

# 基于高斯过程的回归分析

yfreedomliTHU THU EE

**摘要：**本文实现了基础部分的四种回归方法：A. 标准线性回归模型的最小二乘求解；B. 标准线性回归模型的概率求解；C. 引入多项式基函数构成广义线性回归模型；D. 以平方指数函数为核函数的高斯过程回归分析。本文也对对各种回归方法和实验结果进行了分析。为了提高 GPR 的回归精度，本文对不同的核函数（SE, PER, LIN, RQ 等）进行了仿真及结果分析，最终采用平方指数函数（SE）作为核函数。此外，本文还实现了两种 GPR 的变种方法，Additive GPR 和 Warped GPR, 并对其进行了分析。为了进一步提高预测精度，本文对给定的 8 门课程成绩进行相关性分析，发现去除部分相关性弱的课程能得到更好的回归效果。最终，通过选取的相关性更强的课程成绩和合适的模型（GPR, Additive GPR 和贝叶斯模型），得到了效果最好的回归解。

**关键词** 回归分析, GPR, 核函数, SE, PER, RQ, LIN, Additive GPR, Warped GPR

## Gaussian Process Regression(GPR)

Yinfeng Li

(Electronic Engineering, Tsinghua University)

**Abstract** In this paper, I realized the four basic regression methods: A. Standard linear regression model of least squares solution; B. Standard linear regression model of the probability of solution; C. Introduced polynomial basis functions to form generalized linear regression model ;. D. The square exponent Gaussian Process Regression Analysis with Kernel Function. This article also analyzed the various regression methods and experimental results. In order to improve the regression accuracy of GPR, this paper simulated and analyzed the different kernel functions (SE, PER, LIN, RQ, etc.) and finally uses the square exponential function (SE) as the kernel function. In addition, this paper also implements and analyzes two variants of GPR, Additive GPR and Warped GPR. In order to further improve the prediction accuracy, this paper conducted a correlation analysis of the grade of the given eight courses, and found that removing some of the less relevant courses can get a better regression. Finally, the best-performing regression solution was obtained by selecting the more relevant course grades and using the Suitable models (GPR, Additive GPR and Bayesian models).

**Key words :** Regression Analysis , GPR , Kernel Functions , SE , PER , RQ ,LIN,Additive GPR,Warped GPR

## 1 引言(Introduction)

### 基于高斯过程的回归分析 (GPR)

回归分析是统计学、信号处理、机器学习等多领域中的基础研究问题之一。回归分析研究的是变量与变量间的关系。记其中一个变量称为自变量  $x \in S \subset \mathbb{R}^d$ , 另一个变量称为因变量  $y \in \mathbb{R}$ , 假设两者存在如下的关系

$$y = f(x) + e$$

其中,  $e$  为表示误差的随机变量,  $f(x)$  称为回归函数 (或预测函数、拟合函数)。给定一组观测  $D = \{(x_i, y_i) \mid i = 1, 2, \dots, n\}$ ,  $D$  又被称为训练数据, 回归分析希望能就此找出在某个准则下最好的回归函数  $f^*$ 。传统的回归分析包括了两个层次的问题, 一是确定合适的回归函数形式; 二是在给定回归函数形式下, 依据训练数据求出具体的回归函数。一类广泛应用的回归模型是广义线性模型, 它假定回归函数为基函数的线性组合, 即:

$$f(x) = \sum_{i=1}^M w_i \phi_i(x)$$

其中  $\phi_i(x)$  称为基函数, 利用简单的基函数构成较复杂的回归函数形式, 比如:  $\phi_1(x) = x_i$  可以构成线性回归函数;  $\phi_k(x) = x_i^k (k=1, \dots, N)$  可以构成多项式回归函数;  $\phi_i(x) = \cos(w_i^T x)$  可以形成三角函数等。在基于广义线性模型的回归分析中, 首先要利用训练数据集确定基函数中的参数以及基函数系数, 从而求出具体的回归函数 (求解过程通常要处理基于观测数据与预测数据间差别最小导出的优化问题)。

获取具体回归函数以后, 可以以此预测未知自变量  $x$  处的因变量值  $y$ 。因此, 可以在一组有别于训练数据的测试数据  $T = \{(x_i^*, y_i^*) \mid i=1, 2, \dots, m\}$  上, 通过计算预测值与观测值间的差异来评价回归函数的好坏。本文采用均方误差 (Mean Squared Error, MSE) 来衡量, 即:

$$MSE = \frac{1}{m} \sum_{i=1}^m (y_{*,i} - y_i^*)^2$$

本文实现了以下方法: A. 标准线性回归模型的最小二乘求解; B. 标准线性回归模型的概率求解; C. 引入多项式基函数构成广义线性回归模型; D. 以平方指数函数为核函数的高斯过程回归分析。在此基础上, 本文分析了不同的核函数对高斯过程回归的影响, 并对给定的课程进行了相关性分析, 选取部分课程进行回归分析, 一定程度上减小了 MSE, 提高了回归精度。

本文使用 MH 算法对二维高斯分布进行采样, 由采样结果对二维高斯分布的相关系数进行估计。就问题本文提出了二维均匀分布和与原分布类似的二维正态分布, 并针对不同的分布对结果的影响进行了分析。

近年来, 除了标准的高斯回归模型, Additive Gaussian Process 模型和 Warped Gaussian Process 模型也越来越流行, 本文也实现了这两种回归方法。

## 2 基于高斯过程的回归分析(Gaussian Process Regression(GPR))

### 2.1 基本原理 (方法阐述)

对于回归进行一般性的建模如下:

$$y = f(x) + e \quad (1)$$

其中,  $e$  为表示误差的随机变量,  $f(x)$  称为回归函数 (或预测函数、拟合函数),  $x$   $D$  维自变量。回归需要完成的工作是基于给定的自变量, 选取合适的回归函数, 使得预测值  $y$  尽可能接近实际值。

### 2.1.1 标准线性回归模型的最小二乘求解

$$f(x) = \mathbf{x}^T \mathbf{w} \quad (2)$$

对于标准线性回归，回归模型可以写成：

$$\mathbf{y} = \mathbf{x}^T \mathbf{w} \quad (3)$$

而协方差计算公式，有：

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n |y_i - f(x_i)|^2 \quad (4)$$

将(3)带入(4)，MSE对 $\mathbf{w}$ 求导，令导数为0，即可求得：

$$\mathbf{w} = (\mathbf{x} \mathbf{x}^T)^{-1} (\mathbf{x} \mathbf{y}) \quad (5)$$

由 $\mathbf{w}$ 和给定的代入(3)即可得到 $\mathbf{y}$ 的预测值。

### 2.1.2 标准线性回归模型的概率求解（贝叶斯回归）

对于贝叶斯回归模型， $\mathbf{w}$ 满足高斯分布：

$$\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \Sigma_p). \quad (6)$$

根据贝叶斯准则，有：

$$p(\mathbf{w}|\mathbf{y}, X) = \frac{p(\mathbf{y}|X, \mathbf{w})p(\mathbf{w})}{p(\mathbf{y}|X)}, \quad (7)$$

而

$$\begin{aligned} p(\mathbf{y}|X, \mathbf{w}) &= \prod_{i=1}^n p(y_i|\mathbf{x}_i, \mathbf{w}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left(-\frac{(y_i - \mathbf{x}_i^T \mathbf{w})^2}{2\sigma_n^2}\right) \\ &= \frac{1}{(2\pi\sigma_n^2)^{n/2}} \exp\left(-\frac{1}{2\sigma_n^2} |\mathbf{y} - X^T \mathbf{w}|^2\right) = \mathcal{N}(X^T \mathbf{w}, \sigma_n^2 I). \end{aligned} \quad (8)$$

可以进一步求得 $\mathbf{w}$ 的分布：

$$p(\mathbf{w}|X, \mathbf{y}) \sim \mathcal{N}(\bar{\mathbf{w}} = \frac{1}{\sigma_n^2} A^{-1} X \mathbf{y}, A^{-1}), \quad (9)$$

$$A = \sigma_n^{-2} X X^T + \Sigma_p^{-1}$$

由 $\mathbf{w}$ 分布的均值和已知的 $\mathbf{x}$ 代入(3)即可得到预测值 $\mathbf{y}$ 。贝叶斯模型采用了最大化后验分布准则（MAP），最大化给定训练数据的利用率，将各个数据均考虑在内。

### 2.1.3 引入基函数构成广义线性回归模型（基函数为多项式）

引入基函数的广义线性回归解决的实际是非线性回归问题。首先需要对给定的自变量 $\mathbf{x}$ 用基函数进行处理，得到基函数的值，然后对基函数的值进行线性回归。以多项式基函数为例：

$$\phi(x) = (1, x, x^2, x^3, \dots)^T \quad (10)$$

模型即可表示为：

$$f(\mathbf{x}) = \phi(\mathbf{x})^T \mathbf{w}. \quad (11)$$

对基函数采用贝叶斯模型，可以求得预测值 $\mathbf{y}$ 的分布：

$$f_*|\mathbf{x}_*, X, \mathbf{y} \sim \mathcal{N}\left(\frac{1}{\sigma_n^2} \phi(\mathbf{x}_*)^T A^{-1} \Phi \mathbf{y}, \phi(\mathbf{x}_*)^T A^{-1} \phi(\mathbf{x}_*)\right) \quad (12)$$

$$A = \sigma_n^{-2} \Phi \Phi^T + \Sigma_p^{-1}$$

根据分布即可得到预测值的均值。

### 2.1.4 高斯过程回归模型

不妨先假定  $f(\mathbf{x})$  的均值如下：

$$m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})], \quad (13)$$

可以得到相关函数（核函数）：

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))], \quad (14)$$

所以高斯过程回归模型可以表示为：

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \quad (15)$$

对于一般的高斯回归，先验噪声变为：

$$\text{cov}(y_p, y_q) = k(\mathbf{x}_p, \mathbf{x}_q) + \sigma_n^2 \delta_{pq} \quad (16)$$

$$\text{cov}(\mathbf{y}) = K(X, X) + \sigma_n^2 I$$

进一步可以写出函数值和预测值的联合分布：

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K(X, X) + \sigma_n^2 I & K(X, X_*) \\ K(X_*, X) & K(X_*, X_*) \end{bmatrix}\right) \quad (17)$$

因此，可以得到预测值的高斯分布：

$$\mathbf{f}_* | X, \mathbf{y}, X_* \sim \mathcal{N}(\bar{\mathbf{f}}_*, \text{cov}(\mathbf{f}_*))$$

其中

$$\begin{aligned} \bar{\mathbf{f}}_* &\triangleq \mathbb{E}[\mathbf{f}_* | X, \mathbf{y}, X_*] = K(X_*, X)[K(X, X) + \sigma_n^2 I]^{-1} \mathbf{y}, \\ \text{cov}(\mathbf{f}_*) &= K(X_*, X_*) - K(X_*, X)[K(X, X) + \sigma_n^2 I]^{-1} K(X, X_*) \end{aligned} \quad (18)$$

### 2.1.5 核函数的选取

核函数的选取对于高斯过程回归的结果有一定的影响，选择合适的核函数能够提升非线性处理的能力，得到较好的预测效果。文献[3][4]给出了SE, LIN, PER, RQ等可供选择的核函数，其具体形式如下：

$$\begin{aligned} k_{\text{SE}}(x, x') &= \sigma^2 \exp\left(-\frac{(x-x')^2}{2\ell^2}\right) \\ k_{\text{PER}}(x, x') &= \sigma^2 \exp\left(-\frac{2 \sin^2(\pi(x-x')/p)}{\ell^2}\right) \\ k_{\text{LIN}}(x, x') &= \sigma_b^2 + \sigma_v^2(x-\ell)(x'-\ell) \\ k_{\text{RQ}}(x, x') &= \sigma^2 \left(1 + \frac{(x-x')^2}{2\alpha\ell^2}\right)^{-\alpha} \end{aligned}$$

本文对上述四种核函数进行了MATLAB仿真，将在后文对各个核函数的实验结果进行对比分析。

### 2.1.6 Additive GPR

该模型是GPR的变形，主要体现在核函数的选取上。

首先选取一维的基函数  $K_i(x_i, x'_i)$ ，由此定义各维度的核函数如下：

$$\begin{aligned} K_{\text{add}_1}(\mathbf{x}, \mathbf{x}') &= \sigma_1^2 \sum_{i=1}^D K_i(x_i, x'_i) \\ K_{\text{add}_2}(\mathbf{x}, \mathbf{x}') &= \sigma_2^2 \sum_{i=1}^D \sum_{j=i+1}^D K_i(x_i, x'_i) K_j(x_j, x'_j) \\ K_{\text{add}_n}(\mathbf{x}, \mathbf{x}') &= \sigma_n^2 \sum_{1 \leq i_1 \leq \dots \leq i_n \leq D} \left[ \prod_{d=1}^n K_{i_d}(x_{i_d}, x'_{i_d}) \right] \end{aligned}$$

$$K_{add_D}(\mathbf{x}, \mathbf{x}') = \sigma_D^2 \prod_{d=1}^D K_d(x_d, x'_d) \quad (19)$$

### 2.1.7 Warped GPR

该与Addictive GPR类似，Warped GPR也是通过改变核函数而提出的高效过程的变种。

与传统高斯过程回归不同的是，Warped GPR决定协方差矩阵的核函数为：

$$C_{mn} = v_1 \exp \left[ -\frac{1}{2} \sum_{d=1}^D \left( \frac{x_d^{(m)} - x_d^{(n)}}{r_d} \right)^2 \right] + v_0 \delta_{mn} \quad (20)$$

其中 $r_d$ 是与输入自变量维度一致的参数向量 $\mathbf{r}$ 的某一维度对应的值。

本模型的参数选取主要根据下式：

$$\begin{aligned} \mathcal{L} = & \frac{1}{2} \log \det \mathbf{C}_N + \frac{1}{2} f(\mathbf{t}_N)^\top \mathbf{C}_N^{-1} \mathbf{t}_N \\ & - \sum_{n=1}^N \log \left. \frac{\partial f(t)}{\partial t} \right|_{t_n} + \frac{N}{2} \log 2\pi \end{aligned} \quad (21)$$

## 2.2 算法实现

### 2.2.1 标准线性回归

---

#### Algorithm 1 Prediction for Linear Regression

---

Input  $\mathbf{X}(\text{train}), \mathbf{Y}(\text{train}), \mathbf{X}^*(\text{test})$   
 Generate weights:  $\mathbf{W} = (\mathbf{X}\mathbf{X}^\top)^{-1}(\mathbf{X}\mathbf{Y})$   
 Generate prediction:  $\mathbf{Y}^* = \mathbf{X}^\top \mathbf{W}$   
 Output  $\mathbf{Y}^*$

---

线性回归算法直接由给定的训练值得到权重 $\mathbf{w}$ ，然后作用于 $\mathbf{X}^*$ ，对于给定的仿真数据，程序需要遍历给定的每个人的成绩，故时间复杂度为 $O(N)$ 。

### 2.2.2 贝叶斯回归

---

#### Algorithm 2 Prediction for Bayesian Regression

---

Input  $\mathbf{X}(\text{train}), \mathbf{Y}(\text{train}), \mathbf{X}^*(\text{test})$   
 Initialize  $\sigma_n, \text{Conv}$   
 Generate  $\mathbf{A} = \sigma_n^{-2} \mathbf{X}\mathbf{X}^\top + \text{Conv}^{-1}$   
 Generate weights:  $\mathbf{E}_w = \sigma_n^{-2} \mathbf{A}^{-1} \mathbf{X}\mathbf{Y}$

---

Generate prediction:  $Y^* = X^T E_w$

Output  $Y^*$

---

贝叶斯回归于线性回归过程类似，时间复杂度为  $O(N)$ 。

### 2.2.3 基于多项式基函数的广义线性回归

---

#### **Algorithm 3** Prediction for Polynomial Regression

---

Input  $X(\text{train}), Y(\text{train}), X^*(\text{test})$

Initialize  $\sigma_n, \text{Conv}$

Generate  $\phi(x)$

Generate  $A = \sigma_n^{-2} \phi(X) \phi(X)^T + \text{Conv}^{-1}$

Generate prediction:  $Y^* = \sigma_n^{-2} \phi(X^*)^T A^{-1} \phi(X) Y$

Output  $Y^*$

---

分析算法可知，该算法并无嵌套循环，因此时间复杂度为  $O(N)$ 。

### 2.2.4 高斯过程回归

---

#### **Algorithm 4** Prediction for GP Regression

---

Input  $X(\text{train}), Y(\text{train}), X^*(\text{test}), K(X_1, X_2)$  (covariance function)

Initialize  $\sigma_n$

Generate  $E(Y^* | X, Y, X^*) = K(X^*, X) [K(X, X) + \sigma_n^{-2} I]^{-1} Y$

Generate prediction:  $Y^* = E(Y^* | X, Y, X^*)$

Output  $Y^*$

---

该算法无嵌套循环，程序时间复杂度为  $O(N)$ 。

### 2.2.5 Additive GPR

---

#### **Algorithm 5** Prediction for Additive GPR

---

Input  $X(\text{train}), Y(\text{train}), X^*(\text{test}), k(x_1, x_2)$  (base function)

Initialize  $\sigma_1, \sigma_2, \sigma_D$

Generate  $K(X_1, X_2)$ :

---

```

for i = 1:d
    temp = temp * k(x1, x2)
end for
K(X1, X2) = sigma_D2 * temp
Generate E(Y*|X, Y, X*) = K(X*, X)[K(X, X) + sigma_n2*I]-1Y
Generate prediction: Y* = E(Y*|X, Y, X*)
Output Y*

```

---

程序时间复杂度为  $O(N)$ 。

## 2.2.6 Warped GPR

---

### Algorithm 5 Prediction for Warped GPR

---

```

Input X(train), Y(train), X*(test)
Initialize r, v0, v1
Generate K(X1, X2):
for i = 1:d
    flag = flag + ((X1(i, m) - X2(i, n)) / r(i))2
end for
Kmn(X1, X2) = v1 * exp(-1/2 * flag) + v0 deltamn
Generate E(Y*|X, Y, X*) = K(X*, X)[K(X, X) + sigma_n2*I]-1Y
Generate prediction: Y* = E(Y*|X, Y, X*)
Output Y*

```

---

程序时间复杂度为  $O(N)$ 。

## 2.3 实验结果及分析

由于最终的测试集占30%，因此在仿真阶段选取训练集的70%作为训练数据，30%作为测试数据。先由选取的训练数据得到模型，然后由测试数据得到预测数据。由于测试的30%数据随机过程的分数，可以与预测数据进行对比，从而便于调节模型参数以及对模型的预测效果有较为客观的估计。（数据分析均采用2010年的数据）

### 2.3.1 标准线性回归

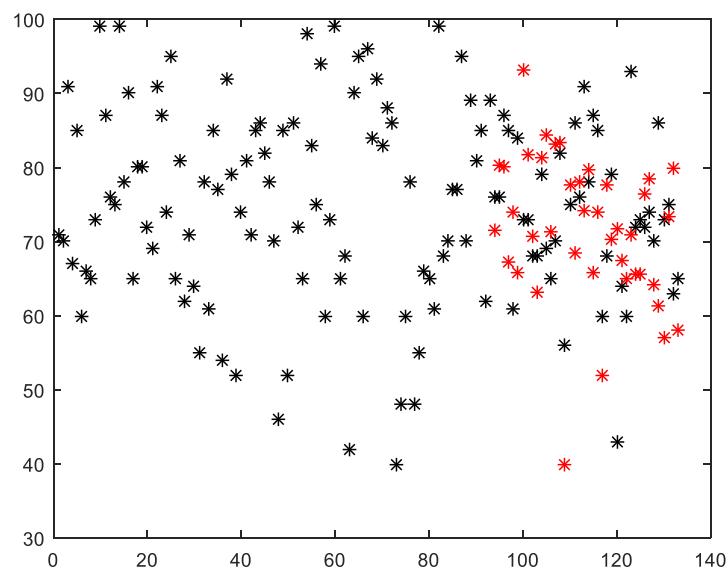


图1

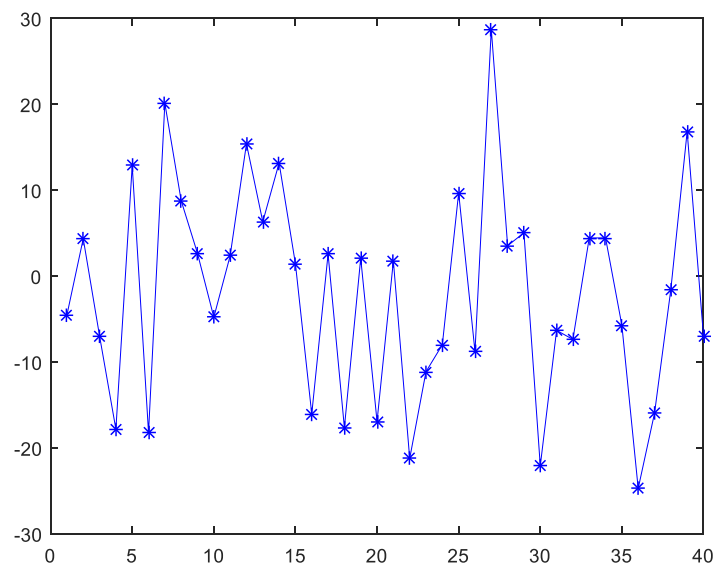


图2

图1是测试数据得到的预测数据与实际值的分布图，‘\*’代表预测值，‘\*’代表实际值。图二是预测值与实际值的差值的变化趋势。（后面的图像类似）

在70%训练数据，30%测试数据的条件下，标准线性回归的均方误差为156.7383，均方误差较大。从图2可以直观看出，预测值与实际值的差值最大接近30。此外由图1可以看出得到的预测值的方差小于实际值，比实际值的分布更密集。

### 2.3.2 贝叶斯回归



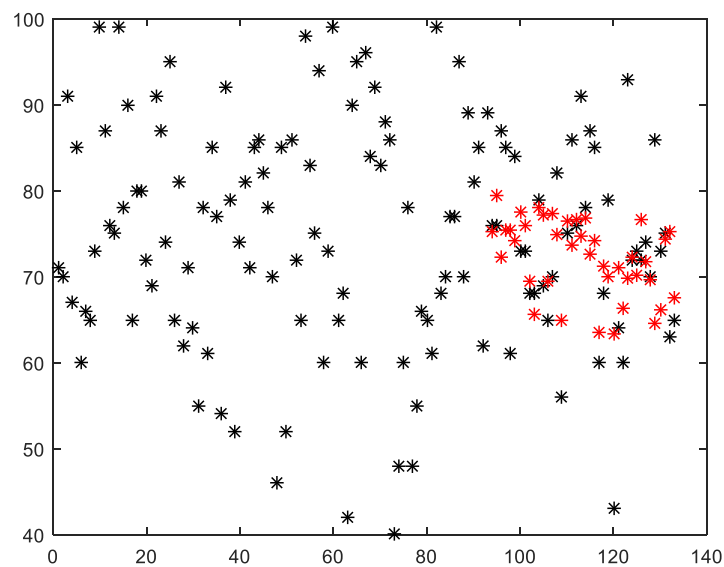


图3

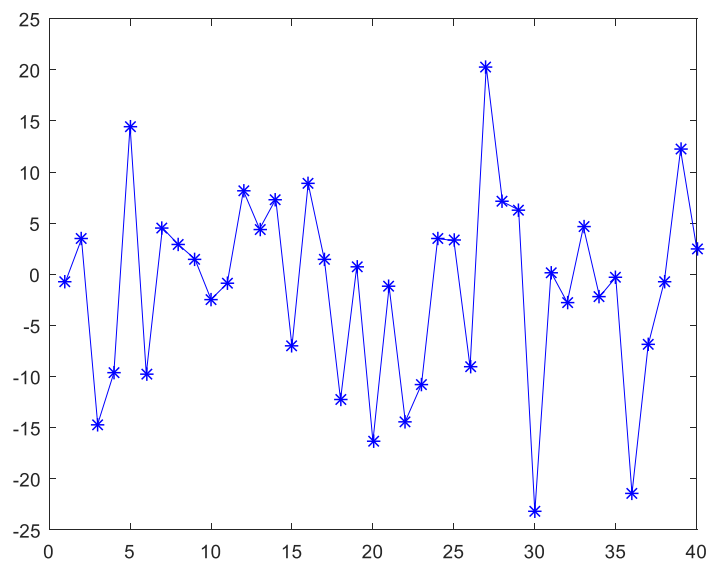


图4

贝叶斯回归模型的均方误差为87.9910，可以看出相比于标准线性回归MSE值减小幅度较大。由图2和图4也可以看出预测值与实际值的差值有一定程度的减小。此外对比图1和图3可以发现贝叶斯回归得到的预测值比标准线性回归更为集中，且更接近实际数据的平均值，因此MSE更小。但预测数据过于集中也可能带来新的问题，如果实际数据方差较大，则也可能导致MSE变大。

### 2.3.3 基于多项式基函数的广义线性回归

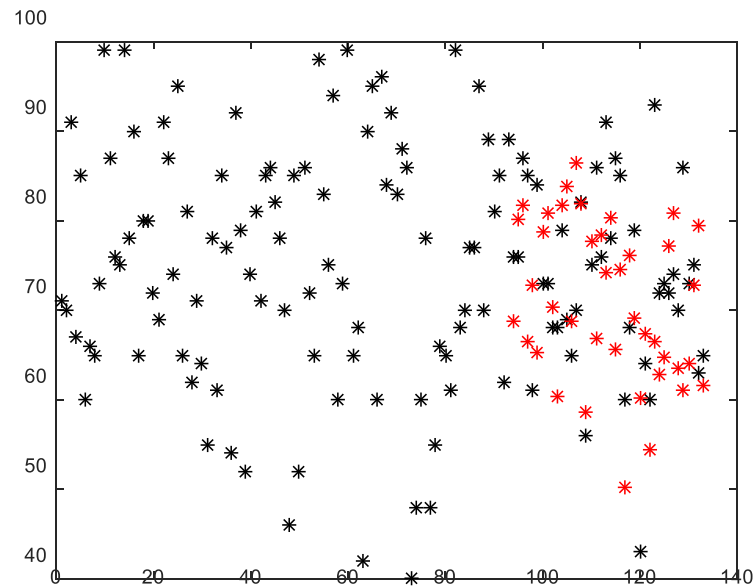


图5

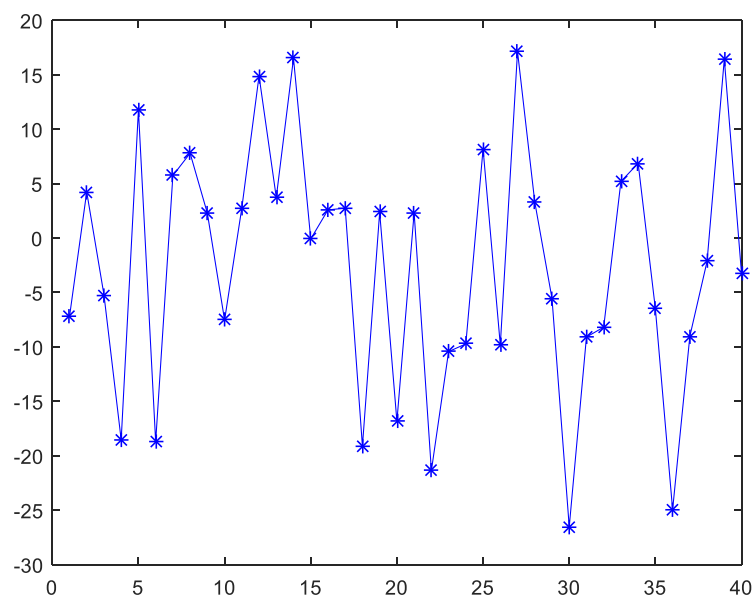


图6

基于多项式基函数的广义线性回归的本质是非线性回归，因此预测值并不集中分布。该模型的均方误差为134.9434。虽然MSE较大，由于分布较为分散，导致预测值与实际值的差值的最大值反而较小，图6较为直观地反映了这一结论。

### 2.3.4 高斯过程回归（核函数为SE）

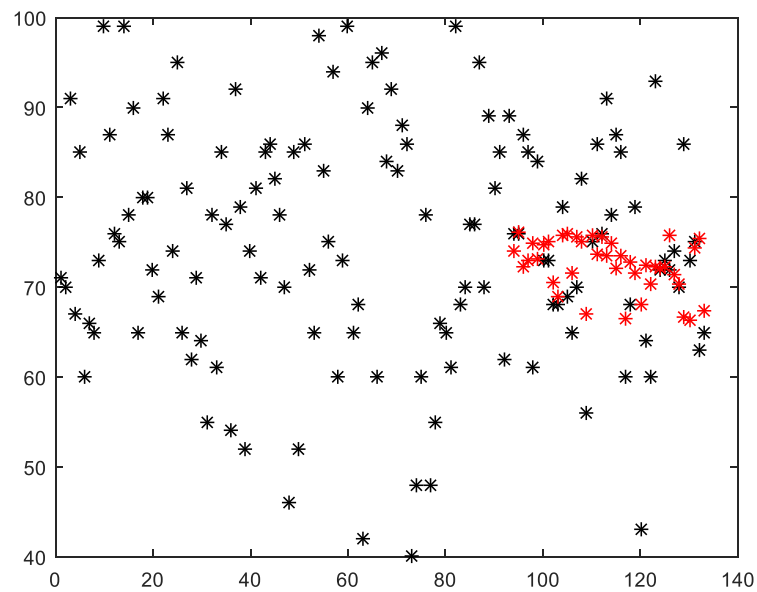


图7

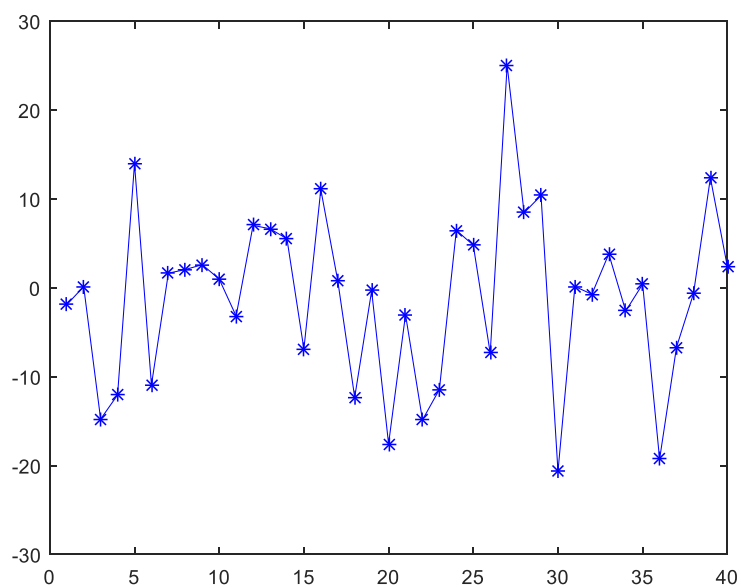


图8

以SE为核函数的高斯过程回归的MSE为94.2678。由图7可知预测值较为集中，如果对于成绩分布较为分散的测试集，则效果较差。与贝叶斯模型对比可知，贝叶斯模型更适合10年的成绩分布。

### 2.3.5 核函数的选取

(a).SE:平方指数协方差函数实验结果见高斯过程回归部分分析。

(b).PER

$$k_{\text{PER}}(x, x') = \sigma^2 \exp\left(-\frac{2 \sin^2(\pi(x-x')/p)}{\ell^2}\right)$$

将上式作为高斯过程回归的核函数，MATLAB仿真结果如下：

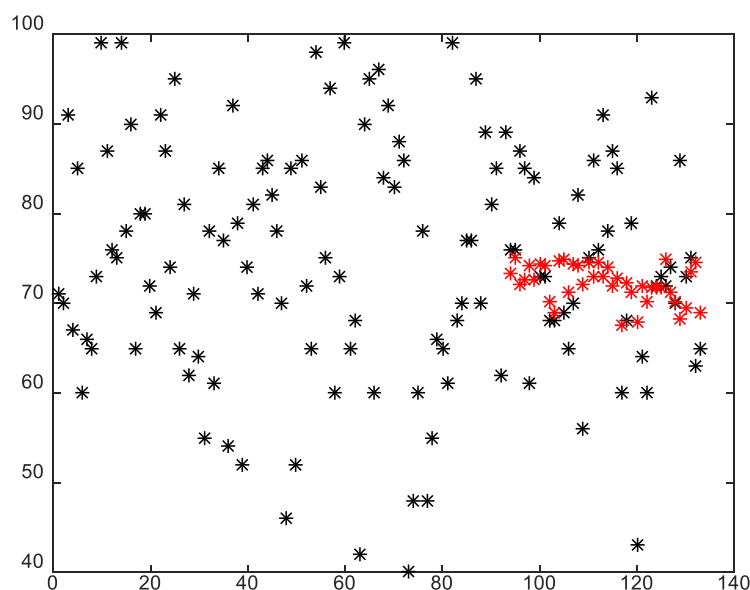


图9

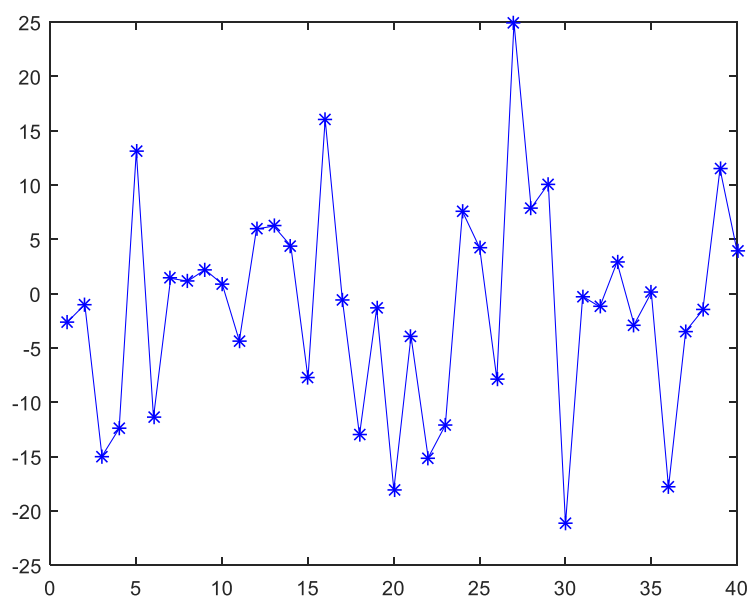


图10

对于10年的数据，以PER为核函数的高斯过程回归的MSE为97.1104，比SE的MSE稍大。在对PER的参数进行调节时，发现PER的分散性比SE好，但对于线性较强的数据，SE比PER效果好。由前面的实验结果，10年的成绩，贝叶斯回归比GPR效果更好，这表明10年数据线性性更强，SE效果好于PER，这与实验结果一致。

### (c).LIN

$$k_{\text{LIN}}(x, x') = \sigma_b^2 + \sigma_v^2(x - \ell)(x' - \ell)$$

将上式作为高斯过程回归的核函数，MATLAB仿真结果如下：

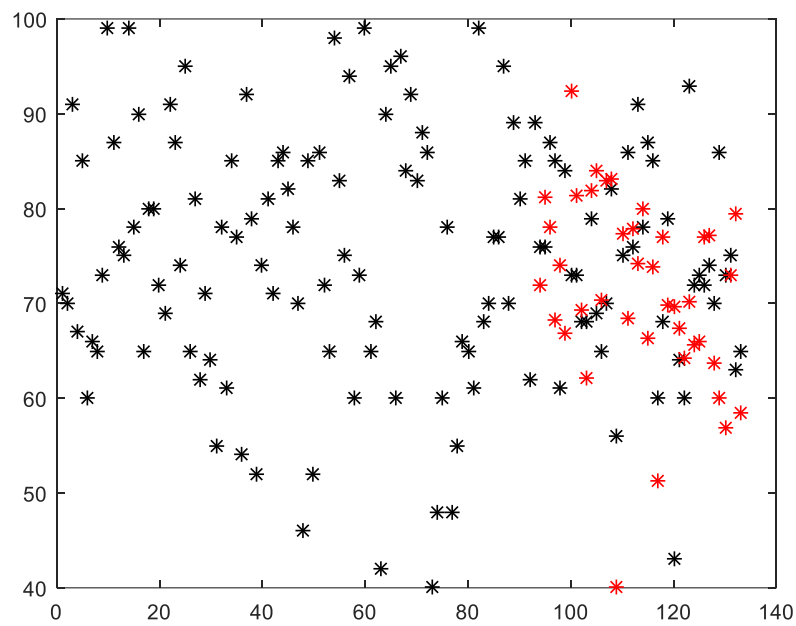


图11

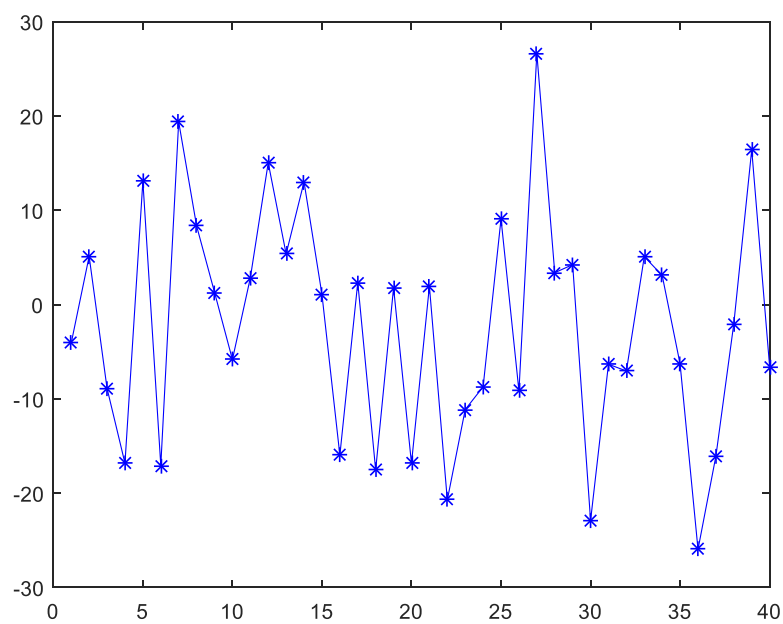


图12

对于10年的数据，以LIN为核函数的高斯过程回归的MSE为152.8682，比SE和PER的MSE都大。由仿真结果可知，LIN的分布过于分散，MSE偏大，不适合用于对成绩的高斯回归。

#### (d).RQ

$$k_{\text{RQ}}(x, x') = \sigma^2 \left( 1 + \frac{(x-x')^2}{2\alpha\ell^2} \right)^{-\alpha}$$

将上式作为高斯过程回归的核函数，MATLAB仿真结果如下：

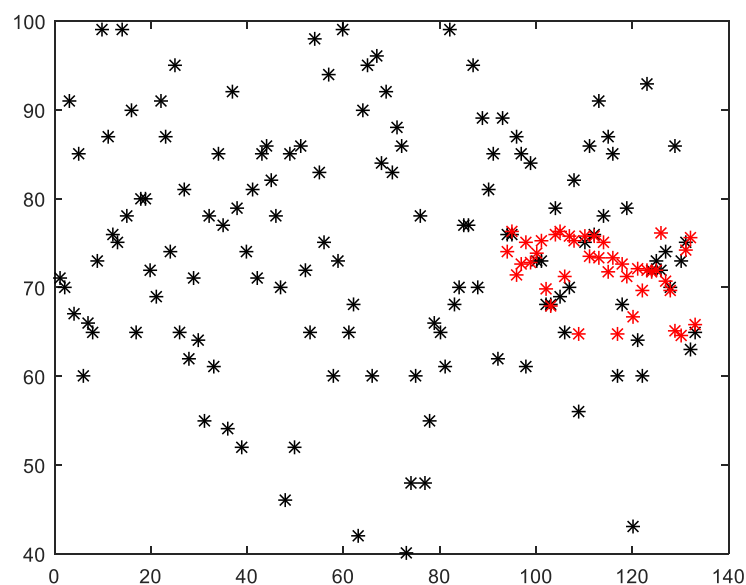


图13

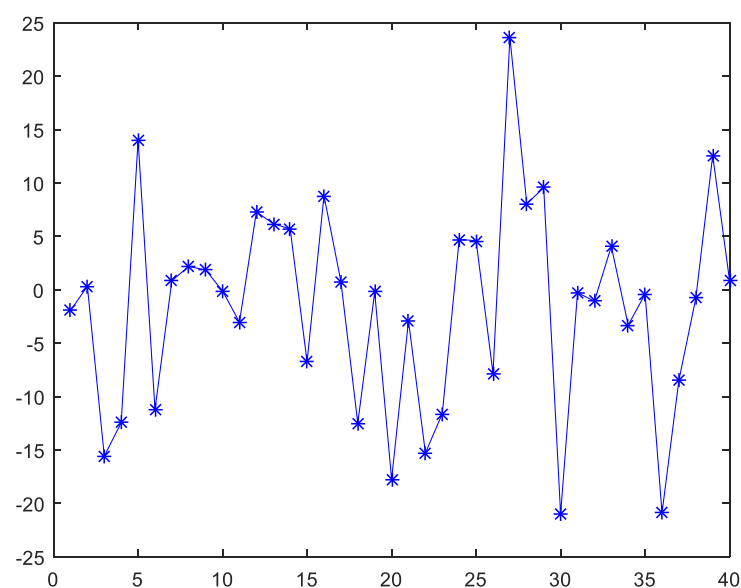


图14

对于10年的数据，以RQ为核函数的高斯过程回归的MSE为94.8520，比SE的MSE稍大，稍小于PER。对比图7和图13可知，相比于SE，RQ分散性稍好。由于给定的待测数据对应的随机过程成绩未知，且成绩方差较大，所以在满足MSE足够小的前提下，预测值较为分散比较好。

Kernel Function	MSE
SE	94.2678
PER	97.1104
LIN	152.8652
RQ	94.8520

综合比较上述四种核函数的仿真结果，可以看出SE, PER, RQ的回归效果较

好且大致相同,均可作为高斯过程回归的核函数。而LIN由于预测值过于分散,MSE较大,不适合作为本次实验中核函数。针对具体的数据,合理利用SE, PER, RQ,将提高GPR的精度。

在此基础上,本文又进一步尝试用上述核函数进行相加或者相乘的组合,得到更多参数的核函数(比如SE<sub>EX</sub>PER, SE+PER等),理论上组合得到的多参数核函数会有更好的回归效果,但由于成绩的不确定度较大,隐藏组合核函数并没有明显的精度提升。经一定的MATLAB仿真后,进一步确认组合核函数并明显无精度提升,故不在此赘述组合得到的多参数核函数。

### 2.3.6 Additive GPR

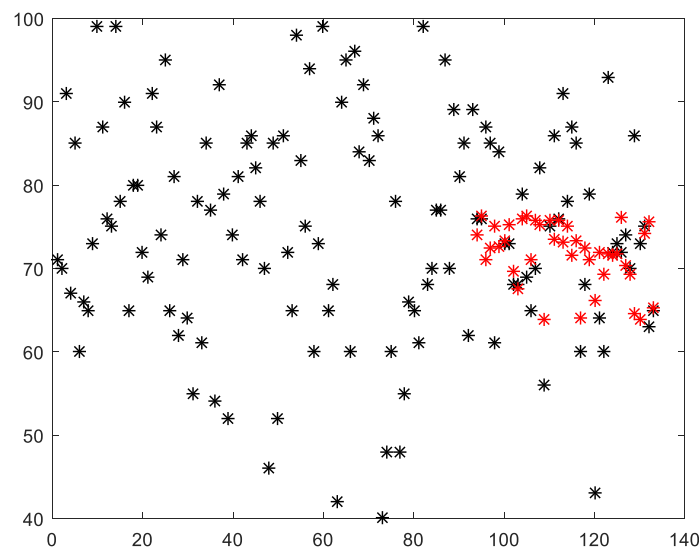


图15

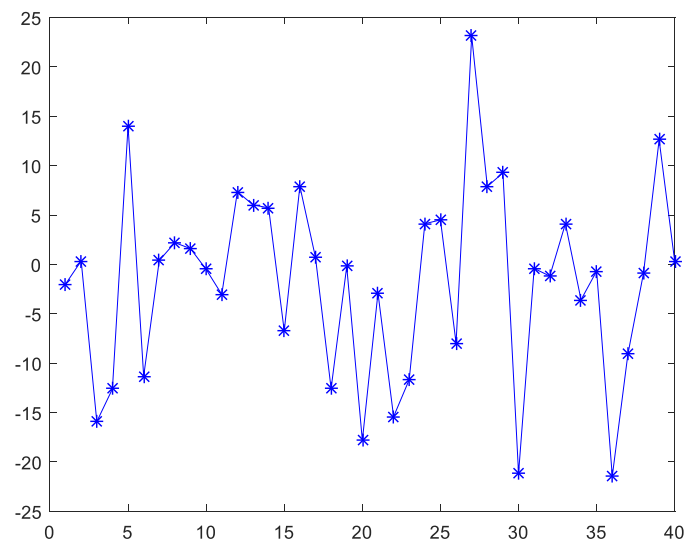


图16

对于10年的数据, Additive GPR的MSE为95.3517, 十分接近以SE为核函数的GPR, 因此也可作为代替GPR的回归模型。由图15、16可以看出, 回

归得到的预测数据较为集中，与SE类似，但是在调节过程中，如果牺牲一定的MSE，也能得到较好的分散度，与PER的特性比较相似。

### 2.3.7 Warped GPR

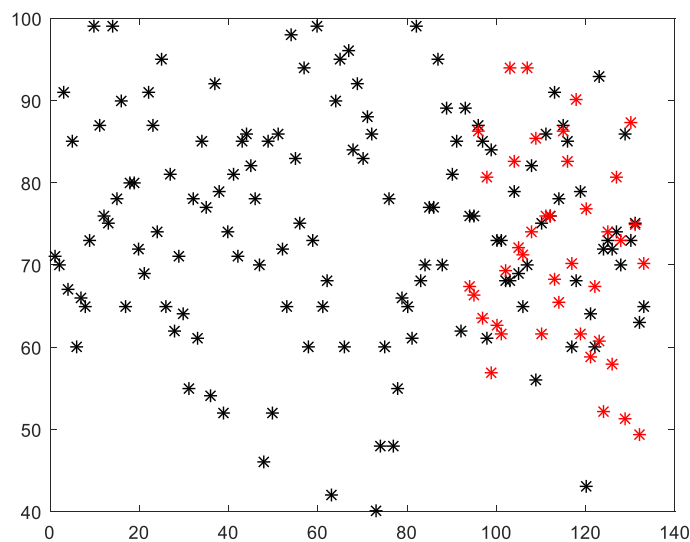


图17

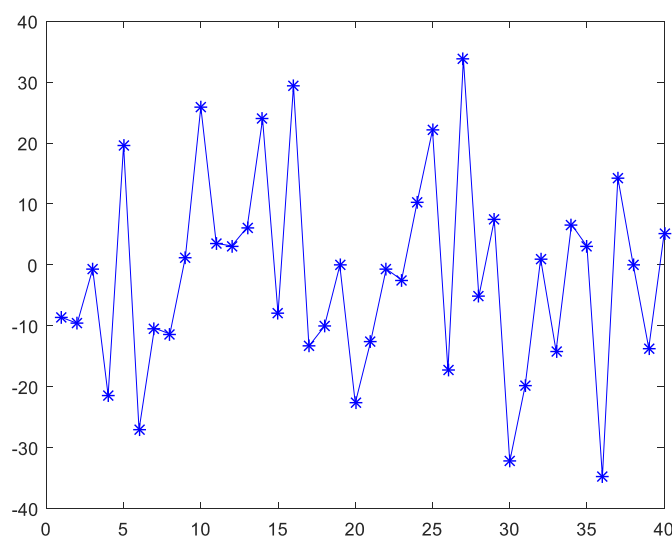


图18

Warped GPR的MSE为265.2551，远大于普通GPR。由图17、18可以看出，预测值分布过于分散，预测值与实际值的差值大，说明该模型并不适用于成绩预测。

## 3 误差分析及实验改进(Error Analysis and Improvement)

### 3.1 误差分析

本文实现了方法A, B, C, D，并针对SE, LIN, PER, RQ等核函数选取进行了



MATLAB仿真。在此基础上，本文还额外实现了两种GPR的改进形式Additive GPR和Warped GPR。针对2010年数据，仿真得到的MSE最小的模型是贝叶斯模型以及GPR和Additive GPR，但最小的GPR也是87.9910，预测值与实际值平均相差9.38分。

考虑到成绩的不确定性过大，对于某个个体，很可能由于发挥失常而分数远低于前面几门课的成绩，不难得出结论：预测每年的平均成绩较为简单，准确预测每个个体成绩很困难，只能尽量减小预测误差。对于GPR，很难找到核函数能准确描述个人随机过程成绩与前面过程的关系，毕竟不确定因素太多。

### 3.2 实验改进

进一步分析给定的8门课程与随机过程的知识关系，结合学习经验，可以推测随机过程与大学物理，离散数学等课程相关度低于其他课程，为了降低回归的误差，本文尝试去除部分相关性较低的的课程后再进行回归预测。

下面同样就2010年的数据进行分析。为了便于描述对各课程进行编号：

一元微积分	信号与系统	数理方程引论	复变函数引论	几何与代数(1)	几何与代数(2)	多元微积分	概率论与随机过程1
1	2	3	4	5	6	7	8

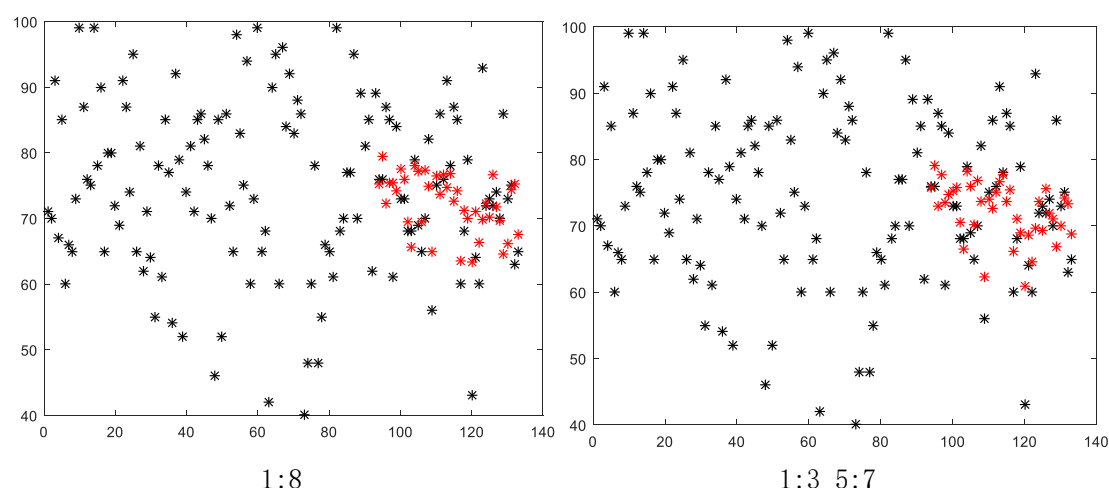
部分组合的仿真结果如下：

课程编号	贝叶斯 (MSE)	GPR (MSE)
1:8	87.9910	94.2678
2:8	89.1381	95.5319
1:7	80.3404	85.3064
1:2 4:7	83.1663	85.9305
1:3 5:7	75.0898	82.1658
1:3 7	80.8170	87.3673

由上表可以看出，不同的课程组合确实对最终的MSE有一定的影响，可以看出去除复变函数和概率论课程后，MSE减小到了75.0898，有较大提升。不过实验结果表明概率论与随机过程的关系并不大，这与实际经验相悖，对后续几年的数据进行分析时，发现概率论与随机过程成绩相关性较高，所以我认为可能是10年的数据有问题，也许概率论标签下对应的成绩并不是学生的概率论成绩。

总之，对于10年的数据，采用贝叶斯模型回归，采用一元微积分，信号与系统，数理方程，几何与代数(1)，几何与代数(2)，多元微积分这几门课能得到较好的回归效果，MSE为75.0899。

回归结果分布如下图：



左图为全部8门课的回归效果，右图为部分课程的回归效果。对比两图可知，右图更为分散，左图过于集中且右图的MSE小于左图。可知去除部分相关性较弱的课程后确实能减小MSE。

#### 4 结论(Conclusion)

本文完整的实现了基础部分的四种回归方法，并在此基础上进一步深入，讨论了核函数的选取对GPR的影响。此外，在高斯过程回归的基础上我还另外实现了近年来比较流行的两种改进的GPR (Additive GPR和Warped GPR)。并针对2010年的数据进行了分析，最后发现GPR (PER, SE, RQ), Additive GPR 和贝叶斯回归模型有较好的结果，可以根据不同年份的数据选取MSE最小的模型，而Warped GPR及其他模型有较大不差，不适合本项目的的成绩预测。此外，本文通过MATLAB仿真发现，去除部分相关性较小的课程，能进一步提高回归精度，最终2010年的数据最小MSE为75.0899, 各年仿真结果如下：

	2010	2011	2012	2013	2014
Linear	135.9019	189.7422	183.3860	196.2939	179.9887
Bayesian	75.0899	151.1366	161.7234	151.5455	121.9324
Polynomial	121.7342	184.3475	203.8783	188.8780	178.0379
GPR	82.1658	153.2535	159.0721	171.5104	148.1548
Additive GPR	85.3032	153.2767	159.9689	171.9337	148.4583
Warped GPR	265.2551	361.6248	520.6140	229.8462	247.3809

根据仿真结果，最终12年数据选取GPR (SE)，其他年份选取贝叶斯模型。

#### 5 致谢(Acknowledgement)

感谢professor OU在课堂上对高斯过程回归的相关讲解，使我对高斯过程回归有了最基本的认识。

感谢bjy在参数调节部分和我的讨论，让我对参数选取有了进一步的认识。

感谢zml在协方差矩阵选取上对我的指导，让我对之前完全硬着头皮试出来的参数有了系统的认识，让我在后续实验的参数调节中避免了很多困难。

## 附录：

- 注：1. 所有实验统一仿真环境：win10操作系统，MATLAB2016a。  
2. 所有数据均保存在GPR.mat文件中，a, b, c, d分别对应方法A, B, C, D的预测数据，e1存放的是Additive GPR的预测数据，e2存放的是Warped GPR预测数据，optimal存放的是由不同模型得到的最好的预测结果。  
3. GPR.mat文件在GPR-MATLAB主文件夹下，不同方法代码见各子文件夹。

## 参考文献

- [1] 茆诗松，程依明，濮晓龙，概率论与数理统计教程. 高等教育出版社，2004.
- [2] 陈家鼎，郑忠国，概率与统计. 北京大学出版社，2007.
- [3] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. MIT Press, 2006.
- [4] D. Duvenaud, J. R. Lloyd, R. Grosse, J. B. Tenenbaum, and Z. Ghahramani, "Structure Discovery in Nonparametric Regression through Compositional Kernel Search," *Creative Commons Attribution-Noncommercial-Share Alike*, 2013
- [5] Duvenaud, D., Nickisch, H., and Rasmussen, C.E. Additive Gaussian processes. In *Advances in Neural Information Processing Systems*, 2011.
- [6] E. Snelson, Z. Ghahramani, and C. Rasmussen. Warped Gaussian processes. In *Advances in Neural Information Processing Systems* 16, 2003.
- [7] Miguel Lázaro-Gredilla. Bayesian Warped Gaussian Processes. *Advances in Neural Information Processing Systems* 25. 2012