Probabilistic Time Series Analysis: Lab 10

Tim Kunisky

November 14, 2018

Review of Gaussian Process Regression

A time series $f(x)$ is modeled as a GP with mean zero and covariance function $K(x_1, x_2; \boldsymbol{\theta})$. Then, observations $\boldsymbol{y} = (f(x_i))_{i=1}^{M}$ are distributed as

$$p(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{y} \mid \mathbf{0}, \boldsymbol{K}_{\boldsymbol{x},\boldsymbol{x}}),$$

where **the matrix $\boldsymbol{K}_{\boldsymbol{x},\boldsymbol{x}}$ has what entries?**

Review of Gaussian Process Regression

A time series $f(x)$ is modeled as a GP with mean zero and covariance function $K(x_1, x_2; \boldsymbol{\theta})$. Then, observations $\boldsymbol{y} = (f(x_i))_{i=1}^{M}$ are distributed as

$$p(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{y} \mid \mathbf{0}, \boldsymbol{K}_{\boldsymbol{x},\boldsymbol{x}}),$$

where **the matrix $\boldsymbol{K}_{\boldsymbol{x},\boldsymbol{x}}$ has what entries?**

If we observe data $(\boldsymbol{x}, \boldsymbol{y})$, then to do prediction for the value at a new point $y' = f(x')$, we condition:

$$\begin{aligned} p(y' \mid x', \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\theta}) \\ = \mathcal{N}(y' \mid \boldsymbol{k}_{x',\boldsymbol{x}}^{\top} (\boldsymbol{K}_{\boldsymbol{x},\boldsymbol{x}})^{-1} \boldsymbol{y}, \\ K(x', x'; \boldsymbol{\theta}) - \boldsymbol{k}_{x',\boldsymbol{x}}^{\top} \boldsymbol{K}_{\boldsymbol{x},\boldsymbol{x}}^{-1} \boldsymbol{k}_{x',\boldsymbol{x}}) \end{aligned}$$

where **the vector $\boldsymbol{k}_{x',\boldsymbol{x}}$ has what entries?**

FITC Part 1: “Independentifying” the Model

Let’s view the conditional formula for the distribution of $f(x')$ as a new model, where *the data* $\boldsymbol{x}$ *and* $\boldsymbol{y}$ *are now among the hyperparameters*: $f(x')$ are independent for different x', having distribution as before.

Let’s say we fix $\boldsymbol{x}' \in \mathbb{R}^N$. **What does the covariance of $(f(x_i'))_{i=1}^N$ look like?**

FITC Part 1: "Independentifying" the Model

Let's view the conditional formula for the distribution of $f(x')$ as a new model, where *the data $\boldsymbol{x}$ and $\boldsymbol{y}$ are now among the hyperparameters*: $f(x')$ are independent for different x', having distribution as before.

Let's say we fix $\boldsymbol{x}' \in \mathbb{R}^N$. **What does the covariance of $(f(x_i'))_{i=1}^N$ look like?** It is diagonal! If $\boldsymbol{y}' = f(\boldsymbol{x}')$, then the distribution is

$$p(\boldsymbol{y}' \mid \boldsymbol{x}', \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{y}' \mid \boldsymbol{K}_{\boldsymbol{x}',\boldsymbol{x}} \boldsymbol{K}_{\boldsymbol{x},\boldsymbol{x}}^{-1} \boldsymbol{y}, \Lambda),$$

$$\Lambda_{ij} = \begin{cases} K(x_i', x_i'; \boldsymbol{\theta}) - \boldsymbol{k}_{x_i',\boldsymbol{x}}^{\top} \boldsymbol{K}_{\boldsymbol{x},\boldsymbol{x}}^{-1} \boldsymbol{k}_{x_i',\boldsymbol{x}} & : \quad i = j \\ 0 & : \quad i \neq j. \end{cases}$$

FITC Part 2: A Change of Perspective

Now, let's write $\tilde{\boldsymbol{\theta}} = (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\theta})$. What we just derived is just a new gaussian process parametrized by $\tilde{\boldsymbol{\theta}}$!. Some remarks about it:

1. It is not centered.
2. It is not correlated in time (across different x').
3. We are free to choose $(\boldsymbol{x}, \boldsymbol{y})$ to be whatever we want—they are just parameters! We now call them *pseudo-data*.

FITC Part 2: A Change of Perspective

Now, let's write $\tilde{\boldsymbol{\theta}} = (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\theta})$. What we just derived is just a new gaussian process parametrized by $\tilde{\boldsymbol{\theta}}$!. Some remarks about it:

1. It is not centered.
2. It is not correlated in time (across different x').
3. We are free to choose $(\boldsymbol{x}, \boldsymbol{y})$ to be whatever we want—they are just parameters! We now call them *pseudo-data*.

Say we're given $\boldsymbol{x}' \in \mathbb{R}^N$ and observations $f(\boldsymbol{x}') = \boldsymbol{y}'$. We want

$$\operatorname{argmax}_{\tilde{\boldsymbol{\theta}}=(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\theta})} \log p(\boldsymbol{y}' \mid \boldsymbol{x}', \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\theta}),$$

which will factorize nicely by the independence we've built in, and so can be effectively done with gradient descent.

And, we can modulate the difficulty of this task by choosing M!