Министерство образования и науки Российской Федерации

Федеральное государственное автономное образовательное

учреждение высшего образования

«Южно-Уральский государственный университет»

(Национальный исследовательский университет)

Высшая школа электроники и компьютерных наук

Кафедра системного программирования

ОТЧЕТ

По лабораторной работе №3

|  |  |  |  |
| --- | --- | --- | --- |
| Байесовская классификация | | | |
|  | | | |
|  |  | Руководитель: | |
|  |  | | \_\_\_\_\_\_\_\_\_\_\_\_\_М.Л. Цымблер  «\_\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 2020 г.  Автор работы  студент группы КЭ-220  \_\_\_\_\_\_\_\_\_\_\_\_А.В.Витомсков  «\_\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 2020 г. |

Челябинск 2020

1. ***Выполните классификацию набора данных с помощью Байесовской классификации, варьируя соотношение мощностей обучающей и тестовой выборок от 60%:40% до 90%:10% с шагом 5%.***

Будем использовать датасет, полученный в процессе выполнения л/р 2 "Разметка аудиоданных" по предмету "Анализ аудио видео и графической информации".

Данный датасет содержит информацию о музыкальных произведениях 3600 исполнителей (по 300 исполнителей каждого из 12 жанров популярной музыки: рок, поп, регги, шансон и т.д.). У каждого исполнителя было выбрано по 30 музыкальных произведений, из каждого произведения вырезаны аудиофрагменты по 10 секунд. Частота дискретизации фрагментов выбрана   
равной 20 кГц. Затем эти музыкальные файлы были представлены в виде мел-кепстальных коэффициентов: каждые 512 отсчетов представлены 15-ю коэффициентами.

Датасет состоит из двух файлов.

Во втором файле “genres\_pca\_4na12920\_comp\_of\_song.csv” методом главных компонент (см. код в файле 1\_rca.ipynb) из 15 мел-кепстральных коэффициента оставлены 4. Таким образом, на каждого исполнителя приходится 4 х 12920 признака.

В первом файле “genres\_pca\_100\_comp\_of\_song.csv” методом главных компонент из 15 мел-кепстральных коэффициента оставлены 4 с сохранением 90% дисперсии. По временной шкале тоже был применен метод главных компонент с сохранением 90% дисперсии. В итоге, файл genres\_pca\_100\_comp\_of\_song.csv содержит по 100 признаков для каждого из 3600 исполнителей.

Ставится задача классифицировать исполнителей по 12 жанрам. Для простоты считаем, что каждый из исполнителей исполняет только в одном конкретном жанре.

Алгоритм был запущен 2 раза. Первый раз анализировался файл, содержащий 100 главных компонент из каждых 30 10-секундных фрагментов аудиоданных. Второй раз анализировался файл, в котором были сохранены все 1920\*4 компоненты.

Код программы приведен в файле “2\_gaussian.ipynb”.

Будем использовать наивную байесовскую классификацию. Используем модель Gaussian из библиотеки scikit-learn. Модель данного типа используется в случае непрерывных признаков и предполагает, что значения признаков имеют нормальное распределение.

2. ***Вычислите показатели качества классификации: аккуратность (accuracy), точность (precision), полнота (recall), F-мера. Выполните визуализацию полученных результатов в виде диаграмм.***

Вычисление показателей качества было реализовано мной отдельной функцией.

Точность классификатора вычислялась по формуле

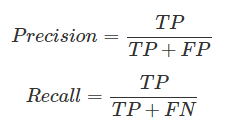


где P – количество объектов, по которым классификатор принял

правильное решение,

N – размер обучающей выборки.

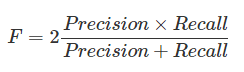
Точность и полнота вычисляются следующим образом:



Для их вычисления мной была построена матрица неточностей (confusion matrix). Матрица неточностей – это матрица размера N на N, где N — это количество классов. Столбцы этой матрицы резервируются за экспертными решениями, а строки за решениями классификатора.

Имея такую матрицу точность и полнота для каждого класса рассчитывается очень просто. Точность равняется отношению соответствующего диагонального элемента матрицы и суммы всей строки класса. Полнота – отношению диагонального элемента матрицы и суммы всего столбца класса.

F-мера представляет собой гармоническое среднее между точностью и полнотой. Она стремится к нулю, если точность или полнота стремится к нулю.



Результаты работы алгоритма для файла genres\_pca\_100\_comp\_of\_song.csv представлены на рисунке 1 и 2.

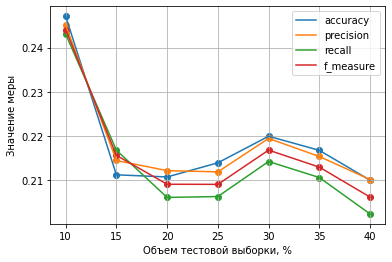


Рисунок 1 – Значения мер качества классификации при различных отношениях объема тестовой и обучающей выборки

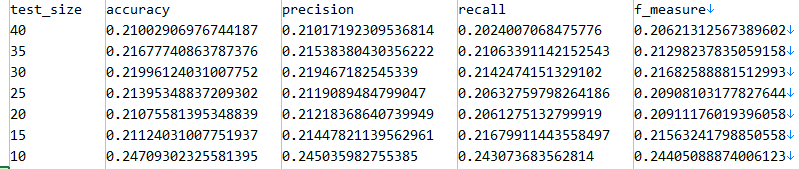


Рисунок 2 - Значения мер качества классификации в виде таблицы

При малых объемах тестовой выборки имеем несколько завышенные значения мер, что говорит не только о том, что модель лучше обучилась на большем объеме данных, но и о недостаточности (малом количестве) тестовых примеров, результаты такого тестирования нестабильны. Наиболее оптимальный результат получается при объеме тестовой выборки около 30%.

Из рисунков 1 и 2 видно, что меры достаточно малы по абсолютному значению, но необходимо учитывать наличие 12 жанров, которые с точки зрения модели не могу коррелировать между собой. Если посмотреть на матрицы неточностей, то видно, что модель в целом справляется с задачей. Для примера, если взять модель, рандомно вычисляющую результат, то для нее accuracy составит 100/12 = 8,33%, а наша модель повышает вероятность такого угадывания в 3 раза.

Матрицы неточностей приведены в таблице 1.

Таблица 1 – Сonfusion matrix

|  |
| --- |
| Модель с мощностью обучающей выборки 60%, тестовой выборки 40%  confusion\_matrix  allrock blues country electronic hip-hop jazz \  allrock 26.0 4.0 10.0 17.0 8.0 2.0  blues 17.0 24.0 20.0 8.0 5.0 11.0  country 12.0 18.0 17.0 7.0 10.0 10.0  electronic 15.0 4.0 6.0 44.0 18.0 5.0  hip-hop 8.0 2.0 5.0 24.0 19.0 3.0  jazz 2.0 7.0 6.0 15.0 5.0 28.0  latin-american 3.0 1.0 8.0 12.0 5.0 7.0  pop 9.0 2.0 4.0 16.0 15.0 1.0  r-n-b 6.0 4.0 5.0 14.0 16.0 3.0  reggae 3.0 5.0 4.0 22.0 22.0 3.0  shanson 2.0 1.0 4.0 1.0 1.0 2.0  world 4.0 9.0 4.0 7.0 12.0 10.0  latin-american pop r-n-b reggae shanson world  allrock 6.0 5.0 8.0 3.0 18.0 11.0  blues 6.0 5.0 7.0 3.0 9.0 10.0  country 9.0 3.0 6.0 4.0 20.0 5.0  electronic 4.0 8.0 2.0 3.0 6.0 4.0  hip-hop 10.0 7.0 10.0 5.0 14.0 9.0  jazz 7.0 3.0 7.0 3.0 14.0 25.0  latin-american 7.0 7.0 8.0 6.0 38.0 12.0  pop 4.0 10.0 9.0 5.0 22.0 3.0  r-n-b 7.0 9.0 13.0 8.0 10.0 5.0  reggae 8.0 0.0 9.0 13.0 22.0 5.0  shanson 12.0 7.0 2.0 5.0 71.0 0.0  world 9.0 5.0 3.0 5.0 32.0 17.0    Модель с мощностью обучающей выборки 65%, тестовой выборки 35%  confusion\_matrix  allrock blues country electronic hip-hop jazz \  allrock 24.0 4.0 9.0 17.0 8.0 1.0  blues 12.0 21.0 16.0 7.0 4.0 12.0  country 10.0 12.0 9.0 5.0 11.0 7.0  electronic 15.0 2.0 5.0 40.0 16.0 5.0  hip-hop 4.0 1.0 4.0 20.0 16.0 5.0  jazz 2.0 4.0 5.0 9.0 5.0 25.0  latin-american 2.0 2.0 7.0 10.0 6.0 6.0  pop 7.0 1.0 1.0 13.0 7.0 1.0  r-n-b 3.0 4.0 5.0 13.0 14.0 3.0  reggae 1.0 5.0 2.0 20.0 15.0 3.0  shanson 2.0 1.0 4.0 2.0 2.0 1.0  world 2.0 8.0 5.0 7.0 8.0 9.0  latin-american pop r-n-b reggae shanson world  allrock 7.0 5.0 7.0 2.0 20.0 10.0  blues 6.0 4.0 10.0 4.0 7.0 8.0  country 8.0 4.0 5.0 4.0 17.0 8.0  electronic 3.0 7.0 4.0 1.0 6.0 4.0  hip-hop 9.0 9.0 7.0 6.0 12.0 6.0  jazz 7.0 2.0 5.0 3.0 9.0 27.0  latin-american 11.0 6.0 7.0 4.0 30.0 9.0  pop 3.0 14.0 10.0 3.0 23.0 4.0  r-n-b 6.0 6.0 8.0 7.0 11.0 4.0  reggae 7.0 1.0 9.0 13.0 18.0 5.0  shanson 5.0 4.0 2.0 7.0 65.0 1.0  world 8.0 4.0 5.0 5.0 27.0 15.0      Модель с мощностью обучающей выборки 70%, тестовой выборки 30%  confusion\_matrix  allrock blues country electronic hip-hop jazz \  allrock 20.0 3.0 9.0 15.0 8.0 2.0  blues 10.0 18.0 16.0 7.0 4.0 9.0  country 9.0 8.0 11.0 4.0 7.0 6.0  electronic 11.0 4.0 4.0 34.0 12.0 4.0  hip-hop 2.0 1.0 4.0 13.0 15.0 4.0  jazz 2.0 5.0 5.0 9.0 2.0 25.0  latin-american 1.0 2.0 7.0 9.0 6.0 7.0  pop 5.0 1.0 1.0 10.0 10.0 0.0  r-n-b 3.0 4.0 6.0 11.0 12.0 1.0  reggae 1.0 4.0 2.0 17.0 15.0 2.0  shanson 1.0 1.0 5.0 2.0 2.0 1.0  world 4.0 8.0 1.0 7.0 8.0 3.0  latin-american pop r-n-b reggae shanson world  allrock 6.0 6.0 6.0 2.0 19.0 8.0  blues 3.0 4.0 9.0 4.0 7.0 9.0  country 8.0 5.0 7.0 2.0 14.0 4.0  electronic 2.0 6.0 4.0 1.0 5.0 5.0  hip-hop 5.0 10.0 6.0 5.0 10.0 5.0  jazz 5.0 2.0 3.0 3.0 7.0 21.0  latin-american 8.0 5.0 6.0 4.0 25.0 7.0  pop 0.0 10.0 8.0 2.0 20.0 4.0  r-n-b 5.0 1.0 9.0 8.0 8.0 5.0  reggae 8.0 2.0 8.0 9.0 14.0 3.0  shanson 4.0 4.0 1.0 5.0 55.0 0.0  world 4.0 3.0 3.0 5.0 26.0 13.0  accuracy = 0.21996124031007752  precision = 0.219467182545339  recall = 0.2142474151329102  f\_measure = 0.21682588881512993  Модель с мощностью обучающей выборки 75%, тестовой выборки 25%  confusion\_matrix  allrock blues country electronic hip-hop jazz \  allrock 16.0 3.0 11.0 13.0 6.0 0.0  blues 10.0 18.0 12.0 6.0 3.0 10.0  country 6.0 9.0 7.0 5.0 7.0 6.0  electronic 9.0 4.0 3.0 28.0 12.0 3.0  hip-hop 2.0 0.0 2.0 10.0 8.0 3.0  jazz 1.0 5.0 2.0 7.0 3.0 21.0  latin-american 1.0 0.0 5.0 6.0 4.0 5.0  pop 4.0 0.0 1.0 9.0 7.0 1.0  r-n-b 2.0 2.0 5.0 11.0 10.0 2.0  reggae 1.0 3.0 1.0 14.0 9.0 0.0  shanson 0.0 1.0 4.0 2.0 2.0 0.0  world 4.0 6.0 1.0 7.0 5.0 4.0  latin-american pop r-n-b reggae shanson world  allrock 4.0 5.0 5.0 3.0 17.0 7.0  blues 2.0 4.0 6.0 3.0 6.0 6.0  country 7.0 3.0 4.0 1.0 14.0 5.0  electronic 4.0 4.0 2.0 3.0 5.0 2.0  hip-hop 3.0 9.0 5.0 4.0 7.0 5.0  jazz 7.0 2.0 3.0 3.0 9.0 15.0  latin-american 7.0 0.0 5.0 8.0 22.0 8.0  pop 1.0 5.0 7.0 1.0 14.0 2.0  r-n-b 5.0 1.0 9.0 6.0 9.0 4.0  reggae 6.0 2.0 7.0 8.0 11.0 3.0  shanson 2.0 4.0 1.0 3.0 47.0 0.0  world 5.0 2.0 3.0 5.0 23.0 10.0    Модель с мощностью обучающей выборки 80%, тестовой выборки 20%  confusion\_matrix  allrock blues country electronic hip-hop jazz \  allrock 16.0 2.0 7.0 11.0 5.0 0.0  blues 8.0 13.0 12.0 4.0 2.0 7.0  country 4.0 9.0 7.0 1.0 6.0 5.0  electronic 8.0 4.0 1.0 20.0 9.0 2.0  hip-hop 1.0 1.0 2.0 8.0 4.0 3.0  jazz 1.0 7.0 2.0 9.0 1.0 14.0  latin-american 1.0 0.0 5.0 5.0 2.0 6.0  pop 2.0 0.0 1.0 6.0 4.0 1.0  r-n-b 1.0 1.0 4.0 9.0 7.0 1.0  reggae 1.0 3.0 0.0 9.0 7.0 0.0  shanson 0.0 0.0 3.0 2.0 2.0 0.0  world 2.0 3.0 2.0 6.0 3.0 3.0  latin-american pop r-n-b reggae shanson world  allrock 5.0 4.0 2.0 3.0 12.0 6.0  blues 2.0 4.0 6.0 2.0 6.0 4.0  country 8.0 2.0 3.0 2.0 13.0 3.0  electronic 3.0 4.0 2.0 2.0 2.0 2.0  hip-hop 4.0 6.0 4.0 3.0 5.0 3.0  jazz 7.0 2.0 1.0 1.0 7.0 10.0  latin-american 6.0 1.0 4.0 5.0 18.0 7.0  pop 1.0 5.0 7.0 1.0 12.0 2.0  r-n-b 4.0 0.0 8.0 8.0 9.0 3.0  reggae 6.0 1.0 5.0 7.0 9.0 2.0  shanson 0.0 2.0 1.0 3.0 36.0 1.0  world 6.0 2.0 2.0 5.0 17.0 9.0    Модель с мощностью обучающей выборки 85%, тестовой выборки 15%  confusion\_matrix  allrock blues country electronic hip-hop jazz \  allrock 14.0 1.0 8.0 9.0 3.0 0.0  blues 4.0 11.0 10.0 5.0 2.0 5.0  country 3.0 5.0 7.0 2.0 2.0 2.0  electronic 5.0 5.0 1.0 16.0 6.0 1.0  hip-hop 0.0 0.0 1.0 7.0 2.0 2.0  jazz 0.0 5.0 3.0 4.0 0.0 10.0  latin-american 1.0 0.0 4.0 5.0 3.0 3.0  pop 1.0 0.0 1.0 5.0 2.0 0.0  r-n-b 1.0 1.0 2.0 7.0 7.0 2.0  reggae 0.0 2.0 0.0 8.0 5.0 0.0  shanson 0.0 0.0 2.0 2.0 2.0 0.0  world 3.0 4.0 1.0 5.0 3.0 2.0  latin-american pop r-n-b reggae shanson world  allrock 4.0 2.0 3.0 1.0 11.0 5.0  blues 1.0 2.0 4.0 3.0 5.0 4.0  country 4.0 2.0 2.0 1.0 9.0 4.0  electronic 4.0 2.0 3.0 1.0 1.0 1.0  hip-hop 2.0 5.0 1.0 3.0 3.0 2.0  jazz 6.0 2.0 1.0 1.0 4.0 9.0  latin-american 4.0 1.0 2.0 4.0 15.0 4.0  pop 1.0 6.0 4.0 0.0 8.0 3.0  r-n-b 3.0 1.0 6.0 5.0 6.0 2.0  reggae 2.0 1.0 3.0 5.0 8.0 1.0  shanson 0.0 2.0 0.0 3.0 23.0 0.0  world 4.0 3.0 1.0 3.0 14.0 5.0    Модель с мощностью обучающей выборки 90%, тестовой выборки 10%  confusion\_matrix  allrock blues country electronic hip-hop jazz \  allrock 9.0 0.0 5.0 3.0 3.0 0.0  blues 5.0 6.0 6.0 2.0 2.0 6.0  country 2.0 5.0 6.0 2.0 2.0 2.0  electronic 3.0 3.0 1.0 14.0 4.0 0.0  hip-hop 0.0 0.0 0.0 4.0 2.0 1.0  jazz 0.0 1.0 3.0 2.0 0.0 10.0  latin-american 1.0 1.0 3.0 2.0 2.0 3.0  pop 1.0 0.0 1.0 4.0 2.0 1.0  r-n-b 1.0 2.0 2.0 5.0 2.0 1.0  reggae 1.0 1.0 0.0 4.0 2.0 0.0  shanson 0.0 0.0 2.0 2.0 1.0 0.0  world 0.0 4.0 1.0 4.0 2.0 1.0  latin-american pop r-n-b reggae shanson world  allrock 4.0 2.0 2.0 0.0 7.0 5.0  blues 1.0 1.0 3.0 2.0 1.0 1.0  country 2.0 1.0 1.0 1.0 4.0 3.0  electronic 2.0 2.0 3.0 1.0 1.0 0.0  hip-hop 0.0 0.0 1.0 2.0 1.0 2.0  jazz 5.0 1.0 0.0 1.0 4.0 5.0  latin-american 4.0 1.0 1.0 2.0 11.0 3.0  pop 0.0 4.0 1.0 0.0 6.0 0.0  r-n-b 2.0 0.0 4.0 4.0 2.0 1.0  reggae 1.0 0.0 3.0 4.0 4.0 1.0  shanson 0.0 0.0 1.0 3.0 17.0 0.0  world 1.0 2.0 2.0 1.0 8.0 5.0 |

Теперь возьмем тот же датасет со всеми 4 \* 12920 признаками.

Результаты работы алгоритма для файла «genres\_pca\_4na12920\_comp\_of\_song.csv» представлены на рисунке 3 и 4.

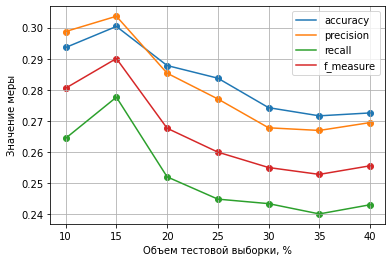


Рисунок 3 - Значения мер качества классификации при различных отношениях объема тестовой и обучающей выборки

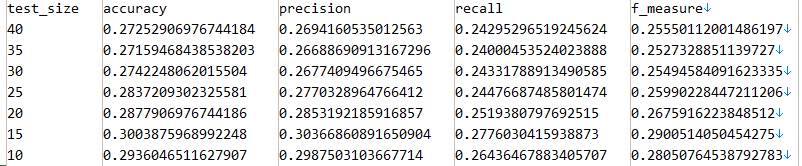


Рисунок 4 - Значения мер качества классификации в виде таблицы

Как видно из рисунков 3 и 4, результаты работы алгоритма на втором файле (содержащем все компоненты) заметно лучше. При увеличении объема тестовой выборки мы будем получать более стабильные результаты тестирования, но вместе с тем получаем меньшие значения ввиду уменьшения объема обучающей выборки.

Матрица неточностей представлена в таблице 2.

Таблица 2 – Сonfusion matrix

|  |
| --- |
| Модель с мощностью обучающей выборки 60%, тестовой выборки 40%  confusion\_matrix  allrock blues country electronic hip-hop jazz \  allrock 52.0 7.0 14.0 9.0 7.0 6.0  blues 31.0 34.0 11.0 3.0 7.0 20.0  country 20.0 23.0 24.0 3.0 3.0 17.0  electronic 14.0 5.0 3.0 21.0 18.0 15.0  hip-hop 8.0 2.0 4.0 20.0 26.0 5.0  jazz 1.0 11.0 4.0 1.0 6.0 76.0  latin-american 6.0 5.0 7.0 3.0 13.0 17.0  pop 5.0 0.0 4.0 4.0 15.0 5.0  r-n-b 9.0 8.0 8.0 5.0 20.0 9.0  reggae 1.0 2.0 2.0 13.0 26.0 5.0  shanson 7.0 2.0 4.0 0.0 3.0 2.0  world 8.0 4.0 7.0 0.0 4.0 26.0  latin-american pop r-n-b reggae shanson world  allrock 2.0 7.0 1.0 3.0 6.0 4.0  blues 3.0 2.0 2.0 0.0 3.0 9.0  country 1.0 8.0 1.0 4.0 7.0 10.0  electronic 2.0 10.0 6.0 11.0 10.0 4.0  hip-hop 5.0 20.0 5.0 12.0 8.0 1.0  jazz 5.0 1.0 3.0 2.0 4.0 8.0  latin-american 5.0 10.0 3.0 11.0 29.0 5.0  pop 5.0 15.0 8.0 9.0 26.0 4.0  r-n-b 1.0 14.0 5.0 12.0 5.0 4.0  reggae 9.0 9.0 12.0 25.0 9.0 3.0  shanson 4.0 5.0 3.0 3.0 74.0 1.0  world 9.0 2.0 6.0 8.0 25.0 18.0    Модель с мощностью обучающей выборки 65%, тестовой выборки 35%  confusion\_matrix  allrock blues country electronic hip-hop jazz \  allrock 49.0 7.0 12.0 6.0 9.0 6.0  blues 27.0 30.0 10.0 2.0 6.0 21.0  country 19.0 18.0 16.0 2.0 3.0 15.0  electronic 15.0 4.0 6.0 19.0 17.0 13.0  hip-hop 8.0 1.0 4.0 13.0 22.0 6.0  jazz 0.0 10.0 6.0 0.0 4.0 65.0  latin-american 5.0 5.0 6.0 4.0 12.0 17.0  pop 4.0 0.0 4.0 2.0 14.0 4.0  r-n-b 8.0 6.0 8.0 4.0 17.0 6.0  reggae 1.0 2.0 1.0 10.0 25.0 4.0  shanson 6.0 2.0 4.0 0.0 3.0 1.0  world 8.0 6.0 5.0 0.0 5.0 23.0  latin-american pop r-n-b reggae shanson world  allrock 4.0 5.0 2.0 4.0 6.0 4.0  blues 0.0 1.0 2.0 0.0 5.0 7.0  country 2.0 7.0 0.0 5.0 5.0 8.0  electronic 2.0 7.0 3.0 11.0 8.0 3.0  hip-hop 3.0 18.0 6.0 11.0 6.0 1.0  jazz 4.0 1.0 5.0 0.0 3.0 5.0  latin-american 3.0 7.0 3.0 11.0 24.0 3.0  pop 2.0 13.0 6.0 8.0 26.0 4.0  r-n-b 1.0 11.0 4.0 13.0 4.0 2.0  reggae 6.0 8.0 9.0 23.0 6.0 4.0  shanson 3.0 4.0 2.0 2.0 68.0 1.0  world 7.0 3.0 4.0 6.0 21.0 15.0    Модель с мощностью обучающей выборки 70%, тестовой выборки 30%  confusion\_matrix  allrock blues country electronic hip-hop jazz \  allrock 43.0 6.0 13.0 6.0 9.0 6.0  blues 24.0 30.0 10.0 1.0 4.0 16.0  country 14.0 14.0 16.0 3.0 2.0 14.0  electronic 14.0 4.0 5.0 18.0 12.0 11.0  hip-hop 6.0 1.0 4.0 12.0 16.0 5.0  jazz 0.0 8.0 5.0 0.0 5.0 55.0  latin-american 4.0 3.0 6.0 4.0 8.0 14.0  pop 3.0 0.0 2.0 4.0 11.0 3.0  r-n-b 6.0 5.0 6.0 2.0 17.0 6.0  reggae 1.0 1.0 1.0 4.0 23.0 5.0  shanson 5.0 0.0 4.0 0.0 3.0 1.0  world 7.0 6.0 3.0 0.0 6.0 16.0  latin-american pop r-n-b reggae shanson world  allrock 5.0 4.0 2.0 4.0 5.0 1.0  blues 1.0 3.0 2.0 0.0 4.0 5.0  country 4.0 6.0 0.0 4.0 4.0 4.0  electronic 2.0 5.0 2.0 8.0 8.0 3.0  hip-hop 4.0 15.0 3.0 10.0 3.0 1.0  jazz 4.0 1.0 2.0 1.0 4.0 4.0  latin-american 3.0 6.0 2.0 9.0 23.0 5.0  pop 2.0 11.0 3.0 5.0 24.0 3.0  r-n-b 1.0 10.0 2.0 11.0 4.0 3.0  reggae 8.0 5.0 9.0 20.0 5.0 3.0  shanson 0.0 4.0 3.0 2.0 58.0 1.0  world 5.0 3.0 2.0 4.0 22.0 11.0  Модель с мощностью обучающей выборки 75%, тестовой выборки 25%  confusion\_matrix  allrock blues country electronic hip-hop jazz \  allrock 40.0 8.0 10.0 6.0 6.0 5.0  blues 20.0 26.0 7.0 2.0 4.0 15.0  country 13.0 15.0 13.0 1.0 2.0 11.0  electronic 12.0 1.0 6.0 12.0 12.0 9.0  hip-hop 3.0 1.0 3.0 5.0 16.0 3.0  jazz 0.0 7.0 2.0 0.0 3.0 50.0  latin-american 1.0 4.0 5.0 1.0 5.0 10.0  pop 2.0 0.0 1.0 2.0 9.0 1.0  r-n-b 7.0 5.0 5.0 3.0 15.0 6.0  reggae 1.0 1.0 0.0 2.0 17.0 3.0  shanson 3.0 0.0 1.0 0.0 4.0 1.0  world 6.0 3.0 4.0 0.0 3.0 16.0  latin-american pop r-n-b reggae shanson world  allrock 4.0 5.0 2.0 1.0 2.0 1.0  blues 1.0 2.0 2.0 0.0 2.0 5.0  country 4.0 5.0 0.0 4.0 4.0 2.0  electronic 3.0 5.0 1.0 7.0 8.0 3.0  hip-hop 1.0 15.0 2.0 6.0 3.0 0.0  jazz 3.0 2.0 3.0 0.0 4.0 4.0  latin-american 3.0 6.0 2.0 8.0 19.0 7.0  pop 4.0 8.0 1.0 4.0 15.0 5.0  r-n-b 1.0 10.0 0.0 8.0 3.0 3.0  reggae 5.0 4.0 7.0 16.0 5.0 4.0  shanson 1.0 2.0 1.0 2.0 50.0 1.0  world 1.0 3.0 3.0 5.0 21.0 10.0    Модель с мощностью обучающей выборки 80%, тестовой выборки 20%  confusion\_matrix  allrock blues country electronic hip-hop jazz \  allrock 32.0 7.0 8.0 3.0 7.0 5.0  blues 19.0 22.0 7.0 1.0 3.0 8.0  country 12.0 10.0 13.0 2.0 1.0 8.0  electronic 10.0 1.0 6.0 9.0 10.0 7.0  hip-hop 2.0 1.0 2.0 4.0 12.0 3.0  jazz 0.0 7.0 1.0 0.0 2.0 38.0  latin-american 2.0 2.0 5.0 1.0 5.0 9.0  pop 2.0 0.0 0.0 1.0 8.0 1.0  r-n-b 6.0 5.0 5.0 2.0 11.0 5.0  reggae 1.0 1.0 0.0 3.0 13.0 1.0  shanson 1.0 0.0 1.0 0.0 3.0 1.0  world 5.0 3.0 2.0 0.0 4.0 13.0  latin-american pop r-n-b reggae shanson world  allrock 3.0 3.0 1.0 1.0 2.0 1.0  blues 1.0 1.0 1.0 0.0 3.0 4.0  country 4.0 5.0 0.0 3.0 4.0 1.0  electronic 3.0 4.0 0.0 3.0 4.0 2.0  hip-hop 1.0 10.0 2.0 4.0 3.0 0.0  jazz 5.0 0.0 2.0 0.0 4.0 3.0  latin-american 3.0 2.0 1.0 7.0 16.0 7.0  pop 3.0 7.0 1.0 3.0 14.0 2.0  r-n-b 2.0 9.0 0.0 6.0 2.0 2.0  reggae 3.0 3.0 5.0 13.0 4.0 3.0  shanson 0.0 1.0 1.0 2.0 40.0 0.0  world 1.0 2.0 3.0 5.0 13.0 9.0  Модель с мощностью обучающей выборки 85%, тестовой выборки 15%  confusion\_matrix  allrock blues country electronic hip-hop jazz \  allrock 27.0 5.0 9.0 1.0 7.0 2.0  blues 18.0 17.0 4.0 1.0 2.0 6.0  country 7.0 8.0 10.0 0.0 0.0 6.0  electronic 8.0 0.0 5.0 9.0 8.0 4.0  hip-hop 2.0 0.0 1.0 2.0 8.0 2.0  jazz 0.0 6.0 1.0 0.0 2.0 31.0  latin-american 1.0 2.0 4.0 0.0 4.0 8.0  pop 0.0 0.0 0.0 1.0 5.0 1.0  r-n-b 5.0 3.0 4.0 2.0 11.0 2.0  reggae 1.0 1.0 0.0 0.0 8.0 1.0  shanson 1.0 0.0 0.0 0.0 3.0 1.0  world 4.0 1.0 3.0 0.0 3.0 9.0  latin-american pop r-n-b reggae shanson world  allrock 3.0 3.0 0.0 1.0 2.0 1.0  blues 0.0 2.0 0.0 0.0 1.0 5.0  country 2.0 5.0 0.0 2.0 3.0 0.0  electronic 2.0 4.0 0.0 2.0 2.0 2.0  hip-hop 1.0 7.0 2.0 2.0 1.0 0.0  jazz 1.0 0.0 1.0 0.0 2.0 1.0  latin-american 3.0 2.0 1.0 5.0 12.0 4.0  pop 3.0 7.0 0.0 3.0 9.0 2.0  r-n-b 1.0 8.0 0.0 5.0 0.0 2.0  reggae 3.0 2.0 3.0 11.0 3.0 2.0  shanson 0.0 0.0 1.0 2.0 26.0 0.0  world 3.0 1.0 3.0 3.0 12.0 6.0    Модель с мощностью обучающей выборки 90%, тестовой выборки 10%  confusion\_matrix  allrock blues country electronic hip-hop jazz \  allrock 17.0 3.0 7.0 2.0 2.0 1.0  blues 12.0 8.0 3.0 1.0 1.0 7.0  country 5.0 5.0 7.0 0.0 0.0 5.0  electronic 4.0 0.0 4.0 6.0 7.0 3.0  hip-hop 0.0 0.0 0.0 0.0 5.0 1.0  jazz 0.0 3.0 1.0 0.0 1.0 22.0  latin-american 1.0 2.0 4.0 0.0 3.0 6.0  pop 0.0 0.0 0.0 0.0 2.0 1.0  r-n-b 4.0 2.0 2.0 1.0 5.0 2.0  reggae 1.0 1.0 0.0 0.0 3.0 1.0  shanson 0.0 0.0 1.0 0.0 2.0 1.0  world 2.0 2.0 2.0 0.0 5.0 6.0  latin-american pop r-n-b reggae shanson world  allrock 3.0 2.0 0.0 1.0 1.0 1.0  blues 0.0 1.0 0.0 0.0 1.0 2.0  country 2.0 5.0 0.0 1.0 1.0 0.0  electronic 0.0 4.0 0.0 2.0 2.0 2.0  hip-hop 0.0 2.0 1.0 3.0 1.0 0.0  jazz 1.0 0.0 1.0 0.0 2.0 1.0  latin-american 1.0 2.0 0.0 4.0 8.0 3.0  pop 2.0 5.0 0.0 2.0 7.0 1.0  r-n-b 1.0 3.0 0.0 6.0 0.0 0.0  reggae 3.0 2.0 2.0 6.0 1.0 1.0  shanson 0.0 0.0 0.0 2.0 20.0 0.0  world 1.0 1.0 1.0 1.0 6.0 4.0 |