Reading Note 9 for Gaussian Process

Xiang Pan

October 16, 2021

1 Introduction

In short, the ability for a model to learn from data is determined by:

1. The support of the model: what solutions we think are a priori possible.
2. The inductive biases of the model: what solutions we think are a priori likely.

The **capacity (flexibility)** of a model $\mathcal{M}_i$ can be defined as the mutual information between the data $\boldsymbol{y}$ (at N locations X) and predictions made by the model $\boldsymbol{y}_*$ (at test locations X_*)

$$I_{i,N} = \sum_{y,y_*} p\left(\boldsymbol{y}, \boldsymbol{y}_* \mid \mathcal{M}_i\right) \log \frac{p\left(\boldsymbol{y}, \boldsymbol{y}_* \mid \mathcal{M}_i\right)}{p\left(\boldsymbol{y} \mid \mathcal{M}_i\right) p\left(\boldsymbol{y}_* \mid \mathcal{M}_i\right)} \tag{1}$$

$$I_{i,N} = p(\boldsymbol{y}) \int p\left(\boldsymbol{y}_* \mid \boldsymbol{y}\right) \log \frac{p\left(\boldsymbol{y}_* \mid \boldsymbol{y}\right)}{p\left(\boldsymbol{y}_*\right)} d\boldsymbol{y}_* \tag{2}$$

2 GP

We are ultimately more interested in – and have stronger intuitions about – the functions that model data than the weights w in a parametric model, and we can express those intuitions with a covariance kernel.

$$\begin{aligned} p\left(\boldsymbol{y}_* \mid \boldsymbol{y}\right) &= \int p\left(\boldsymbol{y}_* \mid f(x)\right) p(f(x) \mid \boldsymbol{y}) df(x) \\ p(f(x) \mid \boldsymbol{y}) &\propto p(\boldsymbol{y} \mid f(x)) p(f(x)) \end{aligned} \tag{3}$$

$$p\left(f_* \mid \boldsymbol{y}\right) = \int p\left(f_* \mid \boldsymbol{f}\right) p(\boldsymbol{f} \mid \boldsymbol{y}) d\boldsymbol{f} \tag{4}$$

Dependency: Hyperparameters → Parameters → Data

$$\log p(\boldsymbol{y} \mid \boldsymbol{\theta}, X) = \overbrace{-\frac{1}{2} \boldsymbol{y}^{\top} \left(K_{\boldsymbol{\theta}} + \sigma^2 I\right)^{-1} \boldsymbol{y}}^{\text{model fit}} - \overbrace{\frac{1}{2} \log \left|K_{\boldsymbol{\theta}} + \sigma^2 I\right|}^{\text{complexity penalty}} - \frac{N}{2} \log(2\pi) \tag{5}$$

Prediction

$$\boldsymbol{f}_* \mid X_*, X, \boldsymbol{y}, \boldsymbol{\theta} \sim \mathcal{N}\left(\overline{\boldsymbol{f}}_*, \operatorname{cov}\left(\boldsymbol{f}_*\right)\right) \tag{6}$$

$$p\left(\boldsymbol{f}_* \mid X_*, X, \boldsymbol{y}\right) = \int p\left(\boldsymbol{f}_* \mid X_*, X, \boldsymbol{y}, \boldsymbol{\theta}\right) p(\boldsymbol{\theta} \mid \boldsymbol{y}) d\boldsymbol{\theta} \tag{7}$$

$$p(\boldsymbol{\theta} \mid \boldsymbol{y}) \propto p(\boldsymbol{y} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) \tag{8}$$

3 Kernel

A kernel is **stationary** if it is invariant to translations of the inputs.

Covariance function	Expression	Stationary
Constant	a_0	Yes
Linear	$x \cdot x'$	No
Polynomial	$(x \cdot x' + a_0)^p$	No
Squared Exponential	$\exp\left(-\frac{\lvert x-x'\rvert^2}{2l^2}\right)$	Yes
Matérn	$\frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{\sqrt{2\nu}\lvert x-x'\rvert}{l}\right)^{\nu} K_\nu\left(\frac{\sqrt{2\nu}\lvert x-x'\rvert}{l}\right)$	Yes
Ornstein-Uhlenbeck	$\exp\left(-\frac{\lvert x-x'\rvert}{l}\right)$	Yes
Rational Quadratic	$\left(1+\frac{\lvert x-x'\rvert^2}{2\alpha l^2}\right)^{-\alpha}$	Yes
Periodic	$\exp\left(-\frac{2\sin^2\left(\frac{x-x'}{2}\right)}{l^2}\right)$	Yes
Gibbs	No	
Spectral Mixture	$\sum_{q=1}^{Q} w_q \prod_{p=1}^{P} \exp\left\{-2\pi^2 \left(x-x'\right)_p^2 v_{qp}\right\} \cos\left(2\pi \left(x-x'\right)_p^P \mu_{qp}\right)$	Yes

4 Mean Function

The mean function is also a powerful way to encode assumptions (inductive biases) into a Gaussian process model, the Gaussian process can leverage the assumptions of a parametric model through a mean function and also reflect the belief that the parametric form of that model will not be entirely accurate.

5 Feature Of GP

- Expressive Kernels
- Exact Efficient Inference: Efficiently determine the eigenvalues of a covariance matrix K
- Multi-Output Gaussian Processes
- Sampling Kernel Hyperparameters

References

[1] Andrew Gordon Wilson. *Covariance kernels for fast automatic pattern discovery and extrapolation with Gaussian processes.* PhD thesis, Citeseer, 2014.