**МИНОБРНАУКИ РОССИИ**

**Санкт-Петербургский государственный**

**электротехнический университет**

**«ЛЭТИ» им. В.И. Ульянова (Ленина)**

**Кафедра БТС**

ИДЗ

**по дисциплине «Технологии и системы принятия решений»**

Тема: Классификация опасных аритмий

|  |  |  |
| --- | --- | --- |
| Студент гр. 9503 |  | Кошаев Е.А. |
| Преподаватель |  | Манило Л.А. |

Санкт-Петербург

2024

**ЗАДАНИЕ**

**на ИДЗ**

|  |  |  |
| --- | --- | --- |
| Студент Кошаев Е.А. | | |
| Группа 9503 | | |
| Тема реферата: Классификация опасных аритмий | | |
| Исходные данные:  кратко указываются исходные данные к реферату и основные требования, предъявляемые к нему | | |
| Предполагаемый объем реферата:  Не менее 00 страниц (обязательны разделы «Содержание», «Введение», «Заключение», «Список использованных источников»). | | |
| Дата выдачи задания: 10.09.2024 | | |
| Дата сдачи реферата: 20.11.2024 | | |
| Дата защиты реферата: 20.11.2024 | | |
|  | | |
| Студент |  | Кошаев Е.А. |
| Преподаватель |  | Манило Л.А. |

# Подготовка данных

# Классификацию данных методом k ближайших соседей

Метод K ближайших соседей основан на гипотезе компактности, которая предполагает, что расположенные близко друг к другу объекты в пространстве признаков имеют схожие значения целевой переменной или принадлежат к одному классу.

В данной работе используется два алгоритма классификации:

1. Объект присваивается тому классу, который является наиболее распространенным среди k соседей данного элемента, классы которых уже известны (т.е. объект имеет наибольшее число соседей в данном классе).
2. Взвешенный способ. Оценивается не только количество объектов, попавших в область близости каждого класса, но и их удаленность от нового объекта. Для каждого класса j определяется оценка близости:

где – расстояние от нового значения до объекта , – количество соседей в данном классе. Объекту присваивается тот класс, у которого выше значение близости.

Чаше всего за метрику расстояния берется евклидово расстояние, которое вычисляется по следующей формуле:

где и – точки, между которыми рассчитывается расстояние.

Для обучения классификаторов используются первые 15 объектов каждого класса. Обучение состоит в запоминании классификатором переданных и размеченных данных. Далее, при классификации объектов, применяется один из описанных выше алгоритмов.

Качество классификации алгоритмов проверялось при помощи подсчета общей точности. Общая точность рассчитывается по формуле:

где – число верно распознанных объектов i-го класса, – число объектов i-го класса, отнесенных к классу j, – число объектов в классе.

На рисунке 1 представлены графики зависимости общей точности от числа ближайших соседей.

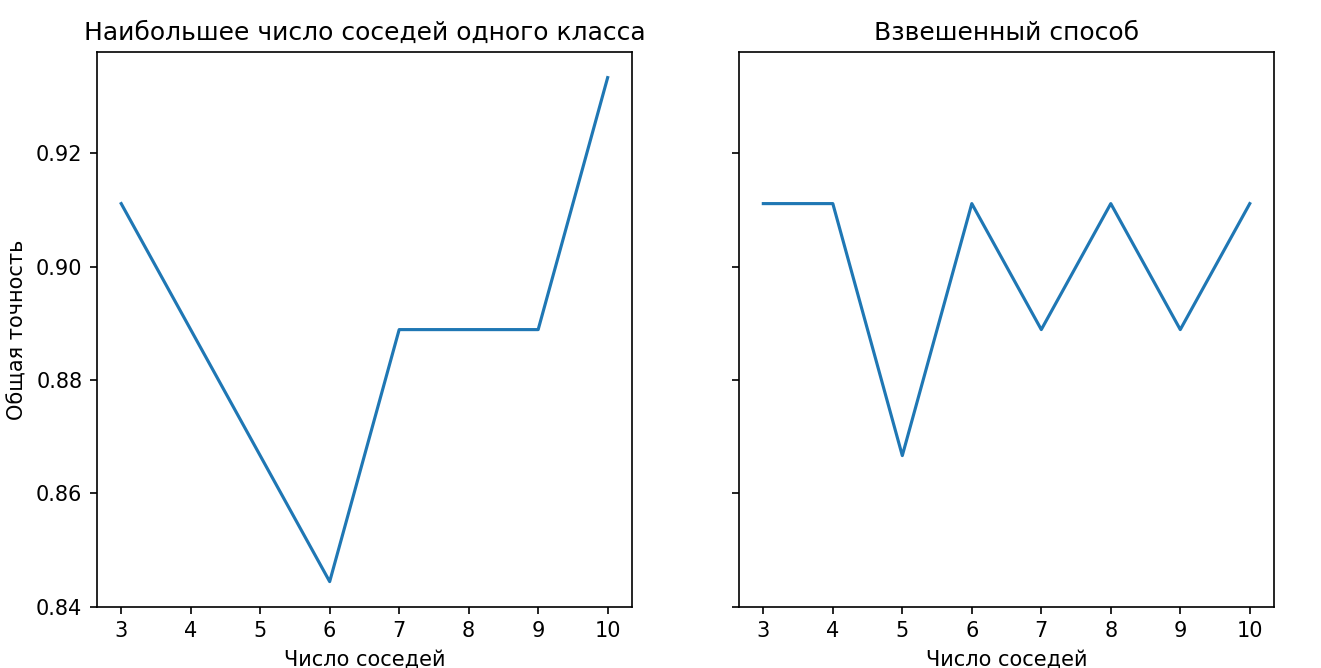


Рисунок 1 – Зависимость общей точности от числа ближайших соседей

В таблицах 1 – 2 представлены результаты классификации с использованием первого и второго алгоритмов соответственно.

Таблица – результаты классификации с использование первого алгоритма.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Количество соседей,  *K* | Истинный класс,  *i* | Число объектов | Результат распознавания (класс), j | | | Общая точность, OA |
| 1 | 2 | 3 |
| 3 | 1 | 15 | 14 | 1 | 0 | 0.91 |
| 2 | 15 | 3 | 12 | 0 |
| 3 | 15 | 0 | 0 | 15 |
| 4 | 1 | 15 | 14 | 1 | 0 | 0.88 |
| 2 | 15 | 3 | 12 | 0 |
| 3 | 15 | 0 | 1 | 14 |
| 5 | 1 | 15 | 14 | 1 | 0 | 0.86 |
| 2 | 15 | 3 | 10 | 2 |
| 3 | 15 | 0 | 0 | 15 |
| 6 | 1 | 15 | 14 | 1 | 0 | 0.84 |
| 2 | 15 | 4 | 11 | 0 |
| 3 | 15 | 0 | 2 | 13 |
| 7 | 1 | 15 | 14 | 1 | 0 | 0.88 |
| 2 | 15 | 4 | 11 | 0 |
| 3 | 15 | 0 | 0 | 15 |
| 8 | 1 | 15 | 14 | 1 | 0 | 0.88 |
| 2 | 15 | 4 | 11 | 0 |
| 3 | 15 | 0 | 0 | 15 |
| 9 | 1 | 15 | 14 | 1 | 0 | 0.88 |
| 2 | 15 | 3 | 11 | 1 |
| 3 | 15 | 0 | 0 | 15 |
| 10 | 1 | 15 | 15 | 0 | 0 | 0.93 |
| 2 | 15 | 3 | 12 | 0 |
| 3 | 15 | 0 | 0 | 15 |

Таблица – Результаты классификации с использованием второго алгоритма

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Количество соседей,  *K* | Истинный класс,  *i* | Число объектов | Результат распознавания (класс), j | | | Общая точность, OA |
| 1 | 2 | 3 |
| 3 | 1 | 15 | 14 | 1 | 0 | 0.91 |
| 2 | 15 | 3 | 12 | 0 |
| 3 | 15 | 0 | 0 | 15 |
| 4 | 1 | 15 | 14 | 1 | 0 | 0.91 |
| 2 | 15 | 3 | 12 | 0 |
| 3 | 15 | 0 | 0 | 15 |
| 5 | 1 | 15 | 14 | 1 | 0 | 0.86 |
| 2 | 15 | 3 | 10 | 2 |
| 3 | 15 | 0 | 0 | 15 |
| 6 | 1 | 15 | 14 | 1 | 0 | 0.91 |
| 2 | 15 | 3 | 12 | 0 |
| 3 | 15 | 0 | 0 | 15 |
| 7 | 1 | 15 | 14 | 1 | 0 | 0.88 |
| 2 | 15 | 4 | 11 | 0 |
| 3 | 15 | 0 | 0 | 15 |
| 8 | 1 | 15 | 14 | 1 | 0 | 0.91 |
| 2 | 15 | 3 | 12 | 0 |
| 3 | 15 | 0 | 0 | 15 |
| 9 | 1 | 15 | 14 | 1 | 0 | 0.88 |
| 2 | 15 | 3 | 11 | 1 |
| 3 | 15 | 0 | 0 | 15 |
| 10 | 1 | 15 | 14 | 1 | 0 | 0.91 |
| 2 | 15 | 3 | 12 | 0 |
| 3 | 15 | 0 | 0 | 15 |

# Метод главных компонент

Для дальнейшего исследования методов классификации, сведем трёхклассовую задачу к поэтапному решению двухклассовых задач. Для этого воспользуемся методом главных компонент.

Метод главных компонент позволяет уменьшит пространство признаков, но не потерять их информативность. Это позволит отобразить классы в пространстве двух первых главных компонент, найти наиболее удаленный (или более плотно сгруппированный) класс, отделив его от двух остальных. Оставшиеся два класса объединяются в общий комплексный класс.

Так, на первом этапе решается задача классификации наиболее удаленного класса и комплексного класса, на втором этапе – классификация классов, составляющих общий.

На рисунке 2 представлена диаграмма рассеяния в пространстве двух первых главных компонент.

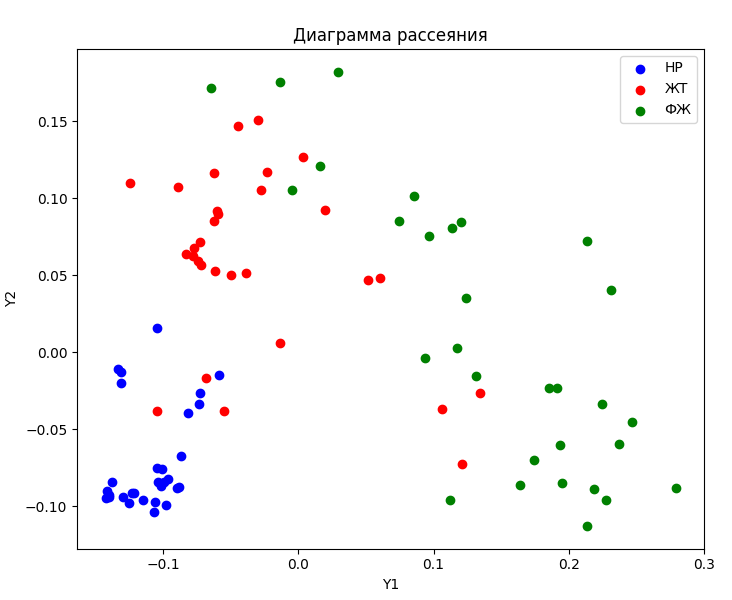


Рисунок – Диаграмма рассеяния в пространстве двух первых главных компонент

Доля дисперсии двух первых главных компонент составляет , . Было решено выделить нормальный ритм (НР) в отдельный класс. Элементы этого класса легко линейно отделимы от двух других классов (фибрилляции желудочков (ФЖ) и желудочковой тахикардии (ЖТ)) как показано на рисунке 2.

Соответственно, на первом этапе будет решаться задача классификации НР и ЖТ + ФЖ, на втором этапе – задача классификации ЖТ и ФЖ.

# Метод классификации по минимуму расстояния

Классификатором по минимуму расстояния называют такой классификатор, оптимальное решающее правило которого формулируется следующим образом: чтобы определить класс вектора *x*, следует измерить евклидово расстояние от *х* до каждого из *c* векторов средних значений и отнести *х* к классу, соответствующему ближайшему среднему значению.

Вычислим весовой вектор из выражения:

где – вектор средних значений по всем признакам для первого класса, – вектор средних значений по всем признакам для второго класса.

Вычисленный весовой вектор необходимо пронормировать.

где – евклидово расстояние. Евклидово расстояние можно рассчитать по формуле .

Процедура распознавания заключается в вычислении проекции вектора на направление весового вектора и сравнении полученной величины с порогом . Порог можно вычислить по следующей формуле:

Так как априорные вероятности по условию равны, то можно представить как . Соответственно:

Однако, более хорошие результаты классификации даёт порог, определенный в месте пресечения функций плотности вероятности. Плотность вероятности можно вычислить следующим образом:

где и – это среднее и стандартное отклонения признака в классе y.

Вычислим для каждого класса среднее и дисперсию проекций объектов на направление W. Результаты вычислений представлены в таблице 3.

Таблица – Результаты вычислений среднего и дисперсии для каждого класса

|  |  |  |  |
| --- | --- | --- | --- |
| Этап классификации | Класс | Среднее, | Дисперсия, |
| Этап 1 (НР и ЖТ+ФЖ) | НР | -0.074 | 0.030 |
| ЖТ+ФЖ | -0.270 | 0.075 |
| Этап 2 (ЖТ и ФЖ) | ЖТ | -0.040 | 0.069 |
| ФЖ | -0.220 | 0.100 |

Разделяющая гиперплоскость определяется, как и ей соответствует следующий алгоритм распознавания:

На рисунке 3 изображены проекции множества классов на весовой вектор для двух этапов.

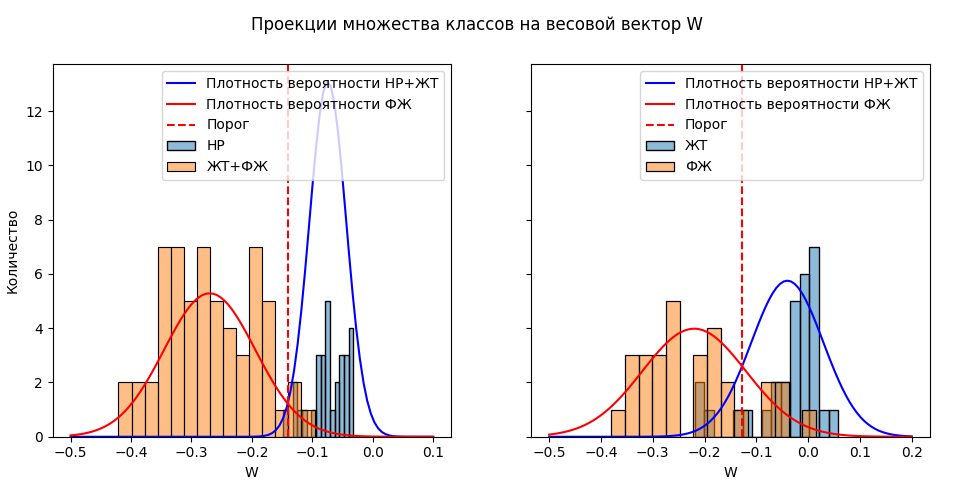


Рисунок – Проекции множества классов на весовой вектор

Проведенные расчеты позволяют получить порог, равный, для первого этапа (НР и ЖТ+ФЖ) , а для второго этапа (ЖТ и ФЖ) .

Запишем уравнения разделяющих гиперплоскостей для первого и второго этапов:

Этап 1 (НР и ЖТ+ФЖ):

Этап 2 (ЖТ и ФЖ):

Решающее правило для каждого из этапов будет иметь следующий вид:

Этап 1 (НР и ЖТ+ФЖ):

Этап 2 (ЖТ и ФЖ):

Можно сформулировать следующий алгоритм классификации: если на первом этапе , то данный объект принадлежит классу НР, иначе объект принадлежит объединённому классу ЖТ+ФЖ и переходит на второй этап классификации. На втором этапе снова сверяем, если , то данный объект принадлежит классу ФЖ, иначе ЖТ.

# Метод классификации по линейному дискриминанту Фишера

Классификатором по линейному дискриминанту Фишера называют такой классификатор, который определяет вектор как линейную функцию с максимальным отношением разброса между классами к «среднему» разбросу внутри классов.

где и – средние значения проекций выборок классов, и – внутриклассовый разброс.

ЛДФ Фишера определяется как такой вектор, для которого критерий максимален.

Среднее значение проекций выборок классов можно представить, как:

где – матрица разброса между классами.

можно рассчитать следующим образом:

Выборочный разброс для двух классов можно определить, как:

где – «усредненная» матрица разброса для двух классов .

Тогда критерий Фишера можно записать в виде:

Искомый вектор можно вычислить, как:

Следует отметить, что полученное выражение вычисления вектора совпадает с выражением для случая нормально распределенных классов с равными ковариационными матрицами, которое записывается следующим образом:

Вычисленный весовой вектор необходимо пронормировать:

Рассчитаем порог, определенный в месте пресечения функций плотности вероятности. Для этого необходимо вычислить для каждого класса среднее и дисперсию проекций объектов на направление W. Результаты вычислений представлены в таблице 4.

Таблица – Результаты вычислений среднего и дисперсии для каждого класса

|  |  |  |  |
| --- | --- | --- | --- |
| Этап классификации | Класс | Среднее, | Дисперсия, |
| Этап 1 (НР и ЖТ+ФЖ) | НР | -0.005 | 0.011 |
| ЖТ+ФЖ | -0.065 | 0.015 |
| Этап 2 (ЖТ и ФЖ) | ЖТ | 0.077 | 0.011 |
| ФЖ | -0.005 | 0.011 |

На рисунке 4 изображены проекции множества классов на весовой вектор для двух этапов.

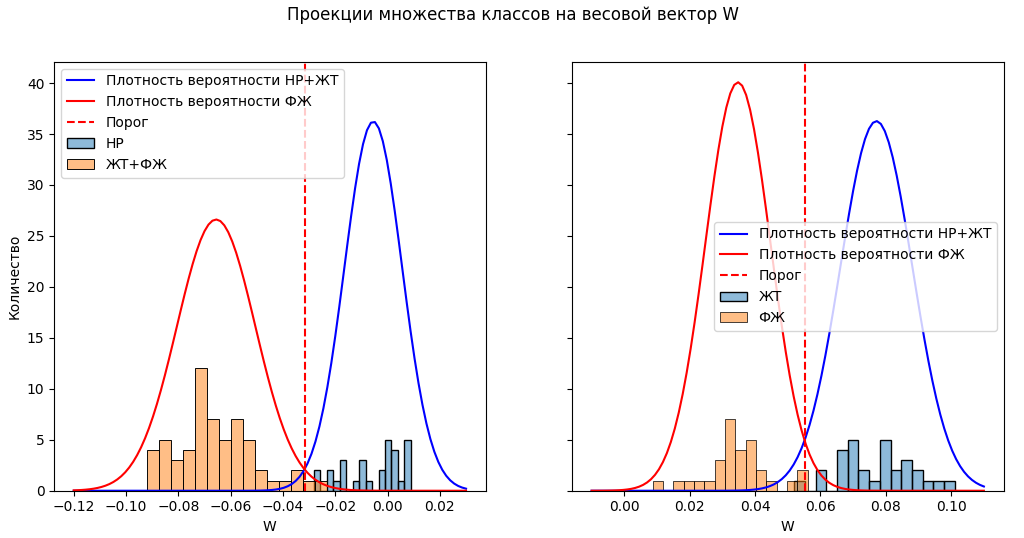


Рисунок – Проекции множества классов на весовой вектор

Проведенные расчеты позволяют получить порог, равный, для первого этапа (НР и ЖТ+ФЖ) , а для второго этапа (ЖТ и ФЖ) .

Запишем уравнения разделяющих гиперплоскостей для первого и второго этапов:

Этап 1 (НР и ЖТ+ФЖ):

Этап 2 (ЖТ и ФЖ):

Решающее правило для каждого из этапов будет иметь следующий вид:

Этап 1 (НР и ЖТ+ФЖ):

Этап 2 (ЖТ и ФЖ):

Можно сформулировать следующий алгоритм классификации: если на первом этапе , то данный объект принадлежит классу НР, иначе объект принадлежит объединённому классу ЖТ+ФЖ и переходит на второй этап классификации. На втором этапе снова сверяем, если, то данный объект принадлежит классу ФЖ, иначе ЖТ.

# Метод классификации по дискриминантному анализу с использованием критерия Фишера

Классификатором по дискриминантному анализу с использованием критерия Фишера называют такой классификатор, который посредствам понижения размерности пространства признаков позволяет производить разделение классов наилучшим образом. Основная идея заключается в предположении о многомерном нормальном распределении признаков внутри классов и поиске их линейного преобразования, которое максимизирует межклассовую дисперсию и минимизирует внутриклассовую.

Данный подход основан на применении того же критерия оптимизации и обобщении определений для матриц разброса между классами и внутри классов. Критерий , оценивающий степень разделения заданных классов сигналов, в общем виде может задаваться следом матрицы

где – матрица рассеяния между классами; – обобщенная матрица рассеяния внутри классов, либо отношением определителей матриц.

Матрица задаётся в следующем виде:

где – частота появления объектов, образующих класс .

Матрица определяется в виде:

где – матрица разброса для i–го класса (находится как выборочная матрица ковариации i-го класса) , – нахождение среднего для i-класса.

Необходимо вычислить матрицы и и далее перейти к нахождению собственных векторов (составляющих матрицы W). Нахождение элементов матрицы сводится к задаче определения собственных значений матрицы . Собственные векторы с ненулевыми собственными значениями определяют то (c-1) c-мерное пространство, в котором могут быть построены дискриминантные функции и определены решающие правила.

Собственные значения вычисленных собственных векторов должны удовлетворять неравенству .

Также, собственные вектора необходимо пронормировать:

На рисунке 5 изображена диаграмма рассеянности классов в уменьшенном пространстве признаков.

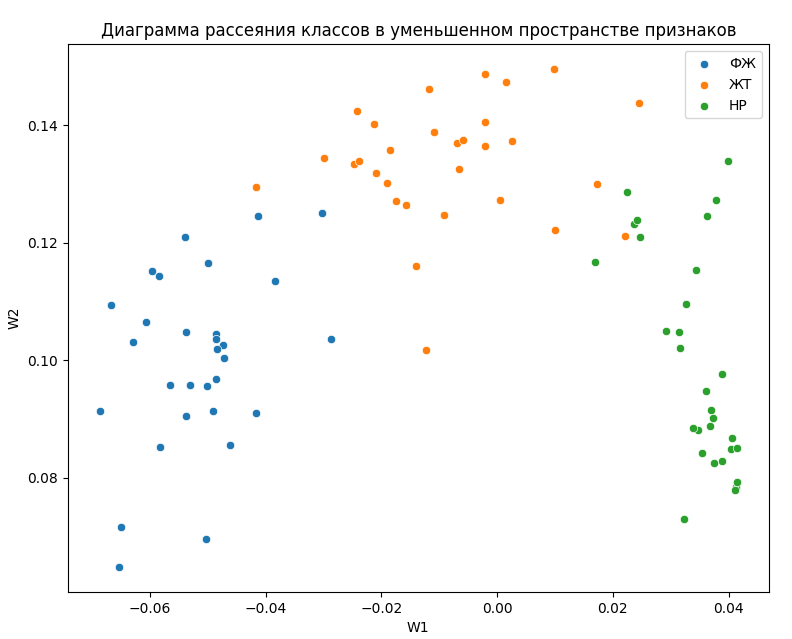


Рисунок – Диаграмма рассеянности классов в уменьшенном пространстве признаков

Решим задачу классификации геометрически. Для этого отделим скопления классов друг от друга при помощи дополнительных геометрических построений.

На рисунке 6 представлена диаграмма рассеянности классов в уменьшенном пространстве признаков с дополнительным построением прямых, разделяющих классы.

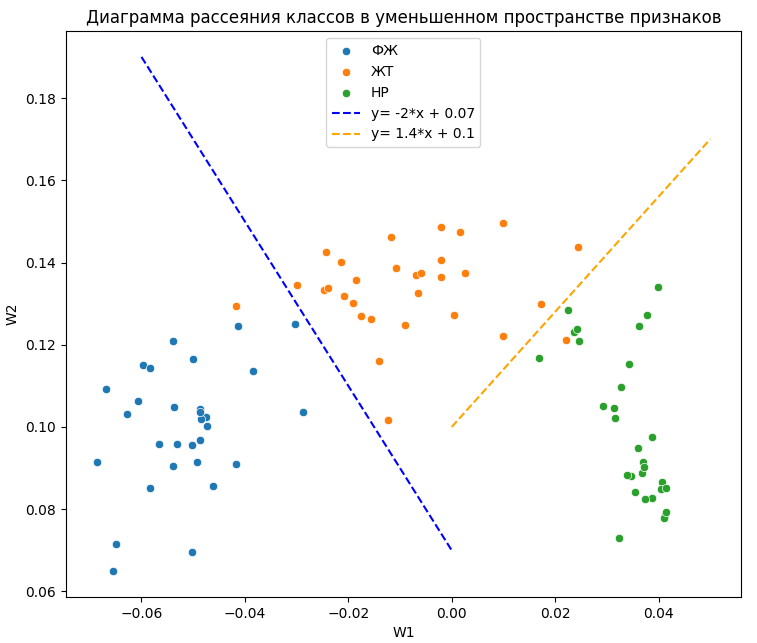


Рисунок – Диаграмма рассеянности классов в уменьшенном пространстве признаков с дополнительным построением прямых, разделяющих классы

Из рисунка 6 видно, что для классификации были построены следующие прямые:

Решающее правило для каждого из этапов будет иметь следующий вид:

Этап 1 (НР и ЖТ+ФЖ):

Этап 2 (ЖТ и ФЖ):

Можно сформулировать следующий алгоритм классификации: если на первом этапе , то данный объект принадлежит классу НР, иначе объект принадлежит объединённому классу ЖТ+ФЖ и переходит на второй этап классификации. На втором этапе снова сверяем, если , то данный объект принадлежит классу ФЖ, иначе ЖТ.

# Сравнение методов

Сравним методы классификации, используемые в данной работе. Для этого построим ROC-кривые для каждого из этапов классификации. Кривая ROC – это график, который иллюстрирует производительность классификационной модели при всех возможных порогах классификации. Ось X данного графика представляет собой специфичность, т.е ложноположительную частоту, а ось Y – чувствительность, т.е истинно положительную частоту.

Чувствительность и специфичность можно рассчитать следующим образом:

где – истинно положительные результаты, – ложно отрицательные результаты, – ложноположительные результаты, – истинно отрицательные результаты.

На рисунке 7 изображены ROC-кривые каждого классификатора для первого и второго этапов.

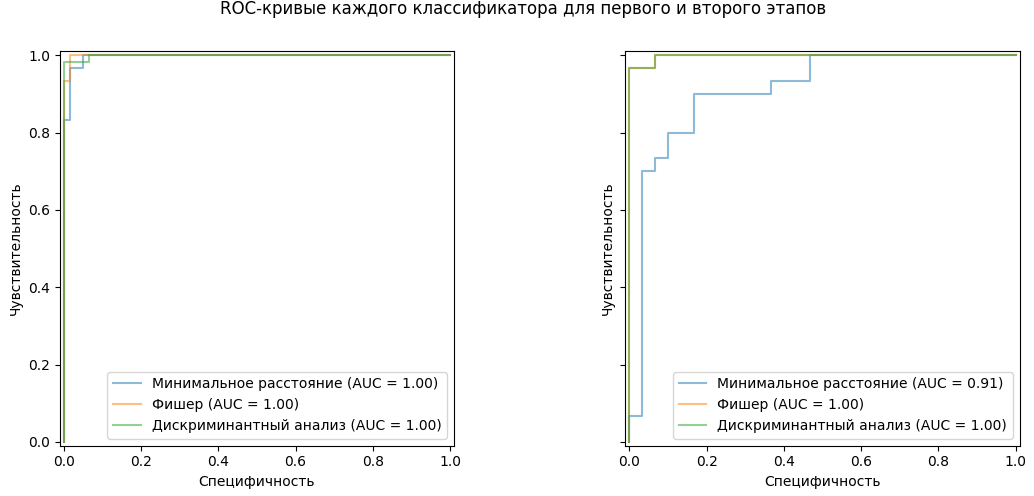


Рисунок – ROC-кривые каждого классификатора для первого и второго этапов

Анализируя результаты, приведенные на рисунке 7, можно заменить, что дискриминантный анализ показывает наилучшие результаты для обоих этапов классификации. Также, следует заметить, что с первым этапом классификации (НР и ЖТ+ФЖ) отлично справились все 3 классификатора, а на втором хорошие результаты также показывает классификатор по Фишеру.

В таблицах 5-7 приведены расчётные значения метрик для каждого классификатора.

Таблица – Метрики классификатора по минимуму расстояния

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Этап классификации | Класс | TP | TN | FN | FP | Чувствительность,  Se | Специфичность,  Sp | Точность,  Acc |
| Этап 1 (НР и ЖТ+ФЖ) | НР | 29 | 57 | 1 | 3 | 96% | 5% | 95% |
| ЖТ+ФЖ | 57 | 29 | 3 | 1 | 95% | 3% |
| Этап 2 (ЖТ и ФЖ) | ЖТ | 26 | 25 | 4 | 5 | 86% | 16% | 85% |
| ФЖ | 25 | 26 | 5 | 4 | 83% | 13% |

Таблица – Метрики классификатора по критерию Фишера

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Этап классификации | Класс | TP | TN | FN | FP | Чувствительность,  Se | Специфичность,  Sp | Точность, |
| Этап 1 (НР и ЖТ+ФЖ) | НР | 30 | 58 | 0 | 2 | 100% | 3% | 97% |
| ЖТ+ФЖ | 58 | 30 | 2 | 0 | 96% | 0% |
| Этап 2 (ЖТ и ФЖ) | ЖТ | 29 | 29 | 1 | 1 | 96% | 3% | 96% |
| ФЖ | 29 | 29 | 1 | 1 | 96% | 3% |

Таблица – Метрики классификатора по дискриминантному анализу

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Этап классификации | Класс | TP | TN | FN | FP | Чувствительность,  Se | Специфичность,  Sp | Точность, |
| Этап 1 (НР и ЖТ+ФЖ) | НР | 30 | 59 | 0 | 1 | 100% | 1% | 98% |
| ЖТ+ФЖ | 59 | 30 | 1 | 0 | 98% | 0% |
| Этап 2 (ЖТ и ФЖ) | ЖТ | 29 | 30 | 1 | 0 | 96% | 0% | 98% |
| ФЖ | 30 | 29 | 0 | 1 | 100% | 3% |

Анализирую данные в таблицах 5-7, можно увидеть, что классификатор по дискриминантному анализу показывает наилучшие результаты по таким параметрам, как: точность, чувствительность, специфичность. Что подтверждает выводе, сделанный ранее.