2019年8月31日 土曜日

ガウス過程回帰

ガウス過程回帰概要

動径基底関数$\phi_h(x) = exp(-\frac{(x-\mu_h)^2}{\sigma^2})$を無限個用意して重みつけて足せば、任意の関数が表せるのでは？というモチベーションのもと、動径基底関数回帰を考えたい。しかし、単純に回帰するとxの次元が増えるに当たって、パラメータの次元が爆発する(10次元でもう17兆くらいになる)。つまり、次元の呪いが発生するためしんどい。でもこれが出来たら大抵の非線形関数は表現できて、しかも各項の共分散も考えられると嬉しいよねみたいな気がする。非線形でとにかく近似したいというのであれば、分析の目的は予測であり、パラメータの値には興味がないので、正直次元の呪いというのも馬鹿馬鹿しい。そこでどうにかうまいことパラメータを消してこの方法を実施し、予測を行いたい。

事前分布を置いてみる

$w \sim N(0, \alpha I_n)$と、重みパラメータのベクトルの事前分布として、期待値０の多変量正規分布を仮定する。ちなみに、今は次のような非線形回帰を考えている。

$$y = \Phi(X)w + \epsilon$$

ここで、ϵは、各要素独立に分散σ^2の正規分布に従うとする。この事前分布の仮定が分析にどういう影響を及ぼすのかは考えないといけないが、それは後述するとして、とりあえず重みパラメータを消したい。まず、$\Phi(X)w$は、Xを条件付けると、期待値0,分散$\alpha\Phi(X)\Phi(X)^T$の多変量正規分布に従うので、もうこの時点でありがたいことにwは消える。あとは、正規分布の再生性から、yの分布は$N(0, \alpha\Phi(X)\Phi(X)^T + \sigma^2 I_n)$である。これの尤度最大化を行うように残りのパラメータを推定すると良いというのがガウス過程回帰の基本的なアイデアとなる。当然、内部の積はカーネルの形になっているので、カーネル法が利用できる。よって、適当なカーネルKを用いて、次のように一般化できる。

$$N(0, \alpha K + \sigma^2 I_n)$$

事前分布の仮定

事前分布については、共分散構造に注目すると、xの値が近い値をとっていれば、yも近い値を取るような構造になっていることが確認できる。これが事前分布による仮定で、そこまできつい仮定には見えない。

新しいデータ点に対する推定

新しいデータ点について考えたい場合、多変量正規分布なので、多変量正規分布の共分散行列のブロック化公式から期待値と分散を導出することになる。

パラメータの最適化

$K_\theta = \alpha K + \sigma^2 I_n$とおくと、yの分布は$N(0, K_\theta)$なので、尤度は

$$p(\theta; y) = \frac{1}{(2\pi)^{\frac{n}{2}}} \frac{1}{|K_\theta|^{\frac{1}{2}}} \exp(-\frac{1}{2} y^T K_\theta^{-1} y)$$

さらに対数尤度をとれば、

$$log(p(\theta; y)) = log(|K_\theta|) - t^T K_\theta^{-1} y + constant$$

対数尤度最大化問題を適当な数値計算で解けば良い。事前分布を仮定しているので、ベイズだと思えば、MCMCを使うこともできる。普通の最適化手法だと局所解にハマるので、素直にMCMC回しといたほうが良さげだが、この場合、高次元パラメータだとしんどいので、どちらもできるようにしておくのがベストか。
例はGPR(Stan).R,GPR.py
Python環境のGPyが一番便利そう。