

Классификация

Беляков Дмитрий

December 27, 2021

1 Описание задачи

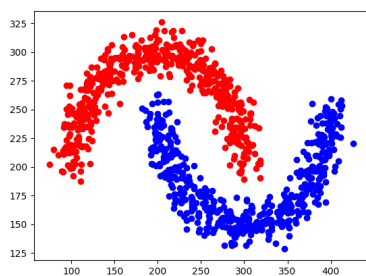
Задача: Сравнить алгоритмы классификации SVM(метод опорных векторов) и KNN (N ближайших соседей) на плоском наборе данных

2 Описание решения

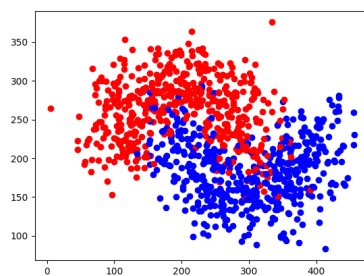
Используем классы SVC и KNeighborsClassifier библиотеки Sklearn. Для SVM будем использовать ядро RBF. Для KNN будем рассматривать 5 ближайших соседей. Сравним процент ошибки, время обучения и время предсказания алгоритмов.

3 Описание данных

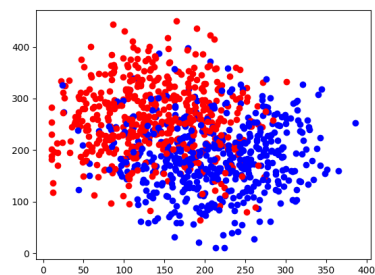
Рассмотрим несколько наборов данных из 1000 точек: два кластера в виде полумесяцев с разным уровнем шума. Три круглых кластера с разным уровнем шума. Три круглых кластера друг внутри друга. 70% данных будем использовать для обучения и 30 для предсказания.



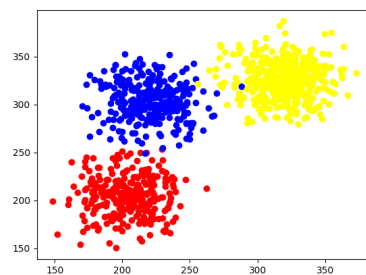
a)



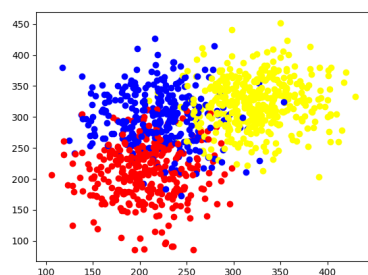
b)



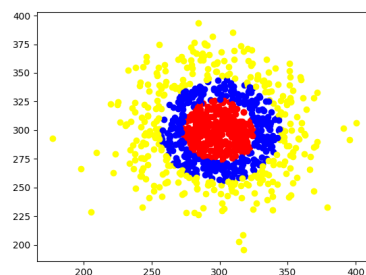
c)



d)



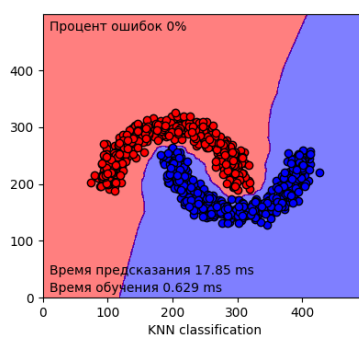
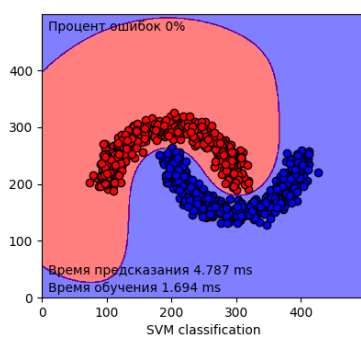
c)

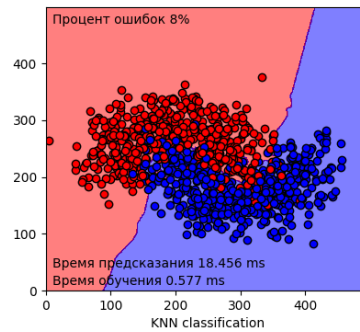
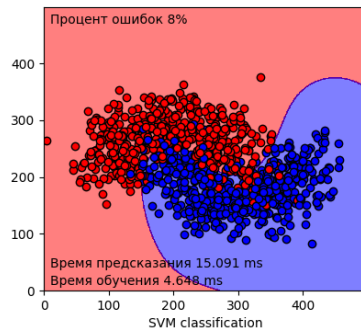
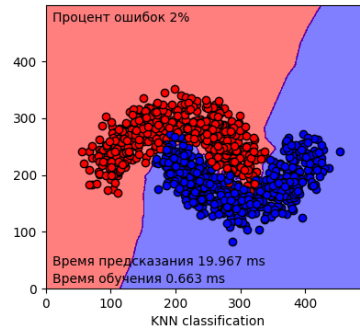
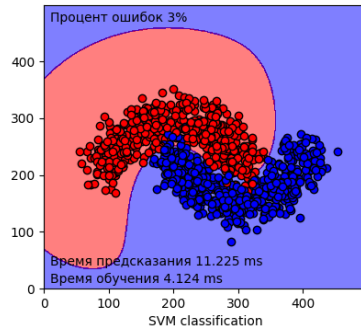


d)

4 Результаты

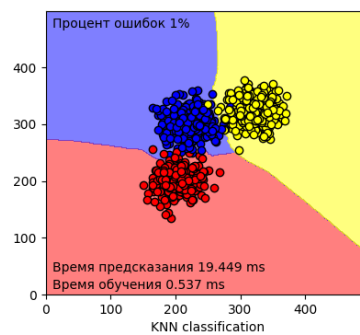
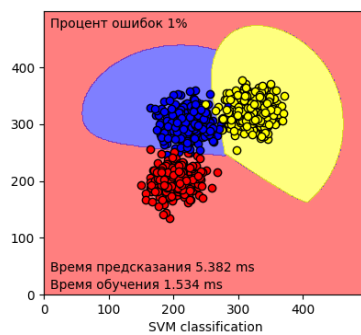
4.1 Тесты а-с

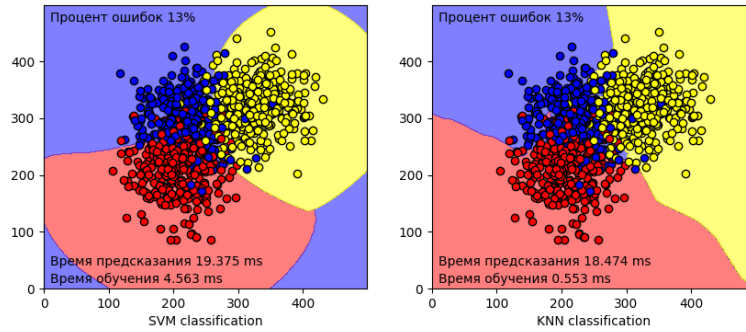




Как мы видим, SVM тратит больше времени на бчение чем KNN, что неудивительно так как KNN не выполняет предпосчета. Но при этом время которое тратит KNN на предсказание значительно больше чем SVM особенно на менее 'шумных' данных. Процент ошибок почти совпадает во всех случаях.

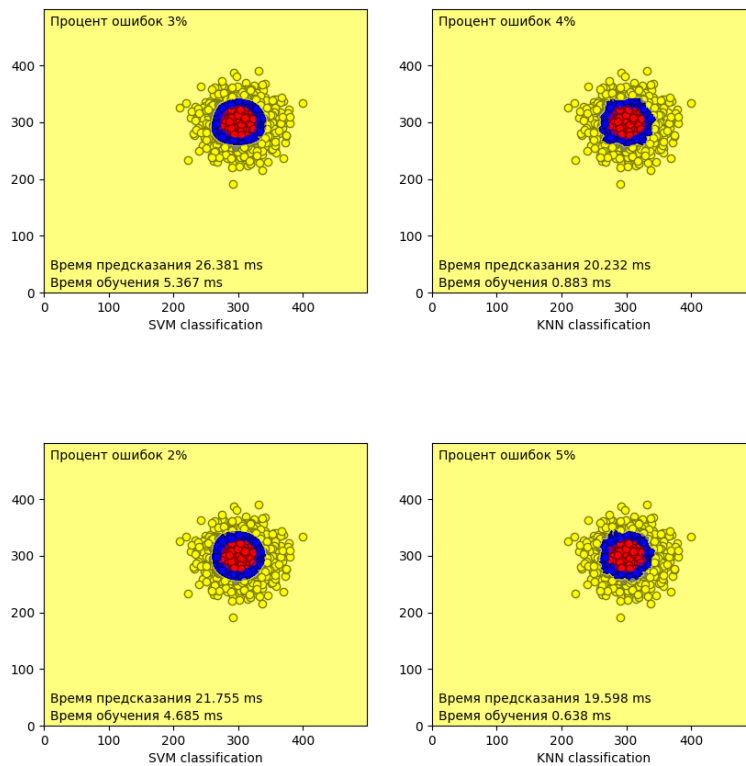
4.2 Тесты d-e





Для для наборов данных из трех кластеров результаты не сильно отличаются. Также подтверждается, что чем менее ровные границы у кластеров - тем больше времени SVM тратит на обучение

4.3 Тест f



В последнем тесте форма кластеров более сложная чем в предыдущих и в результате KNN совершает больше ошибок чем SVM

5 Вывод

Метод SVM занимает меньшее время на кластерах простой формы. Однако на кластерах сложной формы время работы KNN может быть даже ниже, но у SVM будет выше точность