

Лабораторная работа № 2

РАЗДЕЛЕНИЕ ПРОСТРАНСТВА ПРИЗНАКОВ ПО КЛАССАМ

Цель работы: Ознакомиться с наглядной интерпретацией многомерных классов образов на плоскости по произвольно выбранным парам признаков. Освоить статистическое описание образов в многомерном пространстве признаков и процедуры оценки информативности признаков для их минимизации. Исследовать особенности распознавания образов по расстоянию Махаланобиса и методом разделения пространства признаков плоскостями.

Теоретические пояснения

Процесс распознавания включает в себя этап обучения, т.е. определение некоторых характеристик классов по обучающей репрезентативной выборке образов, отнесенных к известным классам, и собственно распознавание по выбранному методу на основе информации, полученной при обучении. Под распознаванием образа понимается отнесение его к конкретному классу или признание образа нераспознанным.

Для одномерных векторов, т.е. в случае распознавания по одному признаку, расстояние между точками l и p в пространстве признаков вычисляется как:

$$d_{lp} = |x_l - x_p|.$$

Расстояние между классами w_l и w_p по одному признаку, учитывающее разброс значений признака для образов, относящихся к одному классу, можно вычислить по критерию Фишера $F_{lp} = \frac{(\mu_l - \mu_p)^2}{D_l + D_p}$, (15)

где μ и D – значения математического ожидания и дисперсии классов w_l и w_p , которые вычисляются по (8) и (9) соответственно.

На рисунке 1 приведено сравнение плотностей распределения вероятности значений признака $p(x)$ для двух пар классов $w_1 - w_2$ и $w_1 - w_3$ при одинаковых значениях математического ожидания ($\mu_2 = \mu_3$) и разных дисперсиях ($D_2 < D_3$) классов w_2 и w_3 . Из рисунка 1 видно, что вероятность правильного распознавания

для пары классов $w_1 - w_2$, больше чем для пары $w_1 - w_3$, что соответствует значениям критерия Фишера для каждой из пар классов согласно (15).

Таким образом, с увеличением критерия Фишера растет вероятность правильного распознавания по данному признаку, т.е. критерий Фишера позволяет еще и оценить информативность признака для распознавания пары классов. В результате анализа оценки информативности признаков по критерию Фишера на этапе обучения системы распознавания можно исключить наименее информативные признаки, что позволит сократить объем информации о классах и время распознавания.

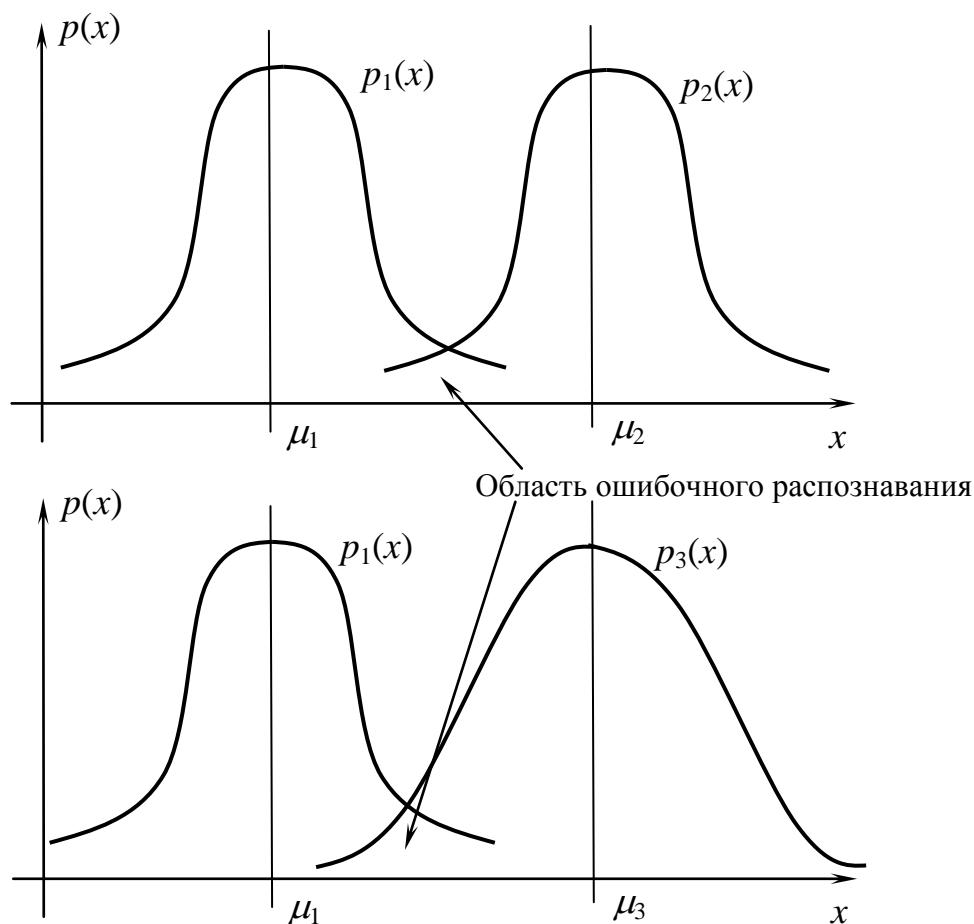


Рисунок 1 – Плотность распределения вероятности значения признака

Расстояние между распознаваемым образом и классом w_1 при одном признаке, учитывающее статистические характеристики класса, можно найти по расстоянию Фишера $d_F = \frac{(x - \mu_l)^2}{D_l}$, где x – значение признака распознаваемого образа,

μ_1 и D_1 – значения математического ожидания и дисперсии класса w_1 .

Обобщением расстояния Фишера на многомерный случай является расстояние Махаланобиса (11), позволяющее определить расстояние между распознаваемым образом и классом с учетом его статистических характеристик.

Методы распознавания в пространстве признаков можно разделить на две группы. К первой группе относятся методы, основанные на последовательном вычислении в пространстве признаков расстояний между распознаваемым образом и классами, например, с применением расстояния Махаланобиса. Образ считается принадлежащим классу, расстояние до которого минимально, если это расстояние не превышает заданного порога. В противном случае образ признается нераспознанным.

Другая группа методов предполагает предварительное разбиение признакового пространства плоскостями или поверхностями более высокого порядка на области, соответствующие известным классам, и определение принадлежности распознаваемого образа одной из областей. Например, для двух признаков производится разбиение плоскости набором прямых и для каждой области плоскости ограниченной отрезками прямых, полученных в результате их пересечения друг с другом и границами плоскости, определяется номер соответствующего класса.

Обе группы методов имеют ряд достоинств и недостатков, и выбор конкретного метода распознавания в каждом случае зависит от таких особенностей, как число признаков, требуемая надежность и скорость распознавания, технические характеристики системы получения значений признаков и т.п.

Порядок выполнения работы

1. Ввести номер варианта в программу KLAster и сформировать значения трех признаков для десяти образов нового класса с учетом представленных статистических характеристик уже имеющихся классов.
2. Проанализировать изменение значения математического ожидания и дисперсии при изменении признаков отдельных образов.

3. Записать вероятность правильного распознавания при кластеризации образов по расстоянию Махаланобиса для всех классов, включая и вновь созданный класс.
4. По критерию Фишера выбрать наиболее информативную пару признаков.
5. Для выбранной пары признаков провести разбиение признакового пространства (плоскости) на области кластеризации с максимально достигнутой вероятностью правильного распознавания.
6. Получить положительную оценку за ответы на тестовые вопросы.

Содержание отчета

1. Значения признаков для всех образов созданного класса.
2. Значения математического ожидания, дисперсии и вероятности распознавания по расстоянию Махаланобиса для всех классов.
3. Таблица информативности признаков по критерию Фишера.
4. Эскиз разбиения признакового пространства для выбранной пары признаков и соответствующие вероятности распознавания образов.
5. Выводы по сравнению двух методов кластеризации.

Контрольные вопросы

1. Чем характеризуется образ в пространстве признаков?
2. Как можно представить класс в пространстве признаков?
3. Как вычисляются статистические характеристики класса?
4. Чем являются диагональные элементы ковариационной матрицы?
5. Что такое расстояние в метрическом пространстве?
6. Как вычисляется расстояние Махаланобиса?