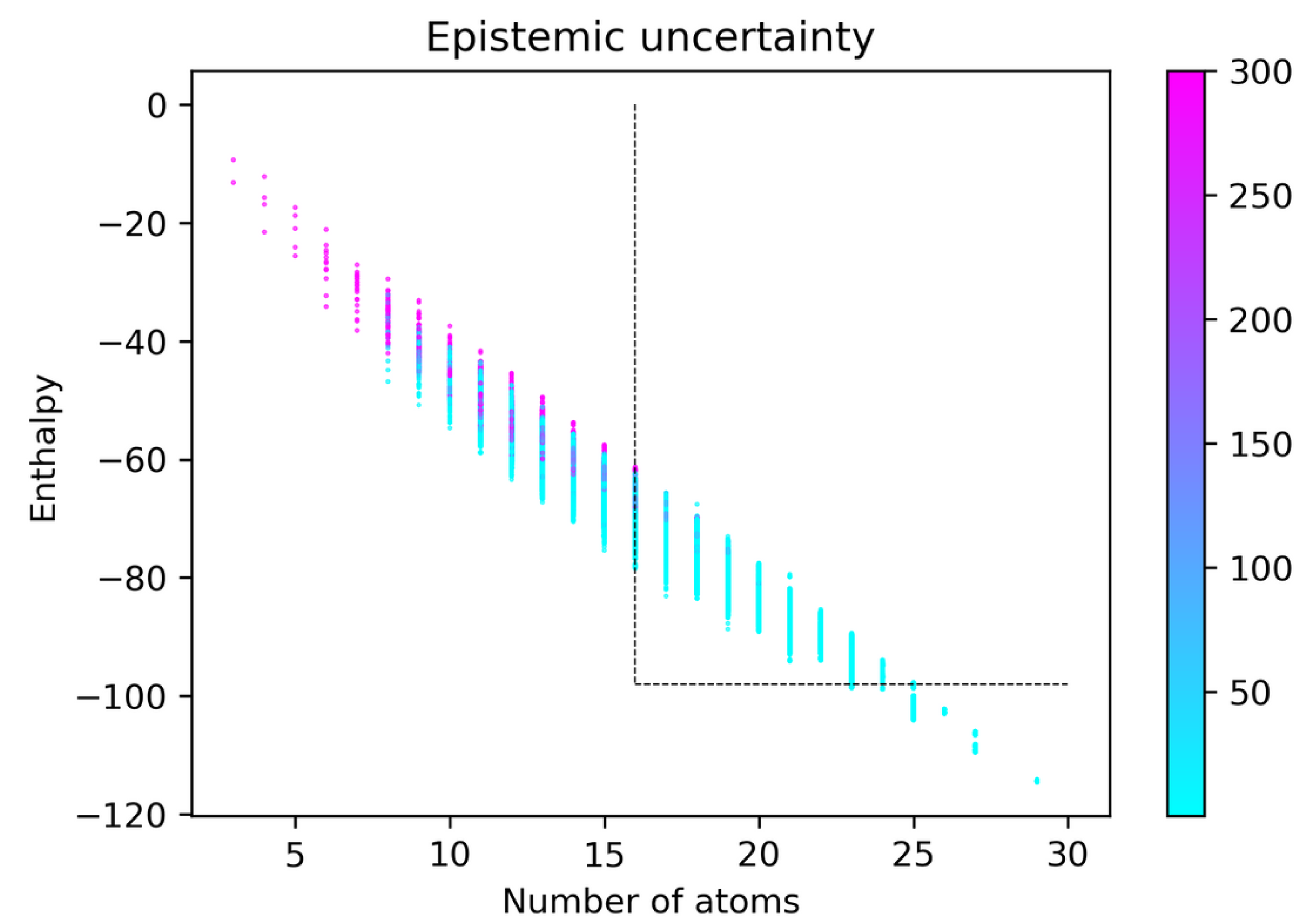
- number of atoms > 16
- enthalpy value > -98



Using best model to predict out-of-distribution data

Predicting using as input the whole QM9 dataset (including the subset used during training). What we aim to see is the model predicting low uncertainty on the data we assumed to be in-distribution and high uncertainty on the rest 1/3 of the dataset, the part we assume as OOD.

Results



Out-of-distribution loss	*In-distribution loss*
1.300	*0.103*

References

- Schütt, Kristof, Oliver Unke, and Michael Gastegger. "Equivariant message passing for the prediction of tensorial properties and molecular spectra." International Conference on Machine Learning. PMLR, 2021.
- Amini, Alexander, et al. "Deep evidential regression." Advances in Neural Information Processing Systems 33 (2020): 14927-14937.
- Amini, A., Schwarting, W., Soleimany, A., & Rus, D. Deep Evidential Regression [Computer software]. https://github.com/aamini/evidential-deep-learning