Министерство образования Российской Федерации
Московский Государственный Технический
Университет им. Н.Э. Баумана

Отчет по лабораторной работе №3
По курсу ««Математическая статистика»

Тема: «Метод наименьших квадратов»

Вариант 11

Студент: **Медведев А.В.**
Группа: **ИУ7-62**

Преподаватель: **Власов П.А.**

Москва, 2019

Содержание

Определения

Постановка **задачи аппроксимации неизвестной зависимости по результатам наблюдений**
Пусть Y - случайная величина, $X_1, ..X_r$ - детерменированные величины (не являются случайными). Но величины $Y_i(X_1, ..X_r)$ при фиксированных значениях $X_1, ..X_r$ являются случайными.
Говорят, что переменная Y **стохастически** зависит от детерменированных переменных $X_1, ..X_r$, если на изменение значений этих переменных Y реагирует изменением своего закона распределения.
Предмет изучения **регрессионного анализа** составляют задачи, связанные с установлением стохастических зависимостей между случайной величиной Y и детерменированными переменными $X_1, ..X_r$
В регрессионном анализе используют модель черного ящика, как наиболее общую модель, ассоциированную с понятием отображения. На вход черного ящика поступает вектор $\vec{X}(X_1, ..X_r)$, который посредством отображения Φ и случайных возумещний $\vec{\epsilon} = (\epsilon_1, ..\epsilon_m)$ преобразуется в выходной вектор $\vec{Y} = (Y_1, ..Y_m)$.

Понятие МНК-оценки параметров линейной модели Имеются результаты n наблюдений:

$$F_n(x) = \begin{cases} \mathrm{y}_1 = & \Phi(t_1) + \epsilon_1 \\ & ... \\ \mathrm{y}_n = & \Phi(t_n) + \epsilon_n \end{cases}$$

Требуется на основании этих данных подобрать функцию $\hat{\Phi}(t)$ так, чтобы она наилучшим образом аппроксимировала неизвестную функцию Φ
Часто в качестве функции $\hat{\Phi}$ рассматривают функцию

$$\hat{\Phi}(t) = \theta_1\Psi_1(t) + .. + \theta_p\Psi_p(t),$$

где функции $\psi_j(t), j=\overline{1;p}$ - известны, и задача сводится к поиску коэффициентов $\theta_1,..\theta_p$

Обозначим $\vec{Y}=\begin{bmatrix} y_1 \\ ... \\ y_n \end{bmatrix}$, $\vec{\theta}=\begin{bmatrix} \theta_1 \\ ... \\ \theta_p \end{bmatrix}$, $\Psi=\begin{bmatrix} \Psi_1(t_1) & ... & \Psi_p(t_1) \\ & ... & \\ \Psi_1(t_n) & ... & \Psi_p(t_n) \end{bmatrix}$, $\vec{\epsilon}=\begin{bmatrix} \epsilon_1 \\ ... \\ \epsilon_n \end{bmatrix}$

Оценка $\vec{\theta}=(\hat{\theta}_1,..\hat{\theta}_p)^T$ вектора $\vec{\theta}$ называется **оценкой, полученной по методу наименьших квадратов**, если она доставляет минимальное значение функционалу

$$S(\vec{\theta}=||\Psi\vec{\theta}-\vec{Y}||^2=(\vec{Y}-\Psi\vec{\theta})^T(\vec{Y}-\Psi\vec{\theta})=\sum_{i=1}^n(y_i-\theta_1\Psi_1(t_i)-..-\theta_p\Psi_p(t_i))^2$$

Формулы для вычисления МНК-оценки

В рассматриваемом случае базисные функции Ψ имеют вид

$$\Psi(t_i)=t_i,\ i=\overline{1;n}$$

Для вычисления оценки вектора параметров $\vec{\theta}$ используется формула

$$\hat{\vec{\theta}}=(\Psi^T\Psi)^{-1}\Psi^T\vec{Y}$$

Среднеквадратичное отклонение полученнй модели:

$$\Delta=\sqrt{\sum_{i=1}^n(y_i-y(t_i))^2},$$

где y_i - значение Y, полученное экспериментально при $t=t_i$, а $y(t_i)$ - значение Y, полученное с помощью оценки методом наименьших квадратов.

Текст программы

Листинг 1: Текст программы

```
function Main()
    Y = importdata('dataY(v11).txt');
    T1 = importdata('dataT(v11).txt');
    T2 = T1.^2;
    T0(1:length(T1), 1)=1;

    psi = horzcat(T0,T1,T2);
    psiT = transpose(psi);
```

```

    theta = inv(psiT*psi)*psiT*Y;
    Yt = theta(1) + theta(2)*T1 + theta(3)*T2;
    fprintf("theta1 = %.2f\n",theta(1));
    fprintf("theta2 = %.2f\n",theta(2));
    fprintf("theta3 = %.2f\n",theta(3));
    Ystr = sprintf("Yt = %.2f + %.2f*t + %.2f*t^2\n", theta
       (1), theta(2), theta(3));
    fprintf(Ystr);

    delta = sqrt(sum((Y-Yt).^2));
    fprintf("delta=%.3f\n",delta);

    figure(1);
    xlabel('T');
    ylabel('Y');
    plot(T1, Y, '.r');
    hold on;
    plot(T1, Yt, 'b');
    legend("                              (t_i,y_i)", Ystr);
    grid on;
end
```

Результаты расчетов и графики

МНК-оценка вектора $\theta = (\theta_0, \theta_1, \theta_2)$ параметров модели $y = \theta_0 + \theta_1 t + \theta_2 t^2$

θ_0	-0.78
θ_1	0.97
θ_2	-0.90

Среднеквадратическое отклонение $\Delta = \sqrt{\sum_{i=1}^{n}(y_i - y(t_i))^2}$ полученной модели от результатов наблюдений

Δ	154.202

Построение на одном графике системы точек $(t_i, y_i), i = \overline{1;n}$ и графика функции $y = y(t), t \in [t_{(1)}; t_{(n)}]$ (для полученной оценки вектора θ)