Кравцова Александра Николаевна, группа 12.1

Лабораторная работа № 5

**Вариант №** **5b**

Исследование непараметрических алгоритмов оценивания плотности распределения случайной величины

**Цель работы**

Исследовать алгоритмы оценивания плотности распределения случайных величин и случайных векторов на основе методов Парзена и k ближайших соседей.

**Форма контроля**

Письменный отчёт (допускается преставление в электронном виде). Опрос в устной форме в соответствии с перечнем контрольных вопросов.

**Количество отведённых аудиторных часов**

4

**Содержание работы**

Получить у преподавателя вариант задания и написать код, реализующий соответствующий алгоритм обработки информации. Для ответа на поставленные в задании вопросы провести численный эксперимент или статистическое имитационное моделирование и представить соответствующие графики. Провести анализ полученных результатов и представить его в виде выводов по проделанной работе.

# Задание

1. Окно какого вида будет обеспечивать оптимальное по критерию среднеквадратичной ошибки оценивание плотности распределения двумерного случайного вектора по методу Парзена? Построить графики зависимостей ошибок от объема обучающей выборки. Сравните следующие виды окон:

b. гауссовская функция c использованием диагональной матрицы, показательная функция и оконная треугольная функция;

Код программы (Желтым выделены отличные от шаблона фрагменты):

clear all; close all;

%%% Пример вар.5. Построить графики зависимостей ошибок от объема обучающей выборки.

% (b. гауссовская функция c использованием диагональной матрицы, показательная функция, оконная треугольная функция;)

%% Здесь только Двумерный случай

% ЗДЕСЬ задаются перебираемые занчения величины N (объема обучающей выборки)

NN = 1000 : 1000 : 10000;

% Массивы значений ошибок для каждого типа оконной функции

err1 = NN \* 0; % массив значений ошибок заполненный нулями

err2 = NN \* 0; % массив значений ошибок заполненный нулями

err3 = NN \* 0; % массив значений ошибок заполненный нулями

% ЗДЕСЬ добавляется цикл по числу элементов NN

for tt = 1 : numel(NN)

% 1. Исходные данные

n=2; %n-размерность вектора наблюдений

% ЗДЕСЬ подставляем очередное значение из массива NN

N=NN(tt); %количество используемых для оценки векторов

r=0.5;

h\_N=N^(-r/n); %расчет параметра размера окна

kl\_kernel=12;%ключ выбора ядра оценки (см. описание функции vkernel)

% 2.Генерация отсчетов эталонной плотности (в виде смеси гауссиан) для двумерного случая

% Параметры распределения смеси гауссовских случайных векторов;

M=3; %количество компонентов в смеси

ps=[0.2,0.2,0.6]; %вероятности появления СВ различных типов в смеси

% Расчет матрицы ковариаций ГСВ смеси

D=0.2; ro=-log(0.7); %дисперсия и коэффициент корреляции cоседних элементов

% Расположение математических ожиданий компонентов смеси

m1=[0;0]; m2=[1;0]; m3=[0;1]; m=[m1,m2,m3];

% Ковариационная матрица компонентов смеси

for i=1:n, for j=1:n,

C(i,j)=D\*exp(-ro\*abs(i-j));

end;end;

x1=[-2:0.1:3]; x2= [-2:0.1:3]; %области значений СВ, для которой визуализируется оценка

[X1,X2]=meshgrid(x1,x2); x=[X1(:) X2(:)]'; % матрицы Х и Y координат отсчётов

% Значения эталонной плотности

p=ps(1)\*mvnpdf(x',m1',C)+ps(2)\*mvnpdf(x',m2',C)+ps(3)\*mvnpdf(x',m3',C);

% pi=reshape(p,length(x1),length(x2));%матрица значений плотности распределения

% 3. Убрать Отображение графика плотности распределения

% 4. Обучающая выборка

XN=zeros(n,N);

for i=1:N,%генерация обучающей выборки

u=rand;

%индекс принадлежности к компоненте смеси

if u<ps(1), t=1; elseif u<ps(1)+ps(2), t=2; else t=3; end;

XN(:,i)=randncor(n,1,C)+m(:,t);

end;

% 5. Оценка плотности по Парзену

p\_=vkernel(x,XN,h\_N,kl\_kernel);%оценка плотности

%Матрица значений оценки плотности распределения

% pv=reshape(p\_,length(x1),length(x2));

% 3. ЗДЕСЬ Оценки плотности по Парзену с заданными видами оконной функции

p1=vkernel(x,XN,h\_N,11);%оценка плотности (гауссовская функция c использованием диагональной матрицы)

p2=vkernel(x,XN,h\_N,2);%оценка плотности (показательная)

p3=vkernel(x,XN,h\_N,3);%оценка плотности (оконная прямоугольная функция)

% ЗДЕСЬ фиксируем среднеквадратичную ошибку

err1(tt) = sqrt(mean((p(:) - p1(:)) .^ 2));

err2(tt) = sqrt(mean((p(:) - p2(:)) .^ 2));

err3(tt) = sqrt(mean((p(:) - p3(:)) .^ 2));

end

% ЗДЕСЬ вместо п.4. выводим зависимость ошибки от объема обучающей выборки

figure;

plot(NN, err1, '-r', NN, err2, '-b', NN, err3, '-g');

xlabel('Объем выборки ','FontName','Courier');

ylabel('Значение среднеквадратичной ошибки','FontName','Courier');

legend('гаус (диаг. ков. мат)', 'показательная функция', 'оконная прямоугольная функция');

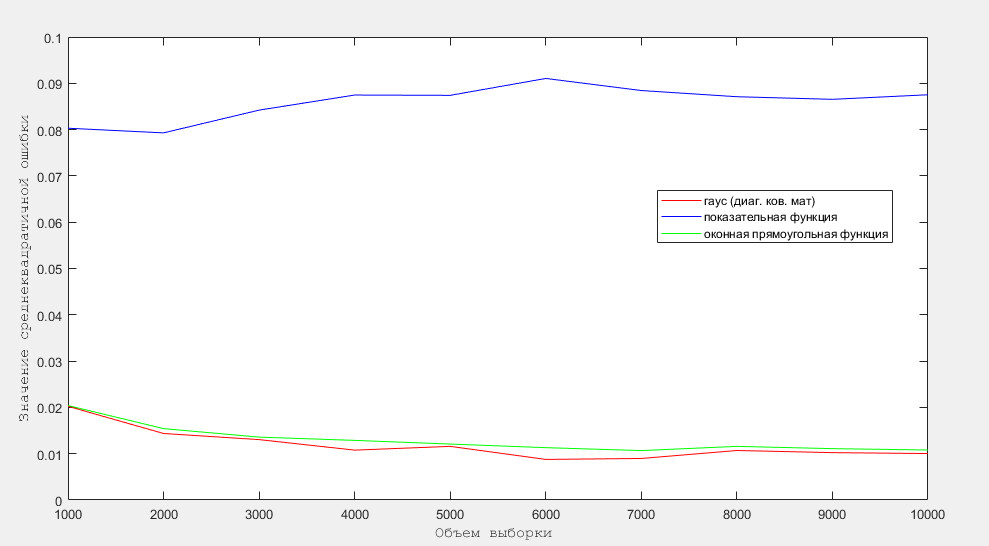
% Какая кривая ниже всех - такое окно и лучше

fprintf('гаус (диаг. ков. мат) = %f\n', err1(end));

fprintf('показательная функция = %f\n', err2(end));

fprintf('оконная прямоугольная функция = %f\n', err3(end));

**Результаты выполнения задания**



*Рис.1*

Кривые на графике отображают зависимость объема выборки (по оси Ox) от среднеквадратичной ошибки (по оси Oy).

Обратимся к полученному нами графику. На графике видно, что оптимальное по критерию среднеквадратичной ошибки оценивание плотности распределения двумерного случайного вектора в соответствии с методом Парзена будет обеспечивать окно вида *гауссовской функции с использованием диагональной ковариационной матрицы*. Так как график именно этой функции расположен ниже окон вида показательной функции и оконной прямоугольной функции, из чего следует, что при различных значениях в указанном объеме выборки (NN) функция устанавливает наименьшее значение среднеквадратичной ошибки.

Вывод показывает следующий результаты:

**гаус (диаг. ков. мат)** = 0.008814 (красный цвет на графике)

**показательная функция** = 0.089806 (синий цвет на графике)

**оконная прямоугольная функция** = 0.009949 (зеленый цвет на графике)

Что так же подтверждает, что окно вида *гауссовской функции с использованием диагональной ковариационной матрицы* обеспечивает оптимальное по критерию среднеквадратичной ошибки оценивание плотности распределения двумерного случайного вектора.

**Выводы**

Исследовали алгоритмы оценивания плотности распределения случайных величин и случайных векторов на основе метода Парзена. Посредством проведенного эксперимента установили, какой вид окна будет обеспечивать оптимальное оценивание плотности распределения двумерного случайного вектора, тем самым графически отобразили все окна вида, данные в задании, а именно: гауссовскую функцию c использованием диагональной матрицы, показательную функцию и оконную треугольную функцию;