Распознавание образов в условиях параметрической неопределенности и обучения с учителем

Цель работы

Синтезировать алгоритмы распознавания образов, описываемых гауссовскими распределениями с неизвестными параметрами. Оценить неизвестные параметры распределений на основе объектов обучающих выборок методом максимального правдоподобия и максимума апостериорной вероятности. Исследовать синтезированные алгоритмы распознавания с точки зрения ожидаемых потерь и ошибок в зависимости от объема обучающей выборки.

1. Постановка задачи машинного обучения с учителем

Образы объектов, для описания которых используются n признаков, представляются реализациями случайного n-мерного вектора $\mathbf{x}$, каждое значение которого $x \in \mathbb{R}^n$ характеризует образ конкретного объекта. Любой образ может принадлежать одному из M классов, для которых имеются индивидуальные обучающие выборки $X^{N_i} = \{x^{(i,1)},..., x^{(i,N_i)}\}$ образов $x^{(i,k)} \in \mathbb{R}^n$, $i = \overline{1,M}$, $k = \overline{1,N_i}$, в совокупности образующие общую выборку $X^N = \bigcup_{i=1}^{M} X^{N_i}$. Для каждого образа рассматриваются альтернативные гипотезы ω_i, $i = \overline{1,M}$, о принадлежности его тому или иному классу, при условии отсутствия статистических описаний классов (априорных вероятностей $p(\omega_i)$ и функций правдоподобия $p(x/\omega_i)$).

Требуется для каждого образа x принять решение, относящее объект к определённому классу, используя оценки неизвестных параметров, основанные на обработке обучающих выборок.

Априорные вероятности можно оценить как $\tilde{p}(\omega_i) = N_i / N$, где $N = \sum_{i=1}^{M} N_i$. Наличие параметрической неопределенности предполагает оценивание неизвестных параметров распределений $\tilde{p}(x/\omega_i)$ известного вида на основе соответствующей обучающей выборки X^{N_i}.

2. Оценка максимального правдоподобия

Для двух классов с гауссовскими условными плотностями, оценки неизвестных параметров распределения, в соответствии с методом максимального правдоподобия (МП), определяются следующим образом:

$$\tilde{p}(x/\omega_i) = N(x, \tilde{m}_i, \tilde{C}_i), \quad \tilde{m}_i = \frac{1}{N_i}\sum_{s=1}^{N_i} x^{(i,s)}, \quad i = \overline{1,2}, \tag{1}$$

$$\tilde{C}_i = \begin{cases} \dfrac{1}{N_i - 1}\displaystyle\sum_{s=1}^{N}(x^{(i,s)} - \tilde{m}_i)(x^{(i,s)} - \tilde{m}_i)^T, & \tilde{C}_1 \neq \tilde{C}_2, \\ \dfrac{1}{N_1 + N_2 - 2}\displaystyle\sum_{i=1}^{2}\sum_{s=1}^{N}(x^{(i,s)} - \tilde{m}_i)(x^{(i,s)} - \tilde{m}_i)^T, & \tilde{C}_1 = \tilde{C}_2. \end{cases}$$

При недостаточном объеме обучающих данных $(\min N_i < n)$ оценка ковариационной матрицы $\tilde{C}_i$ является вырожденной и требует регуляризации:

$$\tilde{C}_i = (1-\varepsilon)\tilde{C}_i + \varepsilon I,$$

где ε – малое относительно диагональных элементов $\tilde{C}_i$ положительное число; I – единичная матрица.

3. Оценка максимума апостериорной вероятности

Рассмотрим случай, в котором функции правдоподобия классов имеют вид нормального распределения с неизвестным математическим ожиданием, так же распределенным нормально. Пусть априорная плотность распределения неизвестного математического ожидания m определяется значениями m_{0_i}, C_{0_i}, а апостериорной плотности соответствуют значения m_{N_i}, C_{N_i}. Условные плотности классов и параметры их распределений, в рамках метода максимума апостериорной вероятности (МАВ), определяются как:

$$\tilde{p}(x/\omega_i) = N(x, m_{N_i}, \tilde{C}_i), \quad m_{N_i} = C_{N_i}(NC_i^{-1}\tilde{m}_i + C_{0_i}^{-1}m_{0_i}),$$

$$\tilde{C}_i = C_i + C_{N_i}, \quad C_{N_i} = C_{0_i}(C_{0_i} + \frac{1}{N_i}C_i)^{-1}\frac{1}{N_i}C_i. \tag{2}$$

4. Оценка ошибок распознавания

Оценки неизвестных параметров условных плотностей, полученные тем или иным методом, используются для определения вероятностей ошибок распознавания вычисляемых на основе известных выражений

$$\alpha = \int_{-\infty}^{l_0'} N(g'', m_{g1}, D_{g1}) dg'', \qquad \beta = \int_{l_0'}^{\infty} N(g'', m_{g2}, D_{g2}) dg''. \qquad (3)$$

В случае если известно, что распределения классов обладают одинаковой матрицей ковариации, в (3) используются следующие значения:

$$m_{g1} = \frac{1}{2}(m_1 - m_2)^T C^{-1}(m_1 - m_2) = h,$$
$$m_{g2} = -h, \qquad D_{g1} = D_{g2} = 2h,$$

где в качестве m_1, m_2 и C используются соответствующие оценки.

В случае различных матриц ковариации классов используются соотношения:

$$m_{g1} = \frac{1}{2} tr(C_2^{-1} C_1 - I) + \frac{1}{2}(m_1 - m_2)^T C_2^{-1}(m_1 - m_2) - \frac{1}{2}\ln\frac{|C_1|}{|C_2|},$$
$$D_{g1} = \frac{1}{2} tr(C_2^{-1} C_1 - I)^2 + (m_1 - m_2)^T C_2^{-1} C_1 C_2^{-1}(m_1 - m_2),$$
$$m_{g2} = \frac{1}{2} tr(I - C_1^{-1} C_2) - \frac{1}{2}(m_1 - m_2)^T C_1^{-1}(m_1 - m_2) + \frac{1}{2}\ln\frac{|C_2|}{|C_1|},$$
$$D_{g2} = \frac{1}{2} tr(C_1^{-1} C_2 - I)^2 + (m_1 - m_2)^T C_1^{-1} C_2 C_1^{-1}(m_1 - m_2).$$

Следует отметить, что в (3) вычисляются условные относительно реализаций X^{N_i}, $i = \overline{1, M}$, вероятности ошибок, усреднение которых относительно реализаций X^{N_i}, позволяет оценить безусловные вероятности.

5. Пример решения задачи

Рассмотрим пример решения задачи синтеза и анализа алгоритма распознавания двух классов образов, описываемых ГСВ с неизвестными математическими ожиданиями и одинаковой известной матрицей ковариации. Обща структура решающего правила в этом случае имеет вид

$$\frac{N_1}{N_1+N_2}\tilde{p}(x/\omega_i)\underset{\omega_2}{\overset{\omega_1}{\gtrless}}\frac{N_1}{N_1+N_2}\tilde{p}(x/\omega_i), \tag{4}$$

где $\tilde{p}(x/\omega_i)$ – плотность, вычисленная в соответствии с (1) (МП), или на основе (2) (МАВ).

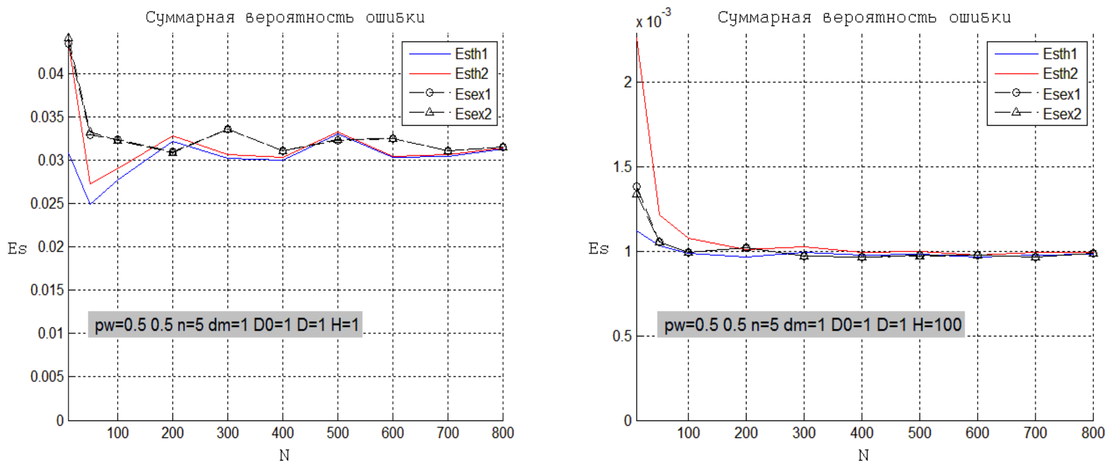
В приведенной ниже программе для получения условных относительно общей обучающей выборки вероятностей ошибок $\tilde{\alpha}(X^N)$, $\tilde{\beta}(X^N)$, а так же суммарной вероятности ошибки E_s, в выражение (4) подставляются оценки МП и МАВ неизвестных математических ожиданий классов. Априорные вероятности классов считаются известными, а объем обучающих выборок для обоих классов задается одинаковым.

Пример 1.

```
%Алгоритмы распознавания ГСВ с неизвестными математическими ожиданиями
clear all; close all;
%1.Задание исходных данных
n=5;M=2; %размерность признакового пространства и число классов
H=100;      %количество статистических испытаний процесса обучения
K=10000; %количество статистических испытаний алгоритма распознавания
%Априорные вероятности,математические ожидания и матрица ковариации классов
pw=[0.5,0.5];np=sum(pw); pw=pw/np;
D0=4; C0=D0*eye(n);   %априорные параметры распределения м.о.
dm=1; m0(:,1)=zeros(n,1); m0(:,2)=dm*ones(n,1);
m=zeros(n,2);[xm,px]=randncor(n,M,C0); %генерация значений м.о.ГСВ
m(:,1)=xm(:,1)+m0(:,1); m(:,2)=xm(:,2)+m0(:,2);
C=zeros(n);  %матрица ковариации признаков обоих классов
D=4; ro=0.7; alf=-log(ro); %дисперсия и коэффициент корреляции соседних
элементов
for i=1:n,
    for j=1:n,
        C(i,j)=D*exp(-alf*abs(i-j));
    end;
end;
C_=C^-1;
%2.Генерация обучающих данных в цикле с переменным объемом выборки
Nn=[10,50,100,200,300,400,500,600,700,800]; ln=length(Nn);%объем обучающей
выборки
Esth1=zeros(1,ln);Esth2=Esth1; Esex1=Esth1; Esex2=Esth1;
for nn=1:ln, %цикл с изменением объема обучающей выборки
    N=Nn(nn); m_=zeros(n,2); mN=zeros(n,2);
    CN=C0*(C0+C/N)^-1*(C/N); C__=(C+CN)^-1; %апостериорная матрица
ковариации
```

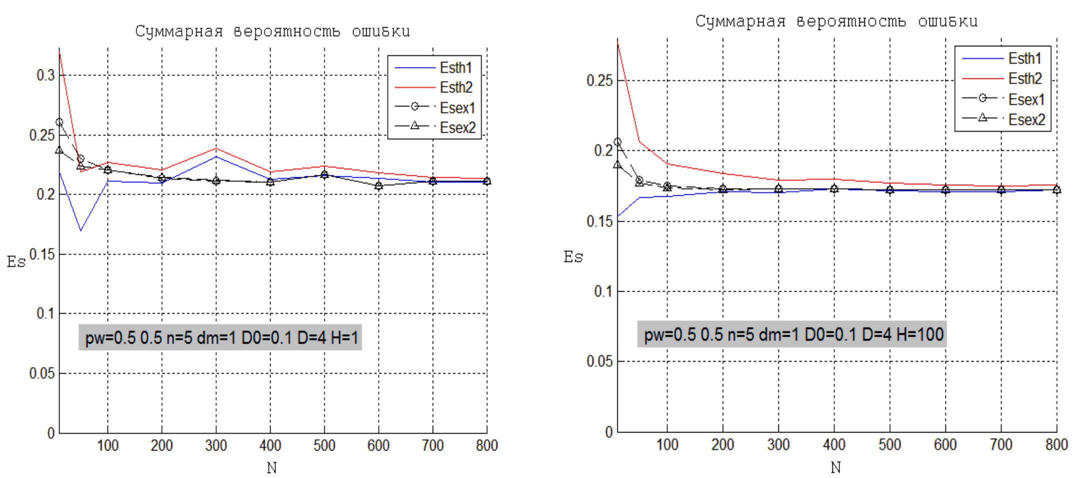
```
    for h=1:H, %цикл статистических испытаний процесса обучения
        [x1,px]=randncor(n,N,C); XN1=repmat(m(:,1),1,N)+x1; %обучающая
выборка
        [x2,px]=randncor(n,N,C); XN2=repmat(m(:,2),1,N)+x2; %обучающая
выборка
        %Обучение на основе оценки неизвестных параметров
        m_(:,1)=sum(XN1,2)/N; %оценка МП
        m_(:,2)=sum(XN2,2)/N; %оценка МП
        mN(:,1)=CN*(N*C^-1*m_(:,1)+C0^-1*m0(:,1)); %оценка МАВ
        mN(:,2)=CN*(N*C^-1*m_(:,2)+C0^-1*m0(:,2)); %оценка МАВ
        %3.Расчет разделяющих функций и вероятностей ошибок распознавания
        G1=zeros(M,n+1);G2=zeros(M,n+1);
        l0_=log(pw(2)/pw(1));
        for i=1:M,
            G1(i,1:n)=(C_*m_(:,i))';   G1(i,n+1)=-0.5*m_(:,i)'*C_*m_(:,i);
            G2(i,1:n)=(C__*mN(:,i))'; G2(i,n+1)=-0.5*mN(:,i)'*C__*mN(:,i);
        end;
        h1=0.5*(m_(:,1)-m_(:,2))'*C_*(m_(:,1)-m_(:,2)); sD1=sqrt(2*h1);
        p1_12=normcdf(l0_,h1,sD1); p1_21=1-normcdf(l0_,-h1,sD1);
        h2=0.5*(mN(:,1)-mN(:,2))'*C__*(mN(:,1)-mN(:,2)); sD2=sqrt(2*h2);
        p2_12=normcdf(l0_,h2,sD2); p2_21=1-normcdf(l0_,-h2,sD2);
        Esth1(nn)=Esth1(nn)+pw(1)*p1_12+pw(2)*p1_21;  %суммарная ошибка
        Esth2(nn)=Esth2(nn)+pw(1)*p2_12+pw(2)*p2_21;  %суммарная ошибка
        %4.Тестирование алгоритма методом статистических испытаний
        x=ones(n+1,1); Pc1_=zeros(M); Pc2_=zeros(M);%экспериментальная
матрица ошибок
        for k=1:K, %цикл по числу испытаний алгоритма распознавания
           for i=1:M, %цикл по классам
              [x_,px]=randncor(n,1,C); x(1:n,1)=m(:,i)+x_; %генерация
образа i-го класса
              u1=G1*x+log(pw'); %вычисление значения разделяющих функций
              [ui1,iai1]=max(u1); %определение максимума
              Pc1_(i,iai1)=Pc1_(i,iai1)+1; %фиксация результата
распознавания
              u2=G2*x+log(pw'); %вычисление значения разделяющих функций
              [ui2,iai2]=max(u2); %определение максимума
              Pc2_(i,iai2)=Pc2_(i,iai2)+1; %фиксация результата
распознавания
          end;
      end;
      Pc1_=Pc1_/K; Pc2_=Pc2_/K;
      Esex1(nn)=Esex1(nn)+pw(1)*Pc1_(1,2)+pw(2)*Pc1_(2,1); %суммарная
ошибка
      Esex2(nn)=Esex2(nn)+pw(1)*Pc2_(1,2)+pw(2)*Pc2_(2,1); %суммарная
ошибка
   end;%по h
end;%по nn
Esth1=Esth1/H; Esth2=Esth2/H;Esex1=Esex1/H;Esex2=Esex2/H;
%5.Визуализация зависимостей вероятностей ошибок
```

На рис.1а,б и 2а,б представлены результаты работы программы для случая $n=5$ при изменении количества испытаний H, по которым усредняется суммарная вероятность ошибки, а также при различных соотношениях D_0 (дисперсия априорного распределения компонент математических ожиданий) и D (дисперсия распределения компонент вектора признаков).



а) б)

Рис.1. Зависимости для суммарной вероятности ошибки от объема обучающих выборок при D0=1 и D=1



а) б)

Рис.2. Зависимости для суммарной вероятности ошибки от объема обучающих выборок при D0=0.1 и D=4

На рис.1 и 2 оценки суммарной вероятности ошибки, вычисленные на основе (3) с последующим усреднением по числу реализаций процесса обучения H обозначены как $Esth1$ и $Esth2$; $Esex1$ и $Esex2$ обозначают оценки суммарной вероятности ошибки, определенные на основе статистического моделирования.

Задание для самостоятельной работы

1. Реализуйте алгоритм распознавания образов, описываемых гауссовским законом распределения с параметрами, соответствующими своему варианту.
2. Вычислите значения ошибок 1-го и 2-го рода для разных объемов выборок и сравните их с теоретическими значениями (3).
3. Проведите имитационное моделирование алгоритма, в ходе которого рассчитайте значения вероятности ошибок распознавания для трех различных случаев априорных вероятностей гипотез:

   - $p(\omega_1) > p(\omega_2)$;
   - $p(\omega_1) = p(\omega_2)$;
   - $p(\omega_1) < p(\omega_2)$ (в случае трёх классов рассмотрите три произвольные комбинации сочетаний $p(\omega_1), p(\omega_2), p(\omega_3)$).

   Сравните полученные вероятности ошибок с их значениями, вычисленными теоретически.

Напишите отчёт о проделанной работе, содержащий следующие пункты:

1. Фамилию исполнителя и номер группы.
2. Название и цель лабораторной работы.
3. Исходные данные (значения параметров и априорные вероятности гипотез).
4. Код для оценок неизвестных параметров (МП и МАВ)
5. Код расчёта разделяющих функций и вероятностей ошибок распознавания
6. Представьте графики значений элементов теоретической и экспериментальной матриц вероятностей ошибок для рассмотренных случаев априорных вероятностей гипотез.

Варианты заданий:

1. m_1=[-2 3], m_2=[10 1], C=[5 -1; -1 4].
2. m_1=[2 1], m_2=[-1 1], C=[7 1; 2 6].
3. m_1=[2 -3 3], m_2=[1 1 0], C=[3 1 1; 1 3 1;1 1 3].
4. m_1=[0 -1], m_2=[-4 2], C=[3 -2; -2 3].
5. m_1=[2 2], m_2=[1 -1], C=[2 -1; -1 3].
6. m_1=[2 -3], m_2=[1 10], C=[3 -1; -1 4].
7. m_1=[10 -2], m_2=[-4 3], C=[3 1; 1 3].
8. m_1=[3 1], m_2=[-1 7], C=[3 1; 1 3].
9. m_1=[0 -1 2], m_2=[2 2 1], C=[4 -1 -1; -1 4 -1;-1 -1 4].
10. m_1=[-3 2], m_2=[0 10], C=[5 1; 1 4].
11. m_1=[5 -1], m_2=[-1 4], C=[6 2; 2 6].
12. m_1=[1 1 2], m_2=[-2 5 1], C=[4 2 2; 2 3 2;2 2 3].
13. m_1=[-1 6], m_2=[1 -4], C=[6 2; 2 6].
14. m_1=[-2 3 -3], m_2=[-1 -1 10], C=[5 -1 0; -1 5 -1;0 -1 5].
15. m_1=[1 -3], m_2=[-1 3], C=[3 1; 1 3].
16. m_1=[-12 3], m_2=[1 7], C=[5 -1; -1 5].
17. m_1=[-2 1], m_2=[1 -1], C=[3 1; 1 3].
18. m_1=[-3 2], m_2=[0 10], C=[5 1; 1 4].
19. m_1=[5 1 2], m_2=[1 4 8], C=[2 1 1; 1 2 1; 1 1 2].
20. m_1=[3 2], m_2=[0 -1], C=[5 -1; -1 5].