Improving the Gaussian Process Sparse Spectrum Approximation by Representing Uncertainty in Frequency Inputs

Yarin Gal, Richard Turner

University of Cambridge
(yg279, ret258)@cam.ac.uk

In short:

- Gaussian process sparse pseudo-input approximations cannot handle complex functions well.
- Alternatively, sparse spectrum approximations are known to over-fit.

We develop variational inference for the sparse spectrum approximation avoiding both issues, treating the covariance function as a random variable.

Background

What is a Gaussian process (GP)?

- A powerful tool in statistics, robust to over-fitting.
- Models distributions over functions.
- Supervised/unsupervised, regression/classification.
- Offers uncertainty estimates over the function values (in blue).

- Given training inputs \( X = \{ x_i \}_{i=1}^N \in \mathbb{R}^{N \times D} \) and outputs \( Y = \{ y_i \}_{i=1}^N \in \mathbb{R}^{N \times D} \), estimate a function \( y = f(x) \) that is likely to have generated \( Y \).

- We place a joint Gaussian distribution over all function values:

\[
  p(Y | X) = \mathcal{N}(m_k, \Sigma_k),
\]

with precision hyper-parameter \( r \) and covariance function \( K(X, X) \).

What is variational inference?

- Condition the model on a finite set of random variables \( \omega \).
- The predictive distribution for a new input point \( x^* \)

\[
  p(y^* | x^*, X, Y) = \int p(y^* | x^*, \omega) p(\omega | X, Y) \text{d}\omega,
\]

- The distribution \( p(\omega | X, Y) \) cannot be evaluated analytically — define an "easier" approximating variational distribution \( q(\omega) \).

- Minimise the Kullback–Leibler (KL) divergence: \( KL(q(\omega) \parallel p(\omega | X, Y)) \).

- Minimising the KL divergence is maximising log evidence lower bound,

\[
  \mathcal{L}_K = \int \log p(Y | X, \omega) \text{d}\omega - KL(q(\omega) \parallel p(\omega | X, Y))
\]

with respect to the variational parameters defining \( q(\omega) \).

Variational Sparse Spectrum Approximation

1. We are given the Fourier transform of the squared exponential covariance function

\[
  K(x - y) = \int \sigma^2 \mathcal{N}(0, I_N) \cos(2\pi w^T(x - y)) \text{d}w
\]

2. Introduce auxiliary variable \( b_k \)

\[
  K(x - y) = \int 2\sigma^2 \mathcal{N}(0, I_N) \text{Unif}(0, 2\pi) \cos(2\pi w^T(x - y) + b_k) \cos(2\pi w^T(y - x_k) + b_k) \text{d}wdb_k
\]

3. Approximate with Monte Carlo integration with \( K \) terms

\[
  \hat{K}(x - y) = \sum_{k=1}^K \frac{2\sigma^2}{K} \sum w^T \mathcal{N}(0, I_N) \text{Unif}(0, 2\pi) \cos(2\pi w^T(x - x_k) + b_k) \sqrt{2} \cos(2\pi w^T(y - x_k) + b_k)
\]

with \( w_k \sim \mathcal{N}(0, I_N) \), \( b_k \sim \text{Unif}(0, 2\pi) \), and variational parameters \( x_k \). This is a random covariance function.

4. The GP is re-parametrised as

\[
  w_k \sim \mathcal{N}(0, I_d), \quad b_k \sim \text{Unif}(0, 2\pi), \quad \omega = \{ w_k, b_k \}_{k=1}^K
\]

\[
  \Phi_{b,w}(\omega) = \sqrt{2\sigma^2} \cos(2\pi w^T(x_i - x_k) + b_k), \quad \Phi \in \mathbb{R}^{N \times K}
\]

\[
  p(Y | X, \omega) = \mathcal{N}(Y | 0, \Phi_{b,w}(\omega)^T + \tau^{-1} I_N)
\]

\[
  p(Y | X) = \int p(Y | X, \omega) p(\omega) \text{d}\omega
\]

5. Use variational distribution \( q(\omega) = \prod_{k=1}^K q(w_k)q(b_k) \) to approximate posterior \( p(\omega | X, Y) \):

\[
  q(\omega_k) = \mathcal{N}(\mu_k, \Sigma_k), \quad q(b_k) = \text{Unif}(\mu_k, \mu_k + \tau)
\]

with \( \Sigma_k \) diagonal.

6. We maximise the log evidence lower bound:

\[
  \mathcal{L}_{VSSGP} = \sum_{k=1}^K \left( -\frac{N}{2} \log(2\pi \tau) - \frac{1}{2} \tau \sum_k \mu_k^2 + \frac{1}{2} \sum_k \Sigma_k \right)
\]

\[
  + \frac{1}{2} \tau \sum_k E_{q^{(\omega)}}(\Phi \Sigma q^{(\omega)} \Phi^T) - KL(q^{(\omega)} \parallel p(\omega))
\]

with \( \Sigma = \mathcal{E}^{(\omega)}(\Phi \Sigma q^{(\omega)} \Phi^T) + \tau^{-1} I_N \). We can evaluate the KL and the expectations analytically using the identity

\[
  E_{q^{(\omega)}}(\cos(w^T x + b)) = e^{-ik\tau}\cos(k^2 \tau + b).
\]

7. This requires \( O(NK^2 + K^3) \) time complexity. Using parallel inference with \( T \) workers this reduces to \( O(NK^2/T + K^3) \).

Factorised Variational Sparse Spectrum Approximation

We often use large \( K \), for which the above is still slow. We can do better.

1. We introduce auxiliary random variables \( a_k \). The GP is re-parametrised as

\[
  a_k \sim \mathcal{N}(0, I_l), \quad A = [a_k]^T \in \mathbb{R}^{l \times D}
\]

\[
  p(Y | X, A, \omega) = \mathcal{N}(Y | 0, \Phi_{b,w}(\omega) A^T + \tau^{-1} I_N)
\]

\[
  p(Y | X) = \int p(Y | X, A^T) q(A) \text{d}A
\]

2. Use variational distribution \( q(\omega) = \prod_{k=1}^K q(w_k)q(b_k) \prod_{i=1}^d q(a_i) \) to approximate posterior \( p(A | X, Y) \):

\[
  q(a_i) = \mathcal{N}(\mu_i, \Sigma_i)
\]

with \( \Sigma_i \) diagonal.

3. This requires \( O(NK^2) \) time complexity — no matrix inverse. Using parallel inference with \( T \) workers this reduces to \( O(NK^2/T) \).

Stochastic Factorised Variational Sparse Spectrum Approximation

With stochastic optimisation, we can do even better.

\[
  \mathcal{L}_{SVSSGP} \approx \sum_{k=1}^K \sum_{d=1}^D \sum_{s=1}^S L_{w,s} - KL(q(A) \parallel p(A)) - KL(q(\omega) \parallel p(\omega))
\]

with random data subset \( S \) is an unbiased estimator of \( \mathcal{L}_{SVSSGP} \).

- Requires \( O(SK^2) \) time complexity with \( S < c \cdot N \) size of random subset, compared to \( O(SK^2 + K^3) \) of GP SVI using sparse pseudo-input approximation.

Approximation Properties

Model uncertainty

- Model uncertainty in interpolation and extrapolation tasks

Extrapolation on the Mauna Loa CO2 concentrations dataset (sum of two SE covariance functions, \( K = 50 \) inducing inputs):

- Interpolation on the reconstructed solar irradiance dataset (SE covariance function, \( K = 50 \) inducing inputs):

Imputation accuracy

- Imputation RMSE on train / test sets, for a speech signal of length 1K (\( K = 100 \)):

- Speed-Accuracy trade-off

- Train error, test error, and running time as functions of number of inducing points \( K \) for a speech signal of length 4K:

Code available at http://github.com/yaringal/VSSGP.