Hybrid Discriminative Models

Zhenwen Dai

Amazon

2019-04-08

Discriminative model

- The aim is to learn a functional relationship:

$$y = f(x) + \epsilon$$

- There are multiple ways to parametrize a functional relationship.
- For example, a basis function model:

$$f(x) = \sum_k w_k \phi_k(x), \quad w_k \sim \mathcal{N}(0, 1)$$

where $\{\phi_k(x)\}_k$ denotes the set of basis functions.

Gaussian process

- Gaussian process has *infinite* number of basis functions.

$$p(\mathbf{y}|\mathbf{X}) = \mathcal{N}\left(\mathbf{y}|0, \mathbf{K}\right)$$

  where the covariance matrix is computed from the set of inputs $\mathbf{X}$ using the kernel function $k(\cdot, \cdot)$.

A hybrid discriminative model

- A discriminative model with a latent input

$$p(\mathbf{y}|\mathbf{X}, \mathbf{H})p(\mathbf{H})$$

- Missing information
  - Missing information in individual data points: flexible uncertainty
  - Missing information shared across multiple data points: multi-output, multi-task, meta-model

Missing information in individual data points

- One latent variable per data point:

$$\mathbf{y} = (y_1, \ldots, y_N), \quad \mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_N), \quad \mathbf{H} = (\mathbf{h}_1, \ldots, \mathbf{h}_N).$$

$$y_n = f(\mathbf{x}_n, \mathbf{h}_n) + \epsilon$$

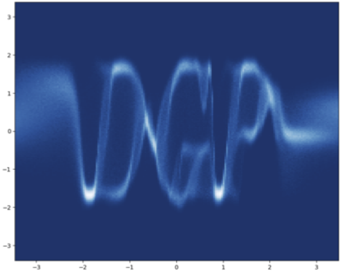


Figure 1: Multi-modal regression (taken from the slides of Hugh Salimbeni)

- This idea has been applied to BNN (Depeweg et al. 2018) and DGP.

Missing information shared across multiple data points

- Clustering of GP: (Hensman, Rattray, and Lawrence 2015), (Lawrence, Ek, and Campbell 2018)
- Multi-output GP with latent space: (Dai, Álvarez, and Lawrence 2017)

A Toy Problem: The Braking Distance of a Car

- To model the braking distance of a car in a *completely data-driven* way.
  - Input: the speed when starting to brake
  - Output: the distance that the car moves before fully stopped
  - We know that the braking distance depends on the friction coefficient.
  - We can conduct experiments with a set of different tyre and road conditions

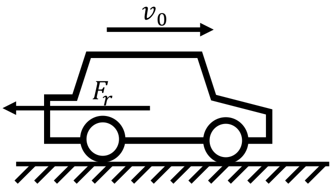


Figure 2: car brakding distance

A non-parametric regression

- GP is the natural choice for such a non-parametric regression problem.

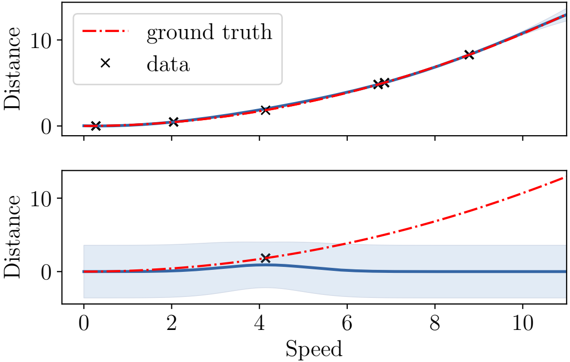


Figure 3: A GP fit

One shot learning

- What if we drive on a different road or changing the tyres?
- Do we need to completely redo the fitting?

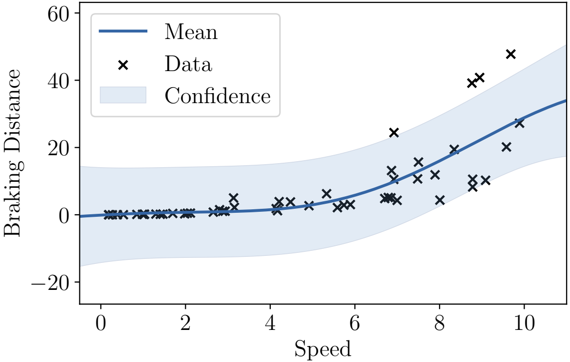
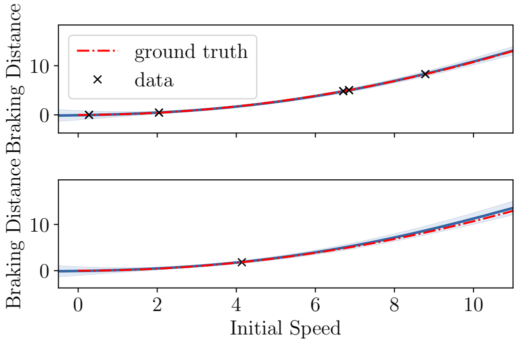


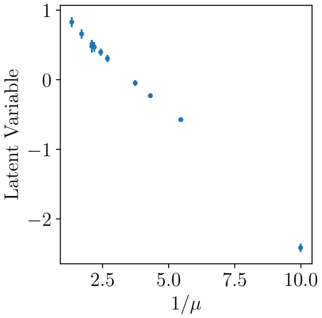
Figure 4: Ignore the difference in condition

Assume a latent variable in the model

- Assume a latent variable representing the road/car condition.

$$y_{n,c} = f(\mathbf{x}_{n,c}, \mathbf{h}_c) + \epsilon, \quad f \sim GP, \quad \mathbf{h}_c \sim \mathcal{N}(0, \mathbf{I})$$





A meta-model

- Modeling beyond a single task has been the focus.
- A generative model for tasks
- A combination of discriminative and generative model
- A generative model with a fansy likelihood (a discriminative model)

Applications

- Meta-model for multi-task Bayesian optimization
- Meta-model for reinforcement learning

References I

Dai, Zhenwen, Mauricio A Álvarez, and Neil D Lawrence. 2017. "Efficient Modeling of Latent Information in Supervised Learning Using Gaussian Processes." In *Advances in Neural Information Processing Systems*.

Depeweg, Stefan, José Miguel Hernández-Lobato, Finale Doshi-Velez, and Steffen Udluft1. 2018. "Decomposition of Uncertainty in Bayesian Deep Learning for Efficient and Risk-Sensitive Learning." In *International Conference on Machine Learning*.

Hensman, James, Magnus Rattray, and Neil D. Lawrence. 2015. "Fast Nonparametric Clusteringof Structured Time-Series." *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE* 37 (2).

Lawrence, Andrew R, Carl Henrik Ek, and Neill D F Campbell. 2018. "DP-Gp-Lvm: A Bayesian Non-Parametric Model for Learning Multivariate Dependency Structures." In *Arxiv*.