

《概率论与随机过程》第二次 Project

无 42 陈佳榕 2014011050

2016 年 1 月 27 日 星期五

目录

| | |
|--------------------------------------|---|
| 1. 摘要 (Abstract) | 2 |
| 2. 介绍 (Introduction) | 2 |
| 3. 模型 (Model) | 2 |
| 4. 基本理论和方法 (Basic Theory and Method) | 3 |
| 5. 方法 (Method) | 3 |
| 6. 方法分析 (Analysis of Method) | 4 |
| 7. 算法 (Algorithm) | 4 |
| 8. 算法分析 (Analysis of Algorithm) | 4 |
| 9. 数值结果 (Numerical Result) | 5 |
| 10. 讨论 (Discussion) | 6 |
| 11. 结论 (Conclusion) | 7 |
| 12. 致谢 (Acknowledgement) | 8 |
| 13. 参考文献 (Reference) | 8 |
| 14. 附录 (Appendix) | 8 |

1. 摘要 (Abstract)

本文首先使用高斯过程回归 (GPR)，通过贪婪法对若干个基本核函数进行加、乘操作得到足以描述数据内在规律的复合核函数，并利用训练数据以训练合适的高斯过程回归器，然后再用训练得到的高斯过程回归器对测试数据进行预测，与实际观测值进行比较，得到均方误差 (MSE) 在 0.35 以下的较好预测。

然后同样使用高斯过程回归的方法，通过选择一个合适的核函数，在此基础上再选择合适的似然函数和贝叶斯推断近似方法，对 F16 飞机的副翼控制问题进行预测，与实际数据比较得到均方误差 (MSE) 在 0.027 左右的较好预测。

2. 介绍 (Introduction)

回归分析研究变量与变量之间的关系，是统计学、信号处理、机器学习等领域的基础研究问题之一，而高斯过程回归 (GPR) 作为其中一者也在近来受到了广泛的关注。

在《Gaussian Processes for Machine Learning》一书的第五章中作者 C. E. Rasmussen 和 C. K. I. Williams 对高斯过程回归器作了一个简单的介绍，给出高斯回归的几种近似算法，但是缺乏对不同近似方法的相对优点的理解；

《Structure Discovery in Nonparametric Regression through Compositional Kernel Search》的作者 Duvenaud D, Lloyd J R, Grosse R 等人提出一种贪婪搜索内核结构的算法，并给出了常用的四种基本核函数；Lloyd J R, Duvenaud D, Grosse R 等人则在《Automatic construction and natural-language description of nonparametric regression models》中介绍了一种自动贝叶斯协方差发现系统及其如何自动生成搜索过程找到的模型的自然语言描述，实现自动化统计建模的目标，且亦给出了五种基本核函数；《The Kernel Cookbook: Advice on Covariance functions》一文作者 David Duvenaud 介绍了五种基本核函数，接着介绍对基本核函数进行的加、乘操作以及一些其他类型的结构；David Kristjansson Duvenaud 在《Automatic Model Construction with Gaussian Processes》中介绍了一种基于高斯过程的自动构造、可视化和描述一个大规模模型并可用于在时间序列、地质构造和物理动力学等领域进行预测以及寻找结构的方法，同时也可以捕获很多类型的统计结构。

根据以上文献，我们实现了贪婪法自动搜索最优复合核函数，辅以合适的似然函数和贝叶斯推断方法，并在测试数据上进行了测试，结果表明预测的性能较好。

3. 模型 (Model)

设高斯过程回归的回归函数为 $f(x)$ ，高斯过程回归将 $\{f(x) | x \in S\}$ 建模为高斯过程 $\mathcal{N}\{m(x), K(x, x')\}$ ，其中 $m(x) = E\{f(x)\}$ ， $K(x, x') = cov(f(x), f(x'))$ ，由训练集数据给出在某些点处的带噪声观测值，利用 $f(x)$ 为高斯过程及观测噪声 e 的 pdf 确定出似然函数，由此可以估计其他任何点处的后验分布。

4. 基本理论和方法 (Basic Theory and Method)

本文利用的 GPR 方法实际上不必直接给定回归函数形式，而是只需确定核函数的函数形式。本文常用的几个核函数如下：

(1)各向同性的平方指数核函数 $K_{SE}(x, x') = \sigma^2 \exp(-\frac{(x-x')^2}{2l^2})$

(2)线性核函数 $K_{LIN}(x, x') = \sigma_b^2 + \sigma_v^2(x-l)(x'-l)$

(3)各向同性有理核函数 $K_{RQ}(x, x') = \sigma^2(1 + \frac{(x-x')^2}{2\alpha l^2})^{-\alpha}$

(4)周期核函数 $K_{PER}(x, x') = \sigma^2 \exp(-\frac{2\sin^2\pi(x-x')}{2\alpha l^2})$

(5)常数核函数 $K_{Const}(x, x') = \sigma^2$

5. 方法 (Method)

总的思路如下：

首先，需要从 N 个选定的基本核函数中选择一个作为核函数，选定的标准是根据所选的这个核函数训练得到的高斯过程回归器再用于测试数据得到的 MSE 最小。

然后再用选定的这个核函数和剩下的 N-1 个基本核函数进行加、乘组合，也就是一共有 2N-2 种组合，根据同样的选定标准选择某种组合作为新的核函数。

接下来在上面得到的核函数的基础上，再和剩下的 N-2 个基本核函数进行加、乘组合，这样又有 2N-4 种组合，同样根据 MSE 最小原则选取其中的一种组合作为新的核函数。

这样一直重复进行，最终可以得到一个由 N 个核函数用加、乘操作组合得到的一个复合核函数。

接着利用已经得到的复合核函数，分别采用不同的似然函数和贝叶斯推断近似方法组合，测试在其上面的 MSE 值，最终选择一个 MSE 最小的组合作为最终的似然函数和贝叶斯推断方法。

最后需要用以上已经确定的核函数、似然函数的形式和贝叶斯推断近似方法以及各种参数，在测试数据上进行测试以得到 MSE。

当然，对于特定的问题有可能简化以上步骤。比如在本次大作业中的第一题就可以省去最后分别采用不同的似然函数和贝叶斯推断方法组合的步骤；第二题则是只进行一次选择核函数操作，而不用进行下面的加、乘组合操作，因为一个核函数已经足够描述数据的内在规律。

6. 方法分析 (Analysis of Method)

方法的预测性能的好坏主要受制于训练时选取的核函数、似然函数形式以及贝叶斯推断方法，这些可以通过最小化 MSE 选择出最适于具体问题的组合。而接下来要提高预测的准确度即降低 MSE 值，最主要的因素是训练阶段优化参数时的最大优化次数。为了提高准确度需要相应地提高最大优化次数，这相应地需要更多的运算时间。所以需要在最大优化次数和运算时间之中取得一个合适的折中。

7. 算法 (Algorithm)

选择 N 个基本的核函数作为核函数集合

for n = 1 : N

 选择第 n 个基本核函数作为当前核函数

 进行优化以得到合适的参数，然后进行训练

 用已训练好的高斯过程回归器来计算 MSE

end

选择 MSE 最小的作为当前核函数并从核函数集合中将其删去

While 核函数集合非空

 依次从核函数集合中选出一个核函数与当前核函数分别进行加、乘操作

 进行参数优化，训练，然后计算 MSE

 选出 MSE 最小的作为当前核函数，并将选中的核函数从核函数集合除去

end

利用以上得到的复合核函数并改用其他的似然函数和贝叶斯推断近似方法组合
重新通过最小化似然来优化参数

重新计算 MSE

最后选择一个 MSE 最小的作为最终的似然函数和贝叶斯推断近似方法

对于具体问题，以上算法可能会略有取舍，比如在本次大作业中的第一题中就不需要进行最后一部分的改用其他的似然函数和贝叶斯推断近似方法组合的步骤；而在第二题中则不需要进行第二部分的重复选择核函数进行加、乘组合操作的部分。总的来说这里提出的算法是具有普遍性的，可以适用于较多场合并且可以根据实际需要进行取舍。

8. 算法分析 (Analysis of Algorithm)

算法的复杂度主要在于训练的时候进行参数的优化，最大优化次数限制了算法的复杂度。虽然减少最大优化次数可以减少时间和空间复杂度，但是同时会减少预测结果的准确度，这是两个无法调和的矛盾，需要在实际中进行适当的折中

以达到较好的效果。

在搜索最优复合核函数的时候假设基本核函数集合有 N 个基本核函数，那么实际上就需要 $N+2(N-1)+2(N-2)+\cdots+2$ 次参数优化，即 $O(N^2)$ 次参数优化，这也会带来较大的时间复杂度。

而空间复杂度则主要由训练数据的规模决定，实际上为了减少空间复杂度，并没有将所有训练数据都用于训练，而只是选择足以让我们判断出所需各种参数的尽可能小的训练数据规模。

9. 数值结果（Numerical Result）

仿真环境采用的 CPU 为 Intel Core i5-4210，操作系统为 Windows10，MATLAB 版本为 R2016a。

(a)question1

训练得到的核函数为 $[(SE+LIN) \times Const + RQ] \times PER$ ；

因为本身利用的就是高斯过程回归，所以似然函数选择了高斯似然，贝叶斯推断近似方法选择了高斯近似；

超参数如下：

| | |
|-----|---|
| cov | [5.415588;4.185325;4.345935;-0.160610;5.566283;2.862116; -13.169638;2.838082;-0.000369;2.701506] |
| lik | -1.493367 |

而根据以上参数，得到的 MSE 为 0.341308。

(b) question2

实际中考虑到机器的计算能力的限制，训练数据的时候没有用到所有的数据，而是截取了其中的一部分（本题中选择训练数据的前 1000 行）。

首先需要选择一个核函数，依次对所选四个基本核函数进行训练然后在测试数据上计算 MSE，得到的结果如下表所示：

| 基本核函数 | MSE |
|-------|----------|
| SE | 0.031554 |
| LIN | 0.034698 |

| | |
|--------------|----------|
| RQ | 0.031666 |
| Const | 0.162503 |

因此选择 MSE 值最小的 SE 作为基本核函数。

然后进行不同的似然函数和贝叶斯推断近似方法的组合，得到不同的 MSE 值如下表所示（最大优化次数为 300，选择训练数据的前 500 行）：

| 似然函数 \ 贝叶斯近似 | 高斯近似 | Laplace 近似 | 变分贝叶斯 | MCMC |
|-------------------|----------|------------|----------|----------|
| likGauss | 0.031777 | 0.031751 | 0.031785 | 0.926849 |
| likLaplace | | 0.926878 | 0.926841 | 0.926850 |
| likSech2 | | 0.926968 | 0.032778 | 0.926850 |

因此选择高斯似然作为似然函数以及 Laplace 近似作为贝叶斯近似。

基于以上的选择，选取更大的训练集（实际上选择了训练数据的前 8000 行），进行参数优化，结果如下：

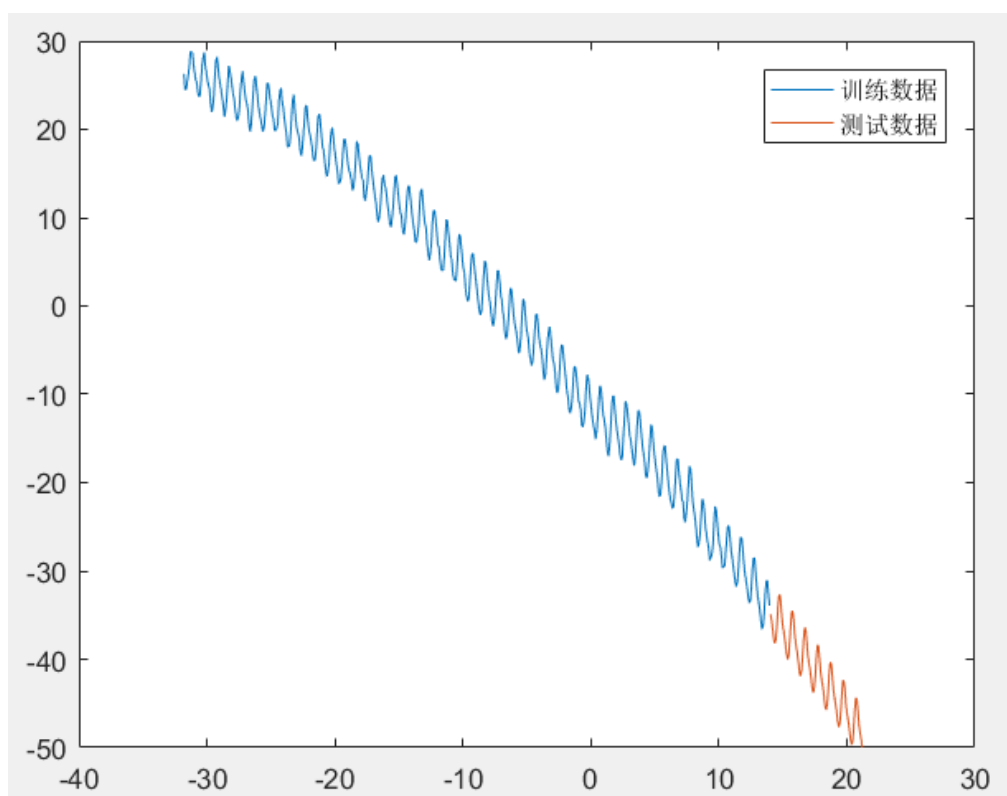
| | |
|-----|----------------------|
| cov | [7.741201; 9.171663] |
| lik | -1.779549 |

在此基础上在测试集上进行预测，得到的 MSE 为 0.027334。

10. 讨论（Discussion）

(a)question1

从结果中可以看出，利用高斯过程回归可以较好地进行预测，说明前面进行的训练得到了一个性能较好的高斯过程回归器。这也可以较为直观地从下图看出，红色部分为预测的数据，可以看到，预测得到的数据较好地符合了原先数据的趋势，并且预测到了数据的走向趋势。



(b)question2

从结果中可以看到，利用高斯过程回归可以较好地在飞机副翼控制问题上进行预测且可以得到较好的预测结果，说明训练阶段得到的参数是比较合适的。同时也可以看到通过更改不同的似然函数和贝叶斯推断近似方法，从中选择最合适的，最终更加精准地在测试数据上进行预测。

11. 结论（Conclusion）

通过高斯过程回归可以根据一定规模的训练数据对高斯过程回归器进行训练得到一个性能较好的高斯过程回归器，经过优化参数、选择合适的核函数和似然函数以及贝叶斯推断近似方法，可以较为精确地对数据进行预测。

本文分别在两个训练数据集上进行了训练，通过分别选择合适的核函数、似然函数和贝叶斯推断方法，建立相应的模型，然后在测试数据上进行 MSE 的计算，发现得到的 MSE 值都较小，说明了预测性能较为优越。

另外在实验中发现算法存在计算速度不够高的问题，这也是一个具有很大突破空间的地方，值得对其进行更进一步的研究，以其改进算法得到更好的性能。

12. 致谢 (Acknowledgement)

最后，要感谢在本次大作业中及时答疑解惑的欧智坚老师、戴音培助教和陈杭助教，以及为我对算法的理解提供了帮助的张元鑫同学。

13. 参考文献 (Reference)

- [1] C. E. Rasmussen, C. K. I. Williams, "Gaussian Processes for Machine Learning", MIT press, 2006.
- [2] Duvenaud D, Lloyd J R, Grosse R, et al. Structure Discovery in Nonparametric Regression through Compositional Kernel Search[J].Creative Commons Attribution-Noncommercial-Share Alike, 2013.
- [3] Lloyd J R, Duvenaud D, Grosse R, et al. Automatic construction and natural-language description of nonparametric regression models[C]// Twenty-Eighth AAAI Conference on Artificial Intelligence.AAAI Press, 2014.
- [4] The Kernel Cookbook:Advice on Covariance functions. <http://www.cs.toronto.edu/~duvenaud/cookbook/index.html>
- [5] David Kristjanson Duvenaud.Automatic Model Construction with Gaussian Processes.PhD Thesis. University of Cambridge. 2014

14. 附录 (Appendix)

最后，说明本文用到的代码文件所有用到的代码均放置于 code 文件夹中，第一题的代码为 exp1_1.m 和 exp1_2.m，exp1_1.m 进行了贪婪法的实现，寻找合适的复合核函数；exp1_2.m 利用寻找得到的复合核函数，进行参数的优化并计算 MSE。第二题的代码为 exp2_1.m 和 exp2_2.m，exp2_1.m 进行核函数的选择，exp2_2.m 在以上得到的核函数的基础上进行似然函数和贝叶斯推断方法的选择，然后进行参数的优化并计算 MSE。最后 data_read_and_MSE.m 是最终版本的程序，已经针对两个问题分别确定了核函数、似然函数的形式和贝叶斯推断近似方法。

其他的代码文件均为 Carl Edward Rasmussen & Hannes Nickisch 所提供的代码，可以从 <http://www.gaussianprocess.org/gpml/code/matlab/doc/> 上得到所需代码。