# Пятое задание по БМСО

Ярослав Аверьянов

Ноябрь 2015

Пусть задана выборка $D = (X, \mathbf{y}) = (\mathbf{x_i}, y_i = y(\mathbf{x_i}))_{i=1}^n$ - выборка из $n$ значений функции $y(\mathbf{x_i})$ в точках $\mathbf{x_i}, i = 1, ..., n$. Задана ковариационная функция гауссовского процесса $k_\theta(\mathbf{x}, \mathbf{x}^{'})$ из параметрического семейства $k_\theta(\mathbf{x}, \mathbf{x}^{'}), \theta \in \Theta$.

## 1 Первая задача

$$p(\mathbf{x}) = N(\mathbf{x}|\mu, \Lambda^{-1}),$$
$$p(\mathbf{y}|\mathbf{x}) = N(\mathbf{y}|A\mathbf{x} + \mathbf{b}, L^{-1})$$

Покажем, что:

$$p(\mathbf{y}) = N(\mathbf{y}|A\mu + \mathbf{b}, L^{-1} + A\Lambda^{-1}A^T),$$
$$p(\mathbf{x}|\mathbf{y}) = N(\mathbf{x}|\Sigma A^T L(\mathbf{y} - \mathbf{b}) + \Lambda\mu, \Sigma),$$
$$\text{где } \Sigma = (\Lambda + A^T LA)^{-1}.$$

У нас $p(x) \sim exp(-\frac{(x-\mu)^T \Lambda(x-\mu)}{2})$
$p(y|x) \sim exp(-\frac{(y-Ax-b)^T L(y-Ax-b)}{2})$
Здесь $x \sim N(\mu, \Lambda^{-1})$ и $\epsilon \sim N(0, L^{-1})$.
Далее возьмем следующую СВ: $y^{'} = Ax + b + \epsilon$ и для нее выполняется равенство: $p(y^{'}|x) = p(y|x) \to y^{'} \sim y$;
$Ax + b \sim N(A\mu + b, A\Lambda^{-1}A^T)$.
$Ax + b + \epsilon$ - также гауссовский процесс с математическим ожиданием $A\mu + b$ и с корреляционной матрицей:

$\mathbf{E}[(Ax - A\mu + \epsilon)(Ax - A\mu + \epsilon)^T] = \mathbf{E}[(Ax - A\mu)(Ax - A\mu)^T] + 2 \cdot E[(Ax - A\mu)\epsilon^T] + \mathbf{E}[\epsilon\epsilon^T] = \mathbf{E}[(Ax - A\mu)(Ax - A\mu)^T] + \mathbf{E}[\epsilon\epsilon^T] = A\Lambda^{-1}A^T + L^{-1} \to y \sim N(A\mu + b, A\Lambda^{-1}A^T + L^{-1})$;

Т.к. $y = Ax+b+\epsilon \to \epsilon = y-b-Ax$. Для данной точки $(x, \epsilon)$ плотность в ней $\sim exp(-\frac{(x-\mu)^T\Lambda(x-\mu)}{2}) \cdot exp(-\frac{(y-Ax-b)^T L(y-Ax-b)}{2}) \sim exp(-\frac{x^T(\Lambda+A^TLA)x-2x^T(\Lambda\mu+A^T(y-b))}{2})$.

$\to p(x|y) \sim N(\Sigma(\Lambda\mu + A^T(y-b)), \Sigma)$, где $\Sigma = (\Lambda + A^TLA)^{-1}$.

## 2 Вторая задача

Покажем, что для $\theta < 0$ функция $k_\theta(x, x^{'}) = exp(-\theta(x-x^{'})^2)$ не может быть ковариационной функцией гауссовского процесса для $x \in \mathbf{R}$

Мы имеем, что $-\theta(x - x^{'})^2 \geq 0 \to exp(-\theta(x - x^{'})^2) \geq 1 \to k_\theta(x, x^{'}) \geq 1$

Далее воспользуемся неравенством Коши-Буняковского:

$k(X(t_1), X(t_2)) = \mathbf{E}[X(t_1), X(t_2)] \leq \sqrt{\mathbf{E}[X^2(t_1)]} \cdot \sqrt{\mathbf{E}[X^2(t_2)]} = 1$

$\to$ Противоречие.

## 3 Третья задача

LR:

$$y(t) = \sum_{q=1}^{Q} x_q(t)\beta_q + \epsilon(t)$$

При $t = t_i \forall i = 1, ..., n$

$$\epsilon(t_i) \sim N(0, \sigma_\epsilon^2) \text{ и } \beta_q \sim N(0, a_q)$$

$\to cov(y(t_i), y(t_j)) = \sum_{q=1}^{Q} a_q x_q(t_i) x_q(t_j) + \delta_{ij}\sigma_\epsilon^2$

$\to LR$ может быть переписана как GPR:

$$y(t) = f(\vec{x}(t)) + \epsilon(t) \text{ или } y(\vec{x}) = f(\vec{x}) + \epsilon(\vec{x})$$

$$\text{где } \vec{x} = (x_1, ..., x_Q), \quad f(\cdot) \sim GP(0, k_{linear}(\cdot, \cdot)),$$

$$k_{linear} = \sum_{q=1}^{Q} a_q x_q x_q^{'}$$

## 4 Четвертая задача

$D_{-j} = [(\mathbf{x_i}, y_i = y(\mathbf{x_i}))]_{i=1,\ldots,j-1,j+1,\ldots,n}$.

$\hat{y_{-j}} = \mathbf{E}[p(y(\mathbf{x_j})|D_{-j}, \theta)]$ в точке $\mathbf{x_j}$ с ковариационной матрицей $k_\theta(\mathbf{x}, \mathbf{x}^{'})$ и выборкой $D_{-j}$.

$$\text{Получить выражение для } \sum_{j=1}^{n} (\hat{y_{-j}} - y_j)^2.$$

Обозначим $K_{ij} = k_\theta(x_i, x_j)$ и введем диагональную матрицу $R_\theta$, на диагонали которой находятся диагональные элементы $K^{-1}$.

Тогда мы пытаемся найти $\hat{\theta} : \hat{\theta} \in argmin \sum_{i=1}^{n} (y_i - \hat{y_{\theta,-i}})^2$.

Тогда если поочередно рассматриваить каждый элемент суммы, то минимум достигается при:

$\hat{y}_{-i} - y_i = \frac{e_i^T K^{-1} y}{e_i^T K^{-1} e_i}$, где $e_i$ - $i$ый единичный орт.

Т.е. $y_i - \hat{y}_{\theta,-i} = \frac{1}{(K^{-1})_{i,i}} (K^{-1}y)_i$. Тогда $\sum_{i=1}^{n} (y_i - \hat{y}_{\theta,-i})^2 = y^T K^{-1} R_\theta^{-1} R_\theta^{-1} K^{-1} y$.

## 5 Пятая задача

В этой задаче нужно было сравнить работу 4 алгоритмов добавления точки на $n$-ом шаге:

1.Новая точка добавляется случайно.

2.Новая точка максимизирует минимальное расстояние до точек текущей выборки $D_n$.

3.Новая точка максимизирует апостериорную дисперсию гауссовского процесса для $D_n$.

4.Новая точка минимизирует ошибку аппроксимации на заданной тестовой выборке $D_{test}$, если добавить ее к текущей выборке $D_n$.

Мною использовалась функция $f(x, y) = -(x^2 + y^2)$.

Начальный дизайн - выборка размера 30 из $\mathbf{U}([0, 1]^2)$.

Точки-кандидаты - выборка размером 30 из $\mathbf{U}([0, 1]^2)$.

Тестовая выборка - выборка размером 40 из $\mathbf{U}([0, 1]^2)$.

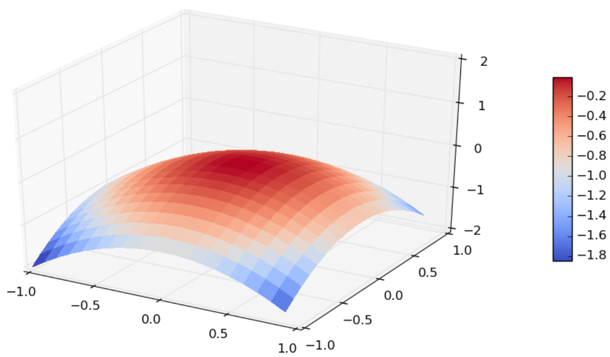


Рис. 1: $f(x,y) = -(x^2 + y^2)$

В результате был построен график среднеквадратичной ошибки на тестовой выборке после добавления новой точки. Лучший алгоритм - четвертый алгоритм, худший - первый алгоритм .

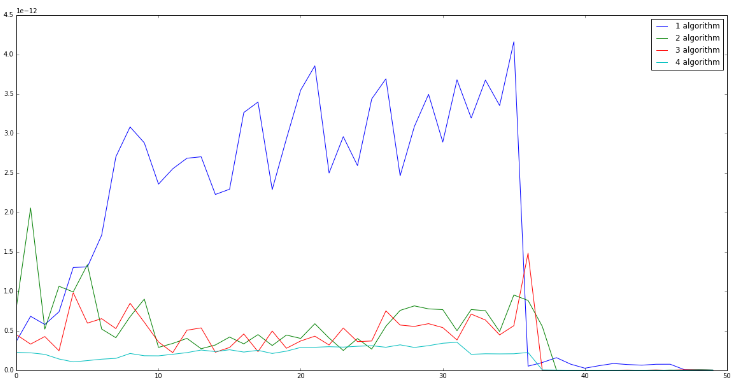


Рис. 2: Среденеквадратичная ошибка на тестовой выборке в зависимости от количества точек, добавляемых в обучаемую выборку