Archipelago: Nonparametric Bayesian Semi-Supervised Learning

Ryan Prescott Adams
Zoubin Ghahramani

University of Cambridge
http://www.inference.phy.cam.ac.uk/rpa23/

ICML, June 2009
The Semi-Supervised Learning Task
A Generative View on SSL

The Clustering Assumption

Semi-Supervised Learning assumes that the gaps between the densities reflect the class boundaries.

The class-wise densities cannot be independent!
Outline

A Generative Gaussian Process Density Model

Archipelago: Coupled Gaussian Process Densities

Inference

Toy Examples

Summary and Future Directions
Outline

A Generative Gaussian Process Density Model

Archipelago: Coupled Gaussian Process Densities

Inference

Toy Examples

Summary and Future Directions
Gaussian Processes

A distribution on functions specified by a kernel function.

The marginalization properties of the Gaussian distribution make it tractable with finite computation.
A Gaussian Process Density Model

1) Draw from the GP:
\[ g(x) \sim \mathcal{GP}(\theta) \]

2) Squash with logistic:
\[ \frac{1}{1 + \exp\{-g(x)\}} \]

3) Multiply by base density:
\[ \frac{\pi(x)}{1 + \exp\{-g(x)\}} \]

4) Normalize into PDF:
\[ f(x) = \frac{1}{\mathcal{Z}} \frac{\pi(x)}{1 + \exp\{-g(x)\}} \]
Samples from the Prior
The Gaussian Process Density Sampler

1) Draw proposals from the base density $\pi(x)$.

2) Draw the function from the GP at the proposals.

3) Squash the function and draw vertical coordinates in $(0, 1)$.

4) Discard proposals above the sampled function.
The Gaussian Process Density Sampler

Generative Process Allows Tractable Inference

- Generally, normalizing GP density models is hard.
- The generative process allows us to avoid this.
- We use MCMC on a latent variable model of the generative process.

The Main Idea

- GPs are already good for classification.
- Combine the GP density model and the GP classification model to solve SSL.
Outline

A Generative Gaussian Process Density Model

Archipelago: Coupled Gaussian Process Densities

Inference

Toy Examples

Summary and Future Directions
Revisit: Gaussian Process Density Sampler

1) Draw from the GP:
   
   \[ g(x) \sim \mathcal{GP}(\theta) \]

2) Squash with logistic:

   \[
   \frac{1}{1 + \exp\{-g(x)\}}
   \]

3) Multiply by base density:

   \[
   \frac{\pi(x)}{1 + \exp\{-g(x)\}}
   \]

4) Normalize into PDF:

   \[
   f(x) = \frac{1}{\mathcal{Z}} \frac{\pi(x)}{1 + \exp\{-g(x)\}}
   \]

\[\mathcal{Z} = \int_{-\infty}^{\infty} \frac{\pi(x)}{1 + \exp\{-g(x)\}} dx\]
Archipelago: Coupled GP Densities

1) \( K \) functions from the GP:

\[
g_k(x) \sim \mathcal{GP}(\theta_k)
\]

2) Apply \( K + 1 \) softmax:

\[
p(k \mid x) = \frac{\exp\{g_k(x)\}}{1 + \sum_{k'} \exp\{g_{k'}(x)\}}
\]

3) Multiply by base density:

\[
\pi(x) \frac{\sum_{k'} \exp\{g_{k'}(x)\}}{1 + \sum_{k'} \exp\{g_{k'}(x)\}}
\]

4) Normalize into PDF:

\[
\frac{1}{\mathcal{Z}} \frac{\pi(x) \sum_{k'} \exp\{g_{k'}(x)\}}{1 + \sum_{k'} \exp\{g_{k'}(x)\}}
\]
Generating Archipelago Data

1) Draw proposals from the base density $\pi(x)$.

2) Draw the functions from the GPs at the proposals.

3) Apply the softmax and sample categories.

4) Discard rejections and keep labels.
Outline

A Generative Gaussian Process Density Model

Archipelago: Coupled Gaussian Process Densities

Inference

Toy Examples

Summary and Future Directions
Marginalizing Out Unknown Labels

We have $N$ labeled data $\{x_n, l_n\}_{n=1}^{N}$ and $Q$ unlabeled data $\{x_q\}_{q=1}^{Q}$. How to deal with unlabeled data?

“The Right Bayesian Thing To Do”

When we don’t know the label, marginalize it out.

Getting the Marginalization for Free

Any generative model for SSL specifies $p(\text{features, label})$. How easy is it to marginalize out the label?

\[
p(\text{feat}, \text{lab}) = p(\text{feat}) p(\text{lab}) \quad \text{No cluster assumption.}
\]

\[
p(\text{feat}, \text{lab}) = p(\text{lab}) p(\text{feat} | \text{lab}) \quad \text{HARD}
\]

\[
p(\text{feat}, \text{lab}) = p(\text{feat}) p(\text{lab} | \text{feat}) \quad \text{EASY}
\]
The likelihood function:

\[
p(\{x_n, l_n\}_{n=1}^{N}, \{x_q\}_{q=1}^{Q} | \{g_k(x)\}_{k=1}^{K}) =
\]

\[
\frac{1}{Z} \prod_{n=1}^{N} \frac{\pi(x_n) \exp\{g_{l_n}(x_n)\}}{1 + \sum_{k'} \exp\{g_{k'}(x_n)\}} \prod_{q=1}^{Q} \frac{\pi(x_q) \sum_{k'} \exp\{g_{k'}(x_q)\}}{1 + \sum_{k'} \exp\{g_{k'}(x_q)\}}
\]

Labeled Data \hspace{2cm} \text{Unlabeled Data}
Doubly-Intractable Posterior

\[
\frac{1}{Z} p^*(\{x_n, l_n\}_{n=1}^N, \{x_q\}_{q=1}^Q | \{g_k(x)\}_{k=1}^K) \frac{p(\{g_k(x)\}_{k=1}^K)}{p(\{x_n, l_n\}_{n=1}^N, \{x_q\}_{q=1}^Q)}
\]

- The likelihood is only available within a constant.
- We cannot use vanilla MCMC methods.
- The generative models saves us.
Inference Via the Latent History

Construct a latent variable model with the unknown state:

- The unknown number of rejections \( M \).
- The locations of the rejections \( \{x_m\}_{m=1}^M \).
- The value of the functions \( \{g_k(x)\}_{k=1}^K \).

\[
p(\{g_k\}_{k=1}^K, M, \{x_m\}_{m=1}^M, \{x_q\}_{q=1}^Q, \{x_n, l_n\}_{n=1}^N) =
\prod_{k=1}^K \mathcal{G} \mathcal{P}(\{g_k(x_n)\}_{n=1}^N, \{g_k(x_q)\}_{q=1}^Q, \{g_k(x_m)\}_{m=1}^M)
\times \prod_{n=1}^N \frac{\exp\{g_{l_n}(x_n)\} \pi(x_n)}{1 + \sum_{k'} \exp\{g_{k'}(x_n)\}} \prod_{q=1}^Q \frac{\pi(x_q) \sum_{k'} \exp\{g_{k'}(x_q)\}}{1 + \sum_{k'} \exp\{g_{k'}(x_q)\}}
\times \prod_{m=1}^M \frac{\pi(x_m)}{1 + \sum_{k'} \exp\{g_{k'}(x_m)\}}
\]
Inference Via Latent History MCMC

Sample the Number of Latent Rejections

- Make birth/death proposals.
- For births, propose locations and function values.
- Accept/reject via Metropolis–Hastings.

Sample the Locations of Latent Rejections

- Accept/reject via Metropolis–Hastings.

Sample the Latent Function Values

- Use Hamiltonian Monte Carlo to avoid random walk behavior.

Also Sample GP and $\pi(x)$ Hyperparameters
Outline

A Generative Gaussian Process Density Model

Archipelago: Coupled Gaussian Process Densities

Inference

Toy Examples

Summary and Future Directions
3 Class Pinwheels

Archipelago

Softmax GP
3 Class Pinwheels

Archipelago

Marginal Entropy
4 Class Pinwheels

Archipelago

Softmax GP
5 Class Pinwheels

Archipelago

Softmax GP
Outline

A Generative Gaussian Process Density Model

Archipelago: Coupled Gaussian Process Densities

Inference

Toy Examples

Summary and Future Directions
Summary and Future Directions

Summary

- Nonparametric Bayesian model for semi-supervised learning.
- Couples together Gaussian process densities.
- Leverages easy label marginalization.
- Uses generative model to avoid intractability.

Future Directions

- Sparse GP methods for larger data sets.
- More sophisticated $\pi(x)$ for higher dimensions.
- Can work with arbitrary positive definite kernels.