ガウス過程回帰 (Gaussian Process Regression)

目次

1 概要

ガウス過程回帰はカーネルを用いて未知の関数の回帰を行う手法の一つである。得られる出力がガウス過程にもとづいているという前提のもとで、関数を推定する。また、回帰によって得られるのは入力に対する出力ではなく、入力に対しての出力の確率分布である。

2 前提

あるパラメータ $\mathbf{w}$ をもち、入力 $\mathbf{x}$ をうけとる関数

$$f(\mathbf{x}, \mathbf{w})$$

にガウスノイズが加わった

$$f(\mathbf{x}, \mathbf{w}) = f(\mathbf{x}, \mathbf{w}) + \varepsilon$$

を考える。ここで $\varepsilon \sim N(0, \sigma^2)$ である。

n 個の入力 $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$ が与えられ、それぞれに対して $y_1, y_2, \cdots, y_n$ が得られたとする。このとき与えられた入力ベクトルからなる行列 X と出力からなるベクトル $\mathbf{y}$ を

$$X = \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{pmatrix}$$

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

と書く。入力 X とパラメータ $\mathbf{w}$ が与えられたときに、出力 $\mathbf{y}$ が得られる確率 (パラメータの尤度) を

$$p(\mathbf{y}|X,\mathbf{w})$$

と書き、これを求める。

関数 f がパラメータ $\mathbf{w}$ の各成分を係数とする関数の線形結合で表されるとすると、ある関数 $\phi_1, \phi_2, \cdots$ によって

$$f(\mathbf{x},\mathbf{w}) = \begin{pmatrix}\phi_1(\mathbf{x}) & \phi_2(\mathbf{x}) & \cdots\end{pmatrix}\mathbf{w} = \phi(\mathbf{x})\mathbf{w}$$

と書くことができ、$p(\mathbf{y}|X,\mathbf{w})$ は

$$\begin{aligned}
p(\mathbf{y}|X,\mathbf{x}) &= \Pi_{i=1}^{n} p(y_i|\phi(\mathbf{x}_i),\mathbf{w}) \\
&= \Pi_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - \phi(\mathbf{x}_i)^t\mathbf{w})^2}{2\sigma^2}\right) \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\left[(y_1 - \phi(\mathbf{x}_1)^t\mathbf{w})^2 + (y_2 - \phi(\mathbf{x}_2)^t\mathbf{w})^2 + \cdots + (y_n - \phi(\mathbf{x}_n)^t\mathbf{w})^2\right]\right) \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\left|\mathbf{y}^t - \Phi^t\mathbf{w}\right|^2\right) \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}(\mathbf{y}^t - \Phi^t\mathbf{w})^t(\sigma^2 I)^{-1}(\mathbf{y}^t - \Phi^t\mathbf{w})\right)
\end{aligned}$$

となる。ここで I は単位行列、Φ は

$$\Phi = \begin{pmatrix}\phi(\mathbf{x}_1) & \phi(\mathbf{x}_2) & \cdots & \phi(\mathbf{x}_n)\end{pmatrix}$$

であり、$o(\mathbf{y}|X,\mathbf{w})$ が平均が $\Phi^t\mathbf{w}$ で分散が $\sigma^2 I$ であるガウス分布であることが示された。

入力として X と対応する出力 $\mathbf{y}$ が与えられたときパラメータ $\mathbf{w}$ の事後確率つまり入力と出力の尤度は

$$\begin{aligned}
p(\mathbf{w}|X,\mathbf{y}) &= \frac{p(\mathbf{y}|X,\mathbf{w})p(\mathbf{w})}{p(\mathbf{y}|X)} \\
&\propto p(\mathbf{y}|X,\mathbf{w})p(\mathbf{w})
\end{aligned}$$

となる。であり、ここで $\mathbf{w}$ が標準正規分布に従うとすれば

$$p(\mathbf{w}) \propto \exp\left(-\frac{1}{2}\mathbf{w}^t\Sigma^{-1}\mathbf{w}\right)$$

であり、また

$$p(\mathbf{y}|X,\mathbf{x}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}(\mathbf{y}^t - \Phi^t\mathbf{w})^t(\sigma^2 I)^{-1}(\mathbf{y}^t - \Phi^t\mathbf{w})\right)$$

であるため

$$
\begin{aligned}
p(\mathbf{y}|X,\mathbf{w})p(\mathbf{w}) &\propto \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2}(\mathbf{y}^t-\Phi^t\mathbf{w})^t(\sigma^2 I)^{-1}(\mathbf{y}^t-\Phi^t\mathbf{w})\right)\exp\left(-\frac{1}{2}\mathbf{w}^t\Sigma^{-1}\mathbf{w}\right) \\
&\propto \exp\left(-\frac{1}{2}(\mathbf{y}^t-\Phi^t\mathbf{w})^t(\sigma^2 I)^{-1}(\mathbf{y}^t-\Phi^t\mathbf{w})\right)\exp\left(-\frac{1}{2}\mathbf{w}^t\Sigma^{-1}\mathbf{w}\right) \\
&= \exp\left(-\frac{1}{2}\left[(\mathbf{y}^t-\Phi^t\mathbf{w})^t(\sigma^2 I)^{-1}(\mathbf{y}^t-\Phi^t\mathbf{w})+\mathbf{w}^t\Sigma^{-1}\mathbf{w}\right]\right) \\
&= \exp\left(-\frac{1}{2}\left[\frac{1}{\sigma^2}(\mathbf{y}^t\mathbf{y}-\mathbf{y}^t\Phi^t\mathbf{w}-\mathbf{w}^t\Phi\mathbf{y}+\mathbf{w}^t\Phi\Phi^t\mathbf{w})+\mathbf{w}^t\Sigma^{-1}\mathbf{w}\right]\right) \\
&= \exp\left(-\frac{1}{2}\left[\mathbf{w}^t\left(\frac{1}{\sigma^2}\Phi\Phi^t+\Sigma^{-1}\right)\mathbf{w}-\frac{1}{\sigma^2}\left(\mathbf{y}^t\Phi^t\mathbf{w}+\mathbf{w}^t\Phi\mathbf{y}\right)+\frac{1}{\sigma^2}\mathbf{y}^t\mathbf{y}\right]\right)
\end{aligned}
$$

となる。ここで

$$
\begin{aligned}
A &= \frac{1}{\sigma^2}\Phi\Phi^t+\Sigma^{-1} \\
\mathbf{b} &= \frac{1}{\sigma^2}\Phi\mathbf{y} \\
2c &= \frac{1}{\sigma^2}\mathbf{y}^t\mathbf{y}
\end{aligned}
$$

とすると

$$
\begin{aligned}
p(\mathbf{y}|X,\mathbf{w})p(\mathbf{w}) &\propto \exp\left(-\frac{1}{2}\left[\mathbf{w}^t\left(\frac{1}{\sigma^2}\Phi\Phi^t+\Sigma^{-1}\right)\mathbf{w}-\frac{1}{\sigma^2}\left(\mathbf{y}^t\Phi^t\mathbf{w}+\mathbf{w}^t\Phi\mathbf{y}\right)+\frac{1}{\sigma^2}\mathbf{y}^t\mathbf{y}\right]\right) \\
&= \exp\left(-\frac{1}{2}\left[\mathbf{w}^tA\mathbf{w}-\left(\mathbf{b}^t\mathbf{w}+\mathbf{w}^t\mathbf{b}\right)+2c\right]\right) \\
&= \exp\left(-\frac{1}{2}\left[\left(\mathbf{w}^tA\mathbf{w}-\mathbf{b}^t\mathbf{w}-\mathbf{w}^t\mathbf{b}\right)+\mathbf{b}^tA^{-1}\mathbf{b}-\mathbf{b}^tA^{-1}\mathbf{b}+2c\right]\right) \\
&= \exp\left(-\frac{1}{2}\left[\left(\mathbf{w}^tA-\mathbf{b}^t\right)A\left(\mathbf{w}-A^{-1}\mathbf{b}\right)+2c-\mathbf{b}^tA^{-1}\mathbf{b}\right]\right) \\
&= \exp\left(-\frac{1}{2}\left(\mathbf{w}^tA-\mathbf{b}^t\right)A\left(\mathbf{w}-A^{-1}\mathbf{b}\right)\right)\exp\left(-\frac{1}{2}\left[2c-\mathbf{b}^tA^{-1}\mathbf{b}\right]\right) \\
&\propto \exp\left(-\frac{1}{2}\left(\mathbf{w}-A^{-1}\mathbf{b}\right)^tA^{-1^{-1}}\left(\mathbf{w}-A^{-1}\mathbf{b}\right)\right)
\end{aligned}
$$

と書くことができる。よって

$$
p(\mathbf{w}|X,\mathbf{y}) \propto p(\mathbf{y}|X,\mathbf{w})p(\mathbf{w}) \propto \exp\left(-\frac{1}{2}\left(\mathbf{w}-A^{-1}\mathbf{b}\right)^tA^{-1^{-1}}\left(\mathbf{w}-A^{-1}\mathbf{b}\right)\right)
$$

となり

$$
p(\mathbf{w}|X,\mathbf{y}) \sim N(A^{-1}\mathbf{b}, A^{-1})
$$

であることがしめされた。これは、複数の入力と対応する出力が与えられると、それらを用いてパラメータの分布を計算できるということである。

参考文献

[1] “ガウス過程回帰の導出（GPR：Gaussian Process Regression）”，https://www.slideshare.net/KenjiUrai/explanation-of-gpr，参照 May.31,2020.

[2] “Gaussian process - Wikipedia”，https://en.wikipedia.org/wiki/Gaussian_process#Definition，参照 May.31,2020.