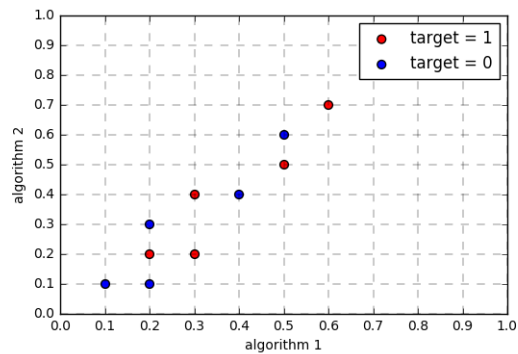


Машинное обучение

Теоретическое домашнее задание №3

Задача 1. Постройте ROC-кривые для двух алгоритмов, предсказания которых изображены на рисунке. Посчитайте AUC-ROC каждого из алгоритмов.



Задача 2. Ответьте на следующие вопросы:

- Какой алгоритм кросс-валидации стоит использовать, если мы имеем дело с сильно несбалансированной выборкой (доля положительного класса около 0.5%)?
- Чем отличается `ShuffleSplit(n_iter = 5)` и кросс-валидация по пяти блокам?
- Как правильно сформировать валидационную выборку, в случае если выборка очень маленькая (100 объектов)?
- Как соотносятся дисперсия качества алгоритма и количество блоков в кросс-валидации?
- Как соотносятся качество алгоритма и размер обучающей выборки?

Задача 3. Пусть имеется линейно разделимая выборка в задаче бинарной классификации. Как будут себя вести веса логистической регрессии при обучении градиентным спуском?