Важная деталь интреф.scikit-learn – возможность вычислить оценки неопределенности прогнозом. Часто интересует не только класс, спрогнозированный моделью, но и степень уверенности модели в правильности прогноза.

В scikit-learn сущ-ет 2 различные ф-ции с пом.кот.м.оценить неопределенность прогнозов:

* Secision\_function
* Predict-Proba

Большая часть классификаторов позволяет использовать по крайней мере 1 из этих ф-ций.

Применим 2 ф-ции к синтетическому набору д-х, построив классификатор GrafientBoostingC, кот.позвол.исп-ть эти 2 метода. [1]

РЕШАЮЩАЯ ФУНКЦИЯ

В бин.классиф.возвращаемое знач.decision\_function имеет форму (n\_samples) [2]

Возвращ.знач.предст.собой число с плав.т.для к.примера.[3]

Знач.показ., насколько сильно модель уверенна в том, что т.д-х принадл. «полож» классу 1.

Полож.знач.указ.на предпочтение в пользу позиционного класса, а отриц знач – на предпочтение в сторону отрицательного класаа.

Мы м.судить о прогнозах, лишь взглянув на знак реш.ф-ции[4]

Для бин.классиф.отриц.знак – первый атрибут эл-та classes, а полож – второй.

Т.о.если вы хотите полностью просмотреть вывод метода predict, нужно воспользоваться атрибутом classes.[5]

Диап.знач.desicion\_function мб произв.и завис.от д-х и пар-ров модели [6]

Это произв.масштабирование часто затрудняет интерпретацию вывода decision\_function.

Построим decision\_function для всех т.2мерной пл-сти, исп-я цветовую кодировку и уже знакомую визуализацию решающей границы. Представим т.обуч.набора в виде кружков, а тест.д-е в виде треугольников [7]

Цветовая кодировка дает доп.инфу. однако в этой визуализации трудно разглядеть границу между 2 классами.

ПРОГНОЗИРОВАНИЕ ВЕРОЯТНОСТЕЙ

Вывод метода predict\_proba – верс-сть к.класса и ее легче понять, чем вывод метода desicicon

Для бин.классиф.им.форму (n\_samples, 2) [8]

Первый эл-т строки – оценка версти 1го классса, второй – оценка вер-сти 2го класса.

Т.к.речь идет о вер-ти, то знач.predict\_proba всегда нах.в между 0 и 1, а сумма их всегда 1 [9]

Т.к.вер-сти 2 классов в сумме дают 1 – 1 из этих классов всегда будет иметь определенность, превышающую 50%. Этот класс и будет спрогнозирован.

**Калиброванная модель –** если вычисленная неопред.соотв.фактической: в калибр.модели прогноз, полученный и 70%ой опред-стью будет правильным в 70% случаев.

Покажем границу принятия решения для набора д-х, а также вер-сти класса 1 [10]

Здесь границы распределены более четко, а небольние участки неопред.отчетливо видны.

НЕОПРЕДЕЛЕННОСТЬ В МУЛЬТИКЛАССОВОЙ КЛАССИФИКАЦИИ

Эти 2 метода м.применить в мультиклассовой классиф. Давайте применим к набору д-х Iris, кот.представл.собой пример 3классовой классиф. [11]

В мультикласс.класиф.decision\_fnction имеет форму (n\_samples, n\_classes) и к.столбец показ. Оценку определенности для к.класса, где высокая оценка ознакает меньшую вер-сть этого класса. Вы м.получить прогнозы, исходя из этих оценок, с пом.ф-ции **np.argmax**. Она возвращ.индекс макс.эл-та массива для к.т.д-х.[12]

Вывод predict\_proba имеет точно такую же форму. И снова вер-сти возм.классов для к.т.в сумме дают 1 [13]

Получим прогнозы вычислив argmax для predict\_proba [14]

Подводя итог, отметим, что predict\_proba и decision\_function всегда имеют форму (n\_samples, n\_classes), за исключением decision\_function в случае бинарной классификации. В бинарной классификации decision\_function имеет только один столбец, соответствующий «положительному» классу classes\_[1].

Для количества столбцов, равного n\_classes, вы можете получить прогноз, вычислив argmax по столбцам

Если хотим сравнить рез-ты полученные с пом.предсказания, с рез-тами эти 2 ф-ций:[15]