Gaussian processes (GPs) provide a nonparametric representation of functions. However, classical GP inference suffers from high computational cost for big data. We propose a new Bayesian approach, EigenGP, that learns both basis dictionary elements—eigenfunctions of a GP prior—and prior precisions in a sparse finite model. It is well known that, among all orthogonal basis functions, eigenfunctions can provide the most compact representation. Unlike other sparse Bayesian finite models where the basis function has a fixed form, our eigenfunctions live in a reproducing kernel Hilbert space as a finite linear combination of kernel functions. We learn the dictionary elements—eigenfunctions—and the prior precisions over these elements as well as all the other hyperparameters from data by maximizing the model marginal likelihood. We explore computational linear algebra to simplify the gradient computation significantly. Our experimental results demonstrate improved predictive performance of EigenGP over alternative sparse GP methods as well as relevance vector machines.

**ABSTRACT**

Gaussian processes (GPs) provide a nonparametric representation of functions. However, classical GP inference suffers from high computational cost for big data. We propose a new Bayesian approach, EigenGP, that learns both basis dictionary elements—eigenfunctions of a GP prior—and prior precisions in a sparse finite model. It is well known that, among all orthogonal basis functions, eigenfunctions can provide the most compact representation. Unlike other sparse Bayesian finite models where the basis function has a fixed form, our eigenfunctions live in a reproducing kernel Hilbert space as a finite linear combination of kernel functions. We learn the dictionary elements—eigenfunctions—and the prior precisions over these elements as well as all the other hyperparameters from data by maximizing the model marginal likelihood. We explore computational linear algebra to simplify the gradient computation significantly. Our experimental results demonstrate improved predictive performance of EigenGP over alternative sparse GP methods as well as relevance vector machines.

**INTRODUCTION**

- **Gaussian Process for Regression**
  - Model: \( y(x) = f(x) + \epsilon \)
  - \( f(x) \sim \mathcal{N}(m(x), K) \) kernel: \( k(x_1, x_2) = k(x_1, x_2) \)
  - \( \epsilon \sim \mathcal{N}(0, \sigma^2) \)
  - e.g. \( k(x, x') = \sigma_0 \exp(-(x - x')^T \text{diag}((x - x')) \)
- **Log Model Evidence:** \( \ln p(y|\theta) = -\frac{1}{2} \ln |C_N| - \frac{1}{2} y^T C_N^{-1} y - \frac{1}{2} \ln(2\pi) \)
- **Predictive Distribution:** \( y^*|x^*, X, y \sim \mathcal{N}(k_{C_N} y, k(x^*, x^*) + k_{C_N}) \)
- **Limitation:** training takes \( O(N^3) \) time and \( O(N^2) \) space.
- **Nyström Method**
  - **Eigenfunctions:** \( \int k(x, x') \varphi(x)\varphi(x')dx = \lambda \varphi(x') \)
  - **Monte Carlo Approximation:** \( \sum_{i=1}^{M} \frac{1}{\text{diag}(w)} k(x_i, h_0) \) \( \approx \lambda \varphi(x) \)
  - **Approximate Eigenfunctions:** \( \varphi(x) = \sqrt{\frac{1}{\lambda}} k(x|h_0) \) \( \approx \varphi(x) \)
- **Model:** \( f(x) = \sum_{i=1}^{M} w_i \varphi_i(x) \)
- **Equivalent kernel:** \( \tilde{k}(x, x') = \sum_{i=1}^{M} w_i \varphi_i(x) \varphi_i(x') \)
- **Advantages:**
  - training takes \( O(NM^2) \) time and \( O(NM) \) space;
  - kernel is approximated by compact basis—eigenfunctions;
  - marginalization of \( \alpha \) prevents over-fitting.

**MODEL OF EIGENGP**

**RELATED WORK**

- **SSGP**
  - Use Fourier basis
- **EigenGP**
  - Use \( k(x, x_i) = \mathcal{N}(1, \mathcal{N}) \) as basis functions
- **SOR**
  - Fix \( \alpha \) to \( A \)
- **RVM**
- Fix \( \alpha \) to \( A \)
- **FITC**

**EXPERIMENTAL RESULTS**

Fig 2. Predictive mean and ± 2 standard derivation on synthetic data

Fig 3. Normalized Mean Square Error (NMSE) vs. time and Mean Negative Log Probability (MNLNP) vs. time on real data

**CONCLUSIONS**

We have presented a simple yet effective sparse Gaussian process method, and applied it to regression. Despite its similarity to the Nyström method, EigenGP can improve its prediction quality by several orders of magnitude.

**REFERENCES**