Линейный дискриминантный анализ (LDA)

Линейный дискриминантный анализ (LDA) является обобщением линейного дискриминанта Фишера.LDA разработан для классификации и преобразования исходного многоразмерного пространства описания объектов в пространство с более низкой размерностью, в котором максимизируется отношение дисперсии между классами к дисперсии внутри классов, тем самым гарантируется максимальная отделимость классов.Решение этой задачи сводится к построению специальных линейных функций, которые называются дискриминантными.

LDA применяется двояко, с одной строны как линейный классификаитор, а с другой, как метод уменьшения размерности описания объектов (уменьшение количества предикторов).Во второй своей роли он относиться к так назваемой группе методов контролируемого снижения размерности для реализации которого необходимо учитывать априорную принадлежность объектов к соотвествующим классам (т.е это метод с учителем).

Условия применимости и допущения LDA:

- Показатели (предикторы) объектов для использования LDA должны быть измерены в метрических шкалах (интервалов или отношений), а выходная переменная(принадлежность каждого объекта к классу) измеряется в номинальной шкале

- Необходимо чтобы число объектов p превышало число исходных предикторов n ($p - 2 > n > 0$);

- Число классов и объектов в каждом классе не менее двух;

- Предикторы должны быть линейно независимыми.

- Генеральные ковариационные матрицы классов равны между собой.

- Закон распределения для каждого класса должен является многомерным нормальным законом

Основные идеи и принципы, лежащие в основн LDA.

Линейный дискриминантный анализ - это общий термин, относящий к нескольким тесно связанным статистическим процедурам. Их можно разделить на методы интерпретации межгрупповых различий в чатности на основе уменьшения размерности исходного описания объектов и методы классификации наблюдений объектов по группам

При интерпритации необходимо ответить на вопросы: возможно ли, и спользуя данный набор показателей (предикторов), отличить один класс от другого, насколько хорошо эти показатели позволяют провести различие, можно (целесообразно) ли осуществить уменьшение размерности

исходного описания объектов и если да, то как это сделать при условии обеспечения высого качества разделения групп.

Методы, относящиеся к классификации , связаны с получением одной или нескольких функций, обеспечивающих возможность отнести объект к одной из групп. Эти функции, называемые дискременантными, зависят от значений показателей таким образом, что появляется возможность отнести каждый объект к одной из групп.

Интерпритация объектов и классов в LDA и задача уменьшения размерности.

Определение канонической дискриминантной функции (КДФ).

Важным понятием LDA является каноническая дискриминантная функция, которая является линейной комбинацией исходных предикторов (дискременантных переменных) и удовлетворяющая определенным условиям. Пусть мы имеем исходную матрицу X из n дискриминантных переменных для p объектов:

$$X = \begin{pmatrix} x_{1,1} & x_{1,2} \ldots & x_{1,\mathrm{n}} \\ x_{2,1} & x_{2,2} \ldots & x_{2,\mathrm{n}} \\ x_{p,1} & x_{p,2} \ldots & x_{p,n} \end{pmatrix} \qquad (1)$$

Тогда каноническая дискременантная функция имеет следующнн математическое выражение :

$$f_{km} = u_0 + u_1 x_{1km} + u_2 x_{2km} + \ldots + u_n x_{1km} \qquad (2)$$

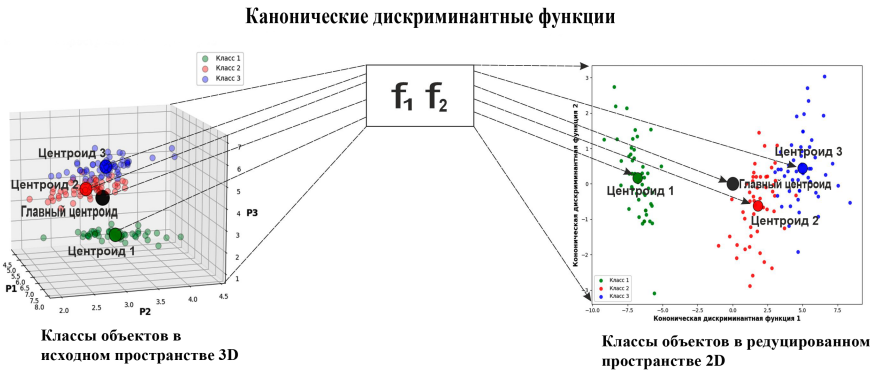
где f_{km} - значение канонической дискреминантной функции для m - го объекта в k -ой группе; x_{ikm} - m - значение i -ой переменной в группе k;u_i -коэффициенты, обеспечивающие выполнение необходимых условий. Коэффициенты для первой функции выбираются такм образом, чтобы обеспечить масимальные различия ее средних значений для разнык классов. То есть она должна обеспечить "максимальные различия между классами"(как это сделать формально определим ниже). Коэффициенты для второй функции выбираются из тех же соображений, но дополнительно должно быть обеспечено условие, что значения второй функции не коррелируют со значениями первой. Аналогично определяется третья функция, которая не должна коррелирова с первой и второй и .т.д Максимальное число дискременантныхз функций, которые могут быть получены таким способом, равно числу классов без единицы или чилу дискреминантных переменных если $n < k - 1$

Геометрическая интерпритация и число канонических дисриминантных функций

Пусть дискриминантные переменные-оси n - мерного евклидова пространства. Каждый объект (наблюдение) представляет собой точку с n координатами в этом пространстве. Если классы отличаются друг от друга, то их можно представить как скопление точек в некоторых областях данного пространства. При этом области занимаемые точками разных классов как не пересикаться так и пересекаться друг с другом. Для определения положения классов можно вычислить их "центроид".Центроид класса есть воображаемая точка, координаты которой есть средние значения переменных в этом классе.Центроиды можно использовать для оценки различий между классами. При этом оказывается, для того чтобы различать относительное положение центроидов не нужна большая размерность описания объектов. Как правило, можно ограничиться размерностью число классов бех единицы.

Роль числа классов становиться понятной, если обратитьсяк к особенностям топологии многооразия с евклидовой метрикой. Для любых пространств, где справедливы аксиомы евклидовой геометрии, две точки определяют прямую, три точки -плоскость, четыре - трехмерную поверхность и.т.д. Принцип состоит в том точки определяют размерность пространства равную числу точек без единицы. Цетроиды задают пространство размерность которого на единицу меньше их числа и в котором отображаются все объекты в виде соотвествующих точек. В этом пространстве можно разместить новую систему коодинат с началом (нулевой точкой) задаваемым, так называемым,"главным центроидом". Главный центроид занимает положение, определяесое средними значениями координат объектов по всей их совокупности. Относительно этого центра существует бесконечное множество положений новых коодинатных осей при условии, что они пренадлежат пространству "натянутому на центроиды". Если напрвить первуюось под углом, при котором средние значения классов разделяются в большей степени чем для любого другого направления,, то эта ось должна быть самой важно для дискриминация классов. Если есть два и более классов, можно оарентировать вторую ось тоже таким образом,чтобы обеспечить максимальное разделение классов и при этом вторая ось должна быть перпендикулярна первой и.т.д.

Расположение осей по такому принципу приводит к критерию для канонических дискриминантных функций.Соотношения (2)задают преобразования n-мерного пространства дискриминантных переменных в q



Графическая интерпретация использования канонических дискриминантных функций для контролируемого сокращения размерности (метод LDA).

Рис. 1: Геометрическая интепритация процесса сокращения размерности

-мерное пространство дискреминантных функций (где q- максимальное число функций). Каждой оси соотвествует свое соотношение(2) , тогда каждые объект будет точкой в этом новом пространсте с координатами равными соотвествующим значения канонической дискриминантной функкции. Графически описанный выше процесс сокращения размерности исходного пространства может быть представлен следующим образом (рис.1):

В том случае когда число переменных n меньше, числа классов, максимальное число функций $q = n$. В этом случае преобразования пространства с большей размерностью в пространство с меньшей размерность не происходит, нужно только сделать замену координат, для удовлетворения некоторому критерию.

Получение коэффициентов КДФ.

Рассмотрим основные принципы получения коэффициентов u_i канонической дискриминантной функции. Прежде всего, необходим статистический метод для измерения степени различия между объектами. Таблицы групповых средних и стандартных отклонений для этого недостаточно. Для этих целей можно использовать следующую квадратную симмитричную матрицу : $T = (t_{ij})^{n \times n}$. Элементы этой матрицы задаются

соотношением:

$$t_{ij} = \sum_{k=1}^{g} \sum_{m=1}^{p} (x_{ikm} - X_{i..})(x_{jkm} - X_{j..}) \qquad (3)$$

Здесь

g -число классов;

p - общее число наблюдений по всем классам;

x_{ikm} -величина переменной i для m-го наблюдения в k-ом классе;

$X_{i..}$ -средняя величина переменной i по всем классам (общее среднее)

Выражения в скобках (3) представляют собой отклонения значения переменных от общего среднего. При $i = j$ сомножители равны и мы имеем кадрат этого отклонения., поэтому на диагонали матрицы T будут распологаться суммы квадратов отклонения соотвествующей переменной от общего среднего. При $i \neq j$ получаем сумму произведений отклонения одной переменной на отклонение другой.Таким образом, в целом данная матрица дает полную информацию о распределении точек по пространству, определяемому переменными. Если разделить каждый элемент матрицы на $n-1$, то получиться ковариационная матрица. Матрицу T также легко приобразовать в корреляционную, поделив каждый элемент матрицы на квадратный корень произведения двух соотвествующих диагональных элементов.

Для измерения разброса внутри классов служит матрица W , которая отличается от T только тем, что ее элементы определяются средними значеними переменных для отдельных классов:

$$w_{ij} = \sum_{k=1}^{g} \sum_{m=1}^{p_k} (x_{ikm} - X_{ik.})(x_{jkm} - X_{jk.}) \qquad (4)$$

здесь

p_k - число наблюдений в классе k;

$X_{ik.}$ -средняя величина переменной i в классе k.

Если разделить каждый элемент матрицы на $n - g$, то получиться внутригрупповая ковариационная матрица, представляющая собой взвешанное среднее ковариационных матриц отдельных классов. Ее также можно преврать в соотвествующую корреляционной матрицу.

Если центроиды классов совпадают друг с другом, то матрицы T и W также совпадают. Если центроиды у классов разные, то элементы

матрицы W будут меньше соотвествующих элентов матрицы T. Тогда вводится матрица B этих различий: $B = T - W$ $(b_{ij} = t_{ij} - w_{ij})$.

Матрица B называется межгрупповой суммой квадратов отклонений и попарных произведений. Матрица B по отношению к матрице W дает меру различия между группами. В совокупности эти матрицы содержат всю основную информацию о зависимости внутри групп и между группами

Можно записать следующее критерий для нахождения коэффициентов канонических дискриминантных функций f:

$$\lambda = B(f)/W(f) \to \max_f \qquad (5)$$

Это означает, что для лучшего разделения наблюдений на группы нужно подобрать коэффициенты канонической дискриминантной функции из условия максимизации отношения межгрупповой матрицы рассеяния к внутригрупповой матрице рассеяния (5) при условии ортогональности дискриминантных плоскостей.

Тогда нахождение коэффициентов дискриминантных функций сводится к решению задачи о собственных значениях и векторах. Это утверждение можно сформулировать так:

если спроектировать g групп n–мерных выборок на $g-1$-мерное пространство, порожденное собственными векторами $(\nu_{1k}, ..., \nu_{nk}), k = 1, ...g-1$, то отношение (5) будет максимальным, т.е. рассеивание между группами будет максимальным при заданном внутригрупповом рассеивании. Если бы мы захотели спроектировать все группы на прямую при условии предельной максимизации рассеивания между ними, то следовало бы использовать собственный вектор, соответствующий максимальному собственному числу λ_1. При этом для канонических дискриминантных функций можно получать нестандартизованные и стандартизованные коэффициенты.

Нестандартизованные коэффициенты.

Пусть $\lambda_1 \geq \lambda_2 .... \lambda_n$ и $\nu_1, \nu_2 .... \nu_n$ соответственно собственные значения и векторы. Тогда условие (5) в терминах собственных чисел и векторов запишется в виде $\lambda = \frac{\sum\limits_k b_{jk}\,\nu_j\,\nu_k}{\sum\limits_k w_{jk}\,\nu_j\,\nu_k}$, а это влечет $\sum\limits_k (b_{jk} - \lambda\, w_{jk})\,\nu_k = 0$ или в матричной записи:

$$(B - \lambda W)\,\nu_i = 0 \quad (6)$$

При этом $\nu_i^T W \nu_j = \delta_{ij}$, где δ_{ij} - символ кронера. Таким образом, решение уравнения $|B - \lambda W| = 0$ позволяет определить компоненты собственных векторов, соответствующих дискриминантным функциям. Если B и W невырожденные матрицы,то собственные корни уравнения $|B - \lambda W| = 0$ такие же, как и у $|W^{-1} B - \lambda I| = 0$.Решение системы уравнений (6) можно получить путем использования разложения Холецкого (метод квадратных корней) $W^{-1} = LL^T$ и решения задачи о собственных значениях имеем:

$$(L^T B L - \lambda_i I) \nu_i , \quad (7)$$

При $\nu_i^T \nu_i = \delta_{ij}$. Каждое решение системы (7), которое имеет свое собственное значение λ_i и собственный вектор ν_i , соответствует одной дискриминантной функции. Компоненты собственного вектора ν_i можно использовать в качестве коэффициентов дискриминантной функции. Однако при таком подходе начало координат не будет совпадать с главным центроидом. Для того, чтобы начало координат совпало с главным центроидом нужно нормировать компоненты собственного вектора:

$$u_i = \nu_i \sqrt{p - g}, \quad u_0 = \sum_{i=1}^{n} u_i X_{i..} \quad (8)$$

Нормированные коэффициенты (8) получены по нестандартизованным исходным данным, поэтому они называются нестандартизованными. Нормированные коэффициенты приводят к таким дискриминантным значениям, единицей измерения которых является стандартное квадратичное отклонение. При таком подходе каждая ось в преобразованном пространстве сжимается или растягивается таким образом, что соответствующее дискриминантное значение для данного объекта представляет собой число стандартных отклонений точки от главного центроида.

Стандартизованные коэффициенты

Их можно получить двумя способами:

- по формуле (8), если исходные данные были приведены к стандартной форме;

- преобразованием нестандартизованных коэффициентов к стандартизованной форме:

$$c_i = u_i \sqrt{\frac{w_{ii}}{p - g}}, \quad (9)$$

где w_{ii} сумма внутригрупповых квадратов i й переменной, определяемой по формуле (4).

Стандартизованные коэффициенты полезно применять для уменьшения размерности исходного признакового пространства переменных. Если абсолютная величина коэффициента для данной переменной для всех дискриминантных функций мала, то эту переменную можно исключить, тем самым сократив число переменных.

Структурные коэффициенты определяются коэффициентами взаимной корреляции между отдельными переменными и дискриминантной функцией. Если относительно некоторой переменной абсолютная величина коэффициента велика, то вся информация о дискриминантной функции заключена в этой переменной.

Структурные коэффициенты

Они определяются коэффициентами взаимной корреляции между отдельными переменными и канонической дискриминантной функцией. Если относительно некоторой переменной абсолютная величина коэффициента велика, то вся информация о дискриминантной функции заключена в этой переменной.

Данные коэффициенты полезны при классификации групп. Структурный коэффициент можно вычислить и для переменной в пределах отдельно взятой группы. Тогда получаем внутригрупповой структурный коэффициент, которыйвычисляется по формуле:

$$s_{ij} = \sum_{k=1}^{n} r_{ik}\, c_{kj} = \sum_{k=1}^{n} \frac{w_{ik}\, c_{kj}}{\sqrt{w_{ii}\, w_{jj}}} \qquad (10)$$

где s_{ij} внутригрупповой структурный коэффициент для i–ой переменной и j–ой функции; r_{ik}?-?внутригрупповые структурные коэффициенты корреляции между переменными i и k;? c_{kj}?- стандартизованные коэффициенты канонической функции для переменной k и функции j.

Структурные коэффициенты по своей информативности несколько отличаются от стандартизованных коэффициентов. Стандартизованные коэффициенты показывают вклад переменных в значение дискриминантной функции. Если две переменные сильно коррелированы, то их стандартизованные коэффициенты могут быть меньше по сравнению с теми случаями, когда используется только одна из этих переменных. Такое распределение величины стандартизованного коэффициента объясняется тем, что при их вычислении учитывается влияние всех перемен-

ных.Структурные же коэффициенты являются парными корреляциями и на них не влияют взаимные зависимости прочих переменных.

Число канонических дискриминантных функций

Общее число КДФ не превышает числа дискриминантных переменных и, по крайней мере, на 1 меньше числа групп. Степень разделения выборочных групп зависит от величины собственных чисел: чем больше собственное число, тем сильнее разделение. Наибольшей разделительной способностью обладает первая дискриминантная функция, соответствующая наибольшему собственному числу λ_1, вторая обеспечивает максимальное различение после первой и т.д. Различительную способность i–ой функции оценивают по относительной величине в процентах собственного числа λ_iот суммы всех λ.

Коэффициент канонической корреляции.

Другой характеристикой, позволяющей оценить полезность дискриминантной функции является коэффициент канонической корреляции . Каноническая корреляция r_i является мерой связи между двумя множествами переменных. Максимальная величина этого коэффициента равна 1. Будем считать, что группы составляют одно множество, а другое множество образуют дискриминантные переменные. Коэффициент канонической корреляции для i–ой дискриминантной функции определяется формулой:

$$r_i = \sqrt{\frac{\lambda_i}{1+\lambda_i}} \qquad (11)$$

Чем больше величина r_i, тем лучше разделительная способность дискриминантной функции.

Остаточная дискриминация.

Так как дискриминантные функции находятся по выборочным данным, они нуждаются в проверке статистической значимости. Дискриминантные функции представляются аналогично главным компонентам. Поэтому для проверки этой значимости можно воспользоваться критерием, аналогичным дисперсионному критерию в методе главных компонент. Этот критерий оценивает остаточную дискриминантную способность, под которой понимается способность различать группы, если при этом исключить информацию, полученную с помощью ранее вычисленных функций. Если остаточная дискриминация мала, то не имеет смысла дальнейшее вычисление очередной АДФ. Полученная статистика носит название "Статистика Уилкса"и вычисляется по формуле:

$$\Lambda = \prod_{i=k+1}^{g} (\frac{1}{1+\lambda_i}) \qquad (12)$$

где k?—?число вычисленных функций. Чем меньше эта статистика, тем значимее соответствующая дискриминантная функция.

Величина: $\chi^2 = -(p - (\frac{n+g}{2} - 1) \cdot \ln \Lambda_k, \qquad k = 0, 1, ...g - 1$ имеет хи–квадрат распределение с $(n-k)$ $(g-k-1)$ степенями свободы.

Вычисления проводим в следующем порядке:

- Находим значение критерия χ^2 при k=0. Значимость критерия подтверждает существование различий между группами. Кроме того, это доказывает, что первая дискриминантная функция значима и имеет смысл ее вычислять.

- Определяем первую дискриминантную функцию и проверяем значимость критерия при $k = 1$. Если критерий значим, то вычисляем вторую дискриминантную функцию и продолжаем процесс до тех пор, пока не будет исчерпана вся значимая информация.

Процедуры классификации.

Выше мы рассматривали задачу интерпретации, которая связана с определением числа и значимости дискриминантных функций и с выяснением их значения для объяснения различий между классами. Классификация это особая задача, в которой либо дискриминантные переменные, либо канонические дискриминантые функции используются для предсказания класса, к которому наиболее вероятно пренадлежит некоторый объект. Существует несколько процедур классификаци, но все они сравнивают положение объекта относительно цетроидов всех классов.

Классифицирующие функции.

Методы классификации обычно требуют определения понятия "расстояния"между объектом и каждым центроидом группы и отненсения его к ближайшей группе. При этом может использоваться редуцированное пространство канонических дискреминантных функций или исходное пространство дискриминантных переменных (предикторов). В этом случае описанный выше дискриминантный анализ может вообще не использоваться.

Простые классифицирующие функции.

Первый классификатор на основе линейной комбинации дискриминантных переменных придложил Фишер (1936). Он предложил линейную функцию, которая максимизирует различие меду классами и мини-

мизирует дисперсию внутри классов. В результате формируется линейная классифицирующая функция для каждого класса. Она имеет следующий вид:

$$h_k = b_{k0} + b_{k1}x_1 + ... + b_{kn}x_n \quad (13)$$

где h_k -значение функции для класса k, а b_{ki}-коэффициенты, которые нужно определить.

Объект относится к классу с наибольшим значением h_k. Коэффициенты определяются слудующим образом:

$$b_{ki} = (p - g) \sum_{j=1}^{n} a_{ij}\, x_{jk}\ , \quad (14)$$

где b_{ki}- коэффициент для переменной i в выражении , соотвествующему классу k, а a_{ij} -элемент матрицы обратной к внутригруппой матрице попарных произведений W. Постоянный член определяется следующим образом:

$$b_{k0} = -0.5 \sum_{j=1}^{n} a_{ij}\, x_{jk}\ , \quad (15)$$

Коэффициенты классифицирующей функции обычно не интерпритируются, потому что они не стандартизованы и каждому классу соответствует своя функция. Точные значения функции не имеет значения важно только определить наибольшую функцииюи отнести объект к классу, которому соответствует эта наибольшая функция. Функции вида (13) называются "простыми классифицирующими функциями"потому,что они предпологают лишь равенство групповых ковариационных матриц и не требуют дополнительных свойств.

Обобщенные функции расстояния и классификация на их основе.

Выбор функций расстояния между объектами для классификации является наиболее очевидным способом введения меры сходства для векторов объектов, которые интерпретируются как точки в евклидовом пространстве. В качестве меры сходства можно использовать евклидово расстояние между объектами. Чем меньше расстояние между объектами, тем больше сходство. Однако в тех случаях, когда переменные коррелированы, измерены в разных единицах и имеют различные стандартные

отклонения, трудно четко определить понятие "расстояния". В этом случае полезнее применить не евклидовое расстояние, а выборочное расстояние Махаланобиса:

$$D^2(x/\,G_k) = (p-g)\cdot\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\,(\,x_i - X_{ik.}\,)\,(\,x_j - X_{jk.}\,)\;, k = 1, ..., g \qquad (16)$$

где $D^2(x|G_k)$ -квадрат расстояния от точки x до центроида класса k. В матричной записи имеем:

$$D^2(x|\,G_k) = (p-g)\cdot(X - \overline{X_k})^T\,W^{-1}\,(\,X - \overline{X_k}\,),\; k = 1, ..., g \qquad (17)$$

где X -объект(точка) с n координатами, $\overline{X_k}$ -вектор средних для точек k-ой группы.

Если вместо W использовать оценку внутригрупповой ковариационной матрицы $\overset{*}{\Sigma} = W/(p-g)$, то получим:

$$D^2(x|\,G_k) = (X - \overline{X_k})^T\,\overset{*}{\Sigma}^{-1}\,(\,X - \overline{X_k}\,),\; k = 1, ..., g \qquad (18)$$

При использовании функции расстояния, объект относят к той группе, для которой расстояние D^2 наименьшее.

Относя объект к ближайшему классу в соответствии с D^2, мы неявно приписываем его к тому классу, для которого он имеет наибольшую вероятность принадлежности $Pr(x|G_k)$. Если предположить, что любой объект должен принадлежать одной из групп, то можно вычислить вероятность его принадлежности для любой из групп:

$$\Pr(G_k\,|x) = \frac{\Pr(x|G_k)}{\sum\limits_{i=1}^{g}\Pr(x|G_i)} \qquad (19)$$

Объект принадлежит к той группе, для которой апостериорная вероятность $\Pr(G_k\,|x)$ максимальна, что эквивалентно использованию наименьшего расстояния.

До сих пор при классификации по D^2 предполагалось, что априорные вероятности появления групп одинаковы. Для учета априорных вероятностей нужно модифицировать расстояние , вычитая из выражений (16-18) удвоенную величину натурального логарифма от априорной вероятности q_k. Тогда, вместо выборочного расстояния Махаланобиса (18), получим:

$$D_q^2(x|\, G_k) = (\, X - \overline{X_k}\,)^T \overset{*}{\Sigma}{}^{-1} (\, X - \overline{X_k}\,) - 2\ln(q_k) \quad (20)$$

Это изменение расстояния математически идентично умножению величины $Pr(x|G_k)$ на априорную вероятность группы q_k.

Отметим, тот факт, что априорные вероятности оказывают наибольшее влияние при перекрытии групп и, следовательно, многие объекты с большой вероятностью могут принадлежать ко многим группам. Если группы сильно различаются, то учет априорных вероятностей практически не влияет на результат классификации, поскольку между классами будет находиться очень мало объектов.

Отбор дискриминантных переменных

Часто при классификации сталкиваются с проблемой отбора дискриминантных переменных. В качестве механизма отбора переменных используется процедура последовательного отбора. Не в даваясь в технику этой процедуры следует отметить, что для ее реализации требуется использовать соответствующий критерий отбора. Существует достаточно много таких критериев: Λ - статистика Уилкса (рассматривалась нами выше), V-статисика Рао, квадрат расстояния Махаланобиса между ближайшими соседями, межгрупповая F-статисика, минимизация остаточной дисперсии и некоторые другие. Приведем некоторую информацию по получению и использованию V-статисики Рао

V–статистика Рао.

Рао, применяя расстояние Махаланобиса, построил статистику, которая является мерой общего разделения классов.

Эта мера, известная как V–статистика Рао, измеряет расстояния от каждого центроида группы до главного центроида с весами, пропорциональными объему выборки соответствующей группы. Она применима при любом количестве классов и может быть использована для проверки гипотезы $H_0 : \mu_1 = ... = \mu_g$ о равенстве математических ожиданий для всех классов. Если гипотеза H_0 верна, а объемы выборок p_i стремятся к ∞ , то распределение величины V стремится к χ^2 $n(g-1)$ с степенями

свободы. Если наблюдаемая величина $\chi^2 > \chi^2_{1-\alpha}(n(g-1))$, то гипотеза H_0 отвергается. V–статистика вычисляется по следующей формуле в матричном виде:

$$V = \sum_{k=1}^{g} n_k \, (\, \overline{X_k} - \overline{X} \,)^T \, \overset{*}{\Sigma}^{-1} \, (\, \overline{X_k} - \overline{X} \,) \quad (21)$$

Отметим, что при включении или исключении переменных V–статистика имеет распределение хи–квадрат с числом степеней свободы, равным $g-1$, умноженное на число переменных, включенных (исключенных) на этом шаге. Если изменение статистики не значимо, то переменную можно не включать. Если после включения новой переменной V–статистика оказывается отрицательной, то это означает, что включенная переменная ухудшает разделение центроидов.

Классификационная матрица

Так как LDA часто используется как линейный классификатор с учителем, то применив его к объектам на которых строилось правило классификации можно оценить по доле правильно классифицированных объектов точность процедуры классификаци. Обычно строится специальная таблица, которая называется классфикационной матрицей K. Строки этой матрицы исходные классы, столбцы предпологаемые классы. Тогда матрица K это квадратная матрица размерности gxg, в которой диагональные элементы содержат количество правильно классифицированных объектов а остальные элементы равны числу ошибочных классификаций. На основе этой информации вычисляются различные показатели характерезующие качество работы классификатора. При этом часто все объекты делятся на основе рандомизованных методов на обучающуб выборку и тестовую. На обучающей строится классификационные правила, а на тестовой оценивается качество этих правил в частности именной для тестовой последовательности строится и исследуется классификационная матрица.

*originalsource*1

*originalsource*2

Реализация метода *LDA* **на** *Python* **и** *R*.

На python и R имеются пакеты, реализующие метод LDA. Рассмотрим примеры их использования для решения задачи управляемого сокращения размерности исходного описания и параллельно задачи классификации с учителем. Для python в качестве исходных данных используем два

набора. Первый изветные всем данные о ирисах, в которых мы для наглядгости перешли от исходного пространства з 4-мерной размерности к 3-мерной, удалив из рассмотрения один их предикторов(дискременантных переменных). Второй набор это уже использовавшиеся ранее данные о российских регинов, для которых априори задано три типа экономики: AT -аграрноторговый, P-промышленный, D-добывающий.

Ниже приведен код python для второго набора данных (для перого набора код аналогичен,он отлличается только в части ввода данных)

```python
from pandas import Series, DataFrame
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
from sklearn import datasets
from sklearn.decomposition import PCA
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis

result =pd.read_table('dan04.txt',sep='\s+')
print(result)
clnam=['AT','P','D']
target_names =np.array (clnam)

y=result['P18']
X =DataFrame(result , columns=['P10','P11','P14'])
y=np.array(y)
X=np.array (X)

SUMX=X[:,0] +X[:,1] + X[:,2]
X[:,0] =X[:,0] /SUMX
X[:,1] =X[:,1] /SUMX
X[:,2] =X[:,2] /SUMX

X1=DataFrame(X,columns=['P10','P11','P14'] )
classy=DataFrame(y, columns=['Class'])
XX = X1.join(classy)

XI1=XX.query("Class in [0]")
```

```
XI2=XX.query("Class in [1]")
XI3=XX.query("Class in [2]")

XI1 =DataFrame(XI1, columns=['P10','P11','P14'])
XI1=np.array (XI1)
XI2 =DataFrame(XI2, columns=['P10','P11','P14'])
XI2=np.array (XI2)
XI3 =DataFrame(XI3, columns=['P10','P11','P14'])
XI3=np.array (XI3)



lda = LinearDiscriminantAnalysis(n_components=2,
 store_covariance = True)
X_r2 = lda.fit(X, y).transform(X)

print ('======= КОЭФФИЦИЕНТЫ=====================')
coefd=lda.coef_
print (lda.coef_)
print ('==========================================')
#print( lda.covariance_)
#print( lda.explained_variance_ratio_)

print ('=======ИСХОДНОЕ ПРОСТРАНСТВО 3D   ==========')
print ('=======ЦЕНТРОИДЫ КЛАССОВ ================')
centroid =lda.means_
print( lda.means_)
print (==========================================')

print ('=======ГЛАВНЫЙ ЦЕНТРОИД ==================')
Gcentr=lda.xbar_
print(lda.xbar_)
print ('==========================================')
centr2D=lda.scalings_
#print(lda.scalings_)

D2mean01 =np.mean(X_r2[y == 0, 0])
D2mean02 =np.mean(X_r2[y == 0, 1])
```

```
D2mean11 =np.mean(X_r2[y == 1, 0])
D2mean12 =np.mean(X_r2[y == 1, 1])

D2mean21 =np.mean(X_r2[y == 2, 0])
D2mean22 =np.mean(X_r2[y == 2, 1])

print ('=== РЕДУЦИРОВАННОЕ ПРОСТРАНСТВО 2D ======')

print ('=======ЦЕНТРОИДЫ КЛАССОВ================')
print(D2mean01,D2mean02 )
print(D2mean11,D2mean12 )
print(D2mean21,D2mean22 )
print ('========================================')

#построение исходного графика 3 D
fig = plt.figure(1, figsize=(12, 10))
ax = fig.add_subplot(111, projection='3d')

xs1 = XI1[:,0]
ys1 = XI1[:,1]
zs1 = XI1[:,2]

xs2 = XI2[:,0]
ys2 = XI2[:,1]
zs2 = XI2[:,2]

xs3 = XI3[:,0]
ys3 = XI3[:,1]
zs3 = XI3[:,2]

ax.scatter(xs1, ys1, zs1,c='g', alpha=0.3, cmap=plt.cm.nipy_spectral,
edgecolor='k',s=80,label= target_names [0])
ax.scatter(xs2, ys2, zs2,c='r',alpha=0.3, cmap=plt.cm.nipy_spectral,
edgecolor='k',s=80,label= target_names [1])
ax.scatter(xs3, ys3, zs3,c='b',alpha=0.3, cmap=plt.cm.nipy_spectral,
edgecolor='k',s=80,label=target_names [2])
ax.scatter(Gcentr[0], Gcentr[ 1] ,Gcentr[ 2], marker='o',
```

```
c='black', alpha=0.9, s=500, edgecolor='k')
ax.scatter(centroid[:, 0], centroid[:, 1], centroid[:, 2],marker='o',
c=['g', 'r','b'],alpha=0.9, s=400, edgecolor='k')

#for i, c in enumerate(centroid):
        #print(str(i) + '-ое значение равно ' + str(c))
        #ax.scatter(c[0], c[1], c[2], marker='$%d$' % i, alpha=1,
                    #s=250, edgecolor='k')
ax.text(centroid[2, 0],centroid[2, 1],centroid[2, 2]+0.05,
'Центроид 3',fontsize=12, fontweight='bold')
ax.text(centroid[1, 0],centroid[1, 1],centroid[1, 2]+0.05,
'Центроид 2',fontsize=12, fontweight='bold')
ax.text(centroid[0, 0],centroid[0, 1],centroid[0, 2]+0.05,
'Центроид 1',fontsize=12, fontweight='bold')
ax.text(Gcentr[0],Gcentr[1],Gcentr[2] +0.06,
'Главный центроид',fontsize=12, fontweight='bold')
ax.set_title("Классы объектов в исходном пространстве 3D",
fontsize=16, fontweight='bold')
ax.set_xlabel('P10',fontsize=14, fontweight='bold')
ax.set_ylabel('P11',fontsize=14, fontweight='bold')
ax.set_zlabel('P14',fontsize=14, fontweight='bold')
ax.legend(loc="best")



plt.figure(2,figsize=(12, 10))
colors = ['g', 'r', 'b']
lw = 2

for color, i, target_name in zip(colors, [0, 1, 2], target_names):
    plt.scatter(X_r2[y == i, 0], X_r2[y == i, 1], alpha=.8, color=color,
                label=target_name)
plt.scatter(D2mean01, D2mean02, marker='o',
c='g', alpha=0.9, s=400, edgecolor='k')
plt.scatter(D2mean11, D2mean12, marker='o',
c='r', alpha=0.9, s=400, edgecolor='k')
plt.scatter(D2mean21, D2mean22, marker='o',
c='b', alpha=0.9, s=400, edgecolor='k')
```

```
plt.scatter(0.0, 0.0, marker='o',
c='black', alpha=0.8, s=600, edgecolor='k')

plt.text(D2mean21+0.2,D2mean22+0.2,'Центроид 3',
fontsize=14, fontweight='bold')
plt.text(D2mean11+0.2,D2mean12+0.2,'Центроид 2',
fontsize=14, fontweight='bold')
plt.text(D2mean01-0.4,D2mean02+0.2,'Центроид 1',
fontsize=14, fontweight='bold')
plt.text(-0.8,0.2,'Главный центроид',
fontsize=14, fontweight='bold')
plt.legend(loc='best', shadow=False, scatterpoints=1)
plt.title('Классы объектов в редуцированном пространсте 2D',
fontsize=14, fontweight='bold' )
plt.xlabel('Каноническая дискриминантная функция f1',
fontsize=12, fontweight='bold')
plt.ylabel('Каноническая дискриминантная функция f2',
fontsize=12, fontweight='bold')
plt.show()

'''
```

Результаты для первого набора данных имеют следующий вид (рис.2,3):

По второму набору нанных получены следующие результаты (рис.4,5):

Пример реализации LDA на R связан со вторым набором данных и здесь мы достаточно подробно реализовали решение задачи классификации, рассчитав таблицу сопряженности (классификационная матрица).

Предлогаем Вашему вниманию следующий скрипт на R:

```r
library(openxlsx)
library(gmodels)
library(lattice)
library(vegan)
library(grDevices)
library(reshape2)
library(ggplot2)
library(ggthemes)
```

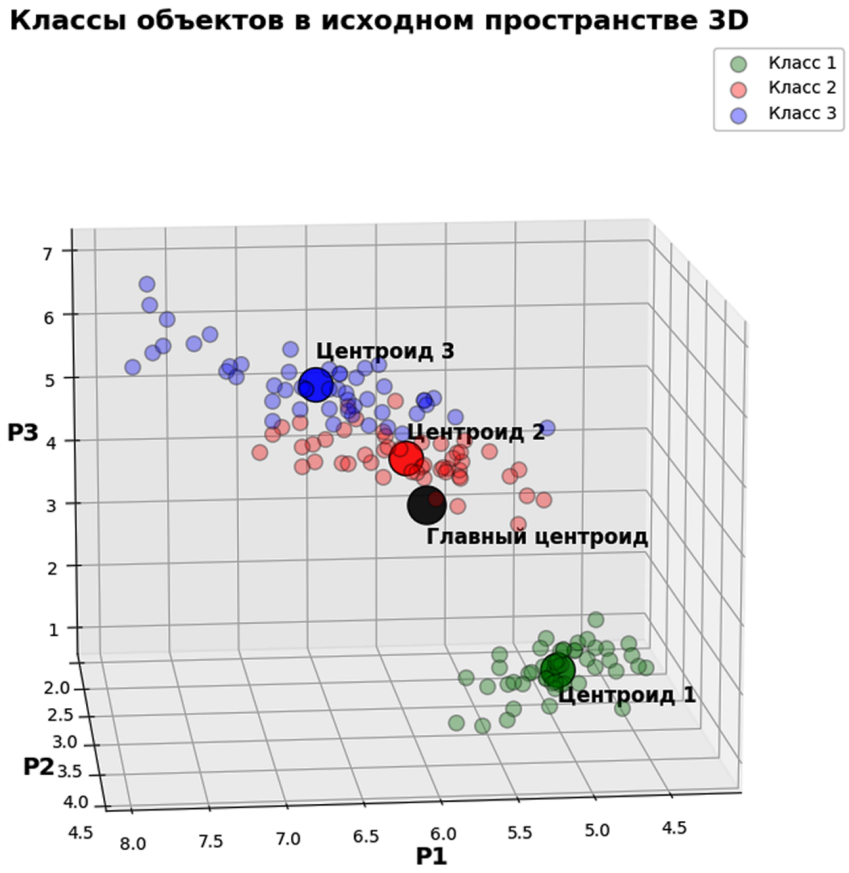


Рис. 2: Данные в исходном трехмерном пространстве 1

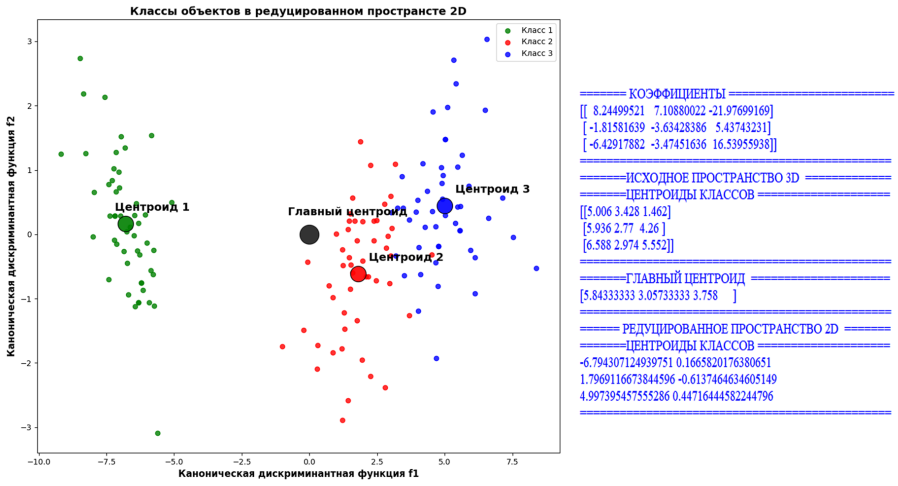


Рис. 3: Данные в редуцированном двухмерном пространстве

```
#ЗАГРУЗКА ИЗ EXSEL ПЕРВИЧНЫХ ДАННЫХ
N=1 # номер листа с данными
x <-read.xlsx("D:/Machine learning/DATA/Region_Data5.xlsx", sheet = N)
#x

#ФОРМИРОВАНИЕ data.frame ОСНОВНЫЕ
СОЦИАЛЬНО-ЭКОНОМИЧЕСКИЕ ПОКАЗАТЕЛИ РЕГИОНОВ РФ
xreg <-data.frame(P1=x$Pok_001,P2=x$Pok_002,
P3=x$Pok_003,P4=x$Pok_004,
P5=x$Pok_005,P6=x$Pok_006,P7=x$Pok_007,
P8=x$Pok_008,P9=x$Pok_009,P10=x$Pok_010,
P11=x$Pok_011,P12=x$Pok_012,P13=x$Pok_013,
P14=x$Pok_014,P15=x$Pok_015,P16=x$Pok_016,
P17=x$Pok_017,P18=x$Pok_018)
#xreg
#Формирование фактора тип экономики региона
xreg_f <- as.factor(x$Pok_018)

#выбор исследуемого подмножества показателей регионов
```

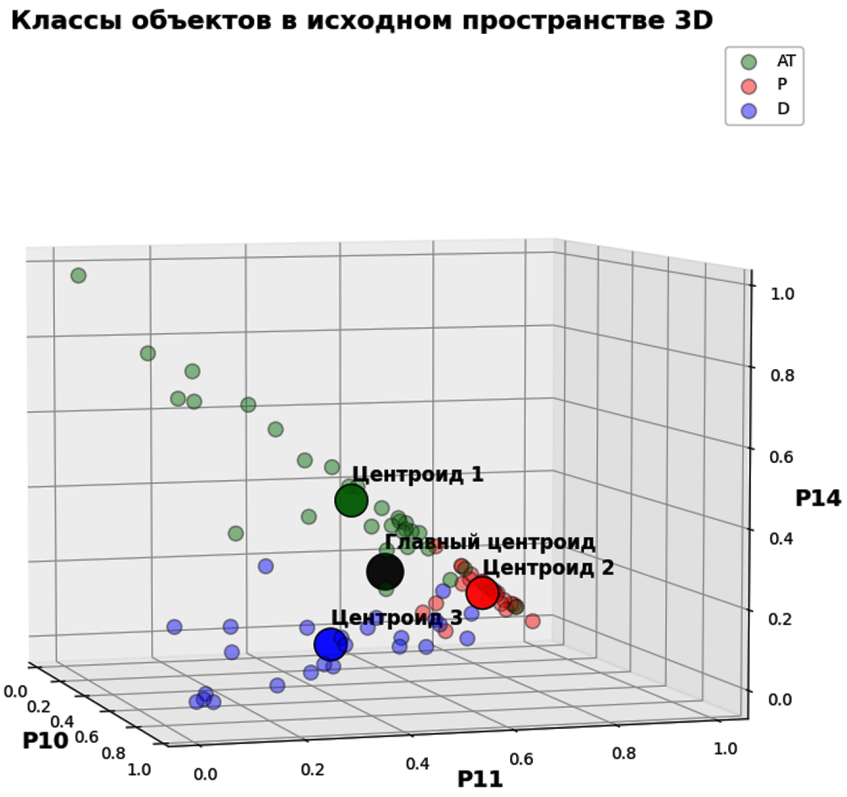


Рис. 4: Данные в исходном трехмерном пространстве 1

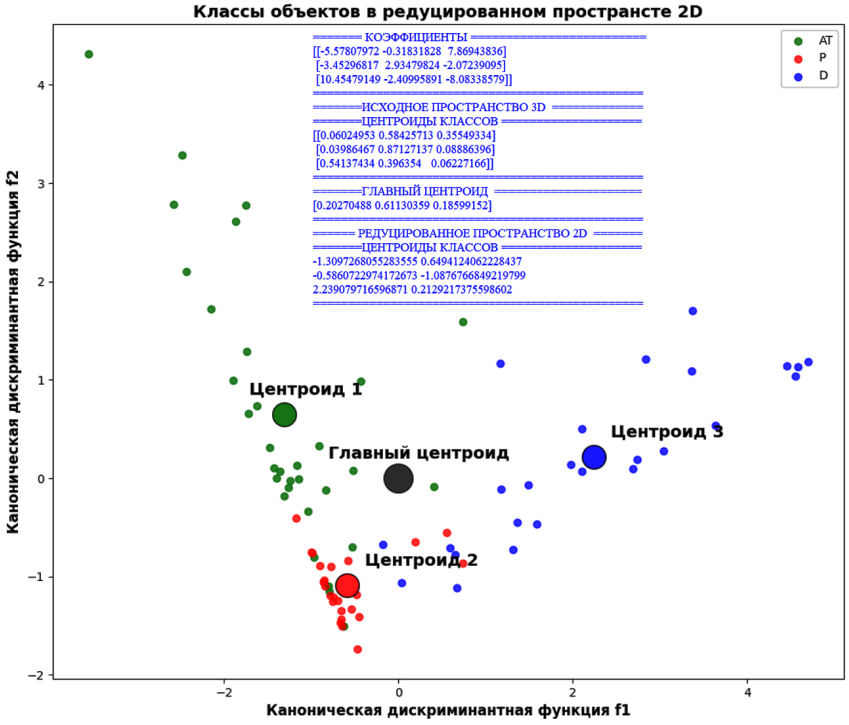


Рис. 5: Данные в редуцированном двухмерном пространстве

```
wpd=c(10,11)
wat=c(14,16)
reg_at <-xreg[,wat]
sum_at <-rowSums(reg_at)
reg.pd <-xreg[,wpd]
sum_pd <-rowSums(reg.pd)
sum_ob <-sum_at+sum_pd
xat_nor <-sum_at/sum_ob
xpd_nor <-reg.pd/sum_ob
xreg_nor <-cbind(xat_nor,xpd_nor)
Y <-xreg_nor



#Метод LDA
library(MASS)
y.lda <-Y
y.train = sample(1:81, 60)
y.lda <-cbind(y.lda,xreg_f)
Rezul.lda <- lda(xreg_f ~ ., y.lda,prior = c(1,1,1)/3,subset = y.train)
Rezul.lda

ypred=predict(Rezul.lda, y.lda [-y.train,])$class
ytest=y.lda [-y.train,4]
CrossTable(x=ytest,y=ypred,prob.chisq = FALSE)

newdat <- predict(Rezul.lda,y.lda)
#newdat$x


y.lda_dr <-data.frame(newdat$x,xreg_f)
newdan1 =subset(y.lda_dr,xreg_f=='AT')
newdan2 =subset(y.lda_dr,xreg_f=='P')
newdan3 =subset(y.lda_dr,xreg_f=='D')
centroid1=sapply(newdan1[,-3],mean)
centroid2=sapply(newdan2[,-3],mean)
centroid3=sapply(newdan3[,-3],mean)
```

```
print('Центроид 1')
print(centroid1)
print('Центроид 2')
print(centroid3)
print('Центроид 3')
print(centroid2)

#График регионов по классам в двухмерном пространстве 2D
x11()
g4 <-ggplot(data= y.lda_dr[,-3], aes(x=LD1,y=LD2,colour = xreg_f),
colors = "Accent")
g4 <-g4 +geom_point(size=4)
g4 <-g4 +geom_point(aes(x=centroid1[1],y=centroid1[2]), size=8,color='red')
g4 <-g4 +geom_point(aes(x=centroid2[1],y=centroid2[2]), size=8,color='blue')
g4 <-g4 +geom_point(aes(x=centroid3[1],y=centroid3[2]), size=8,color='green')
g4 <-g4 +geom_point(aes(x=0.0,y=0.0), size=8,color='black')
g4 <-g4 +labs(title ="Регионы в редуцированном пространстве 2D",
x="Каноническая дискриминантная функция 1",
y="Каноническая дискриминантная функция 2",colour = "Классы")
g4 <-g4 +geom_text(aes(label='Главный центроид'),size =6,
color='black',x=0.0+0.2,y=0.0+0.3)
g4 <-g4 +geom_text(aes(label='Центроид 1'),size =6, color='red',
x=centroid1[1]-0.2,y=centroid1[2]+0.5)
g4 <-g4 +geom_text(aes(label='Центроид 2'),size =6, color='chartreuse4',
x=centroid3[1]-0.5,y=centroid3[2]+0.35)
g4 <-g4 +geom_text(aes(label='Центроид 3'),size =6, color='blue',
x=centroid2[1],y=centroid2[2]+0.35)
g4 <-g4 + theme(text=element_text(size=16,  family="serif"))
#g4 <-g4 + theme_fivethirtyeight() +  scale_color_fivethirtyeight()
#g4 <-g4 + theme_stata() + scale_colour_stata()c
g4

```

Результаты работы скрипта (рис. 6,7):

Классификационная матрица расчитывалась на тестовой части данных (21 регион), а модель строилась на обучаемой части (60 регионов). В тесте из 21 региона правильно классифицированы 18 регионов (87 про-

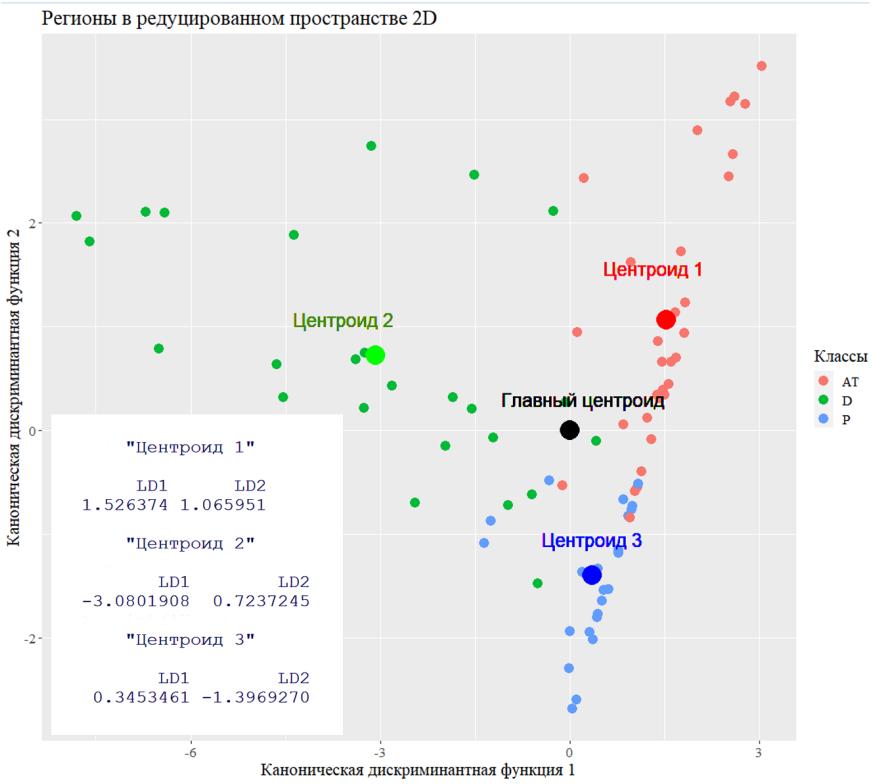


Рис. 6: Данные в редуцированном двухмерном пространстве

```
   Cell Contents
|-------------------------|
|                       N |
| Chi-square contribution |
|           N / Row Total |
|           N / Col Total |
|         N / Table Total |
|-------------------------|


Total Observations in Table:  21

```

ytest	ypred AT	D	P	Row Total
AT	7	0	2	9
	5.333	3.857	0.010	
	0.778	0.000	0.222	0.429
	1.000	0.000	0.400	
	0.333	0.000	0.095	
D	0	9	1	10
	3.333	5.186	0.801	
	0.000	0.900	0.100	0.476
	0.000	1.000	0.200	
	0.000	0.429	0.048	
P	0	0	2	2
	0.667	0.857	4.876	
	0.000	0.000	1.000	0.095
	0.000	0.000	0.400	
	0.000	0.000	0.095	
Column Total	7	9	5	21
	0.333	0.429	0.238	

```
Prior probabilities of groups:
       AT         D         P
0.3333333 0.3333333 0.3333333

Group means:
     xat_nor        P10       P11
AT 0.6494031 0.02024398 0.3303529
D  0.3915010 0.28989595 0.3186031
P  0.3515161 0.03060702 0.6178769

Coefficients of linear discriminants:
               LD1       LD2
xat_nor  4.6382865  4.096909
P10     -8.1587493  2.339246
P11      0.7089649 -4.116258

Proportion of trace:
   LD1    LD2
0.6443 0.3557
```

Рис. 7: Расчетные данные

центов). Процедура контролируемого уменьшения размерности LDA, как видно графически, довольно эффективна. Она на наших данных сохраняет взамное расположения объектов в каждом классе и между классами, что хорошо видно по расположению центроидов в редуцированном пространстве 2D. При этом две первые канонические дискриминантные функции берут почти все 100 процентов общих дискриминантных возможностей. Это говорит, что исходные пременные сильно коррелированы и одна из них лишняя для решения задачи классификации на втором наборе данных.