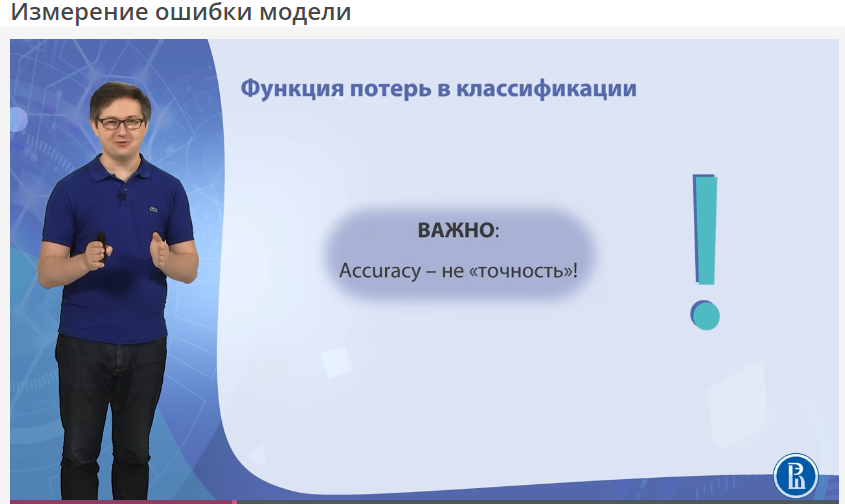


y – ответ, a – значение, выданное моделью. Если они не совпадают, штраф – функция потерь – равна 1, если они совпадают, то штраф равен 0.

Функционал ошибки – среднее значение этого индикатора.

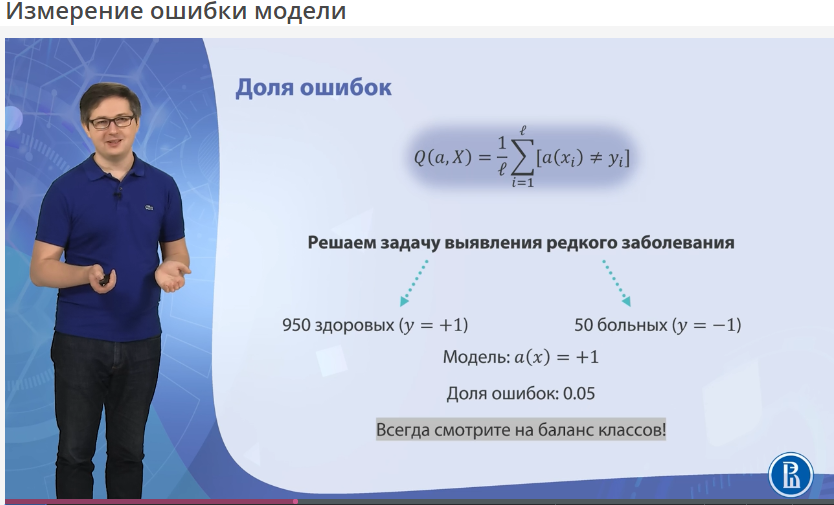


Точность = precision

Доля верных ответов = accuracy



У этого способа измерения ошибки есть проблемы



Если схалтурить, и взять модель, которая всегда выдает +1, то доля правильных ответов будет 95%

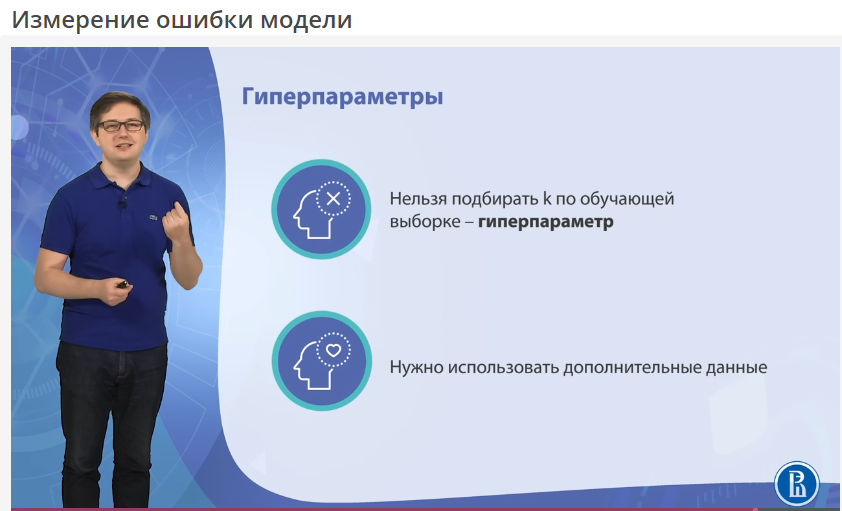
Это потому, что одного из классов очень мало.

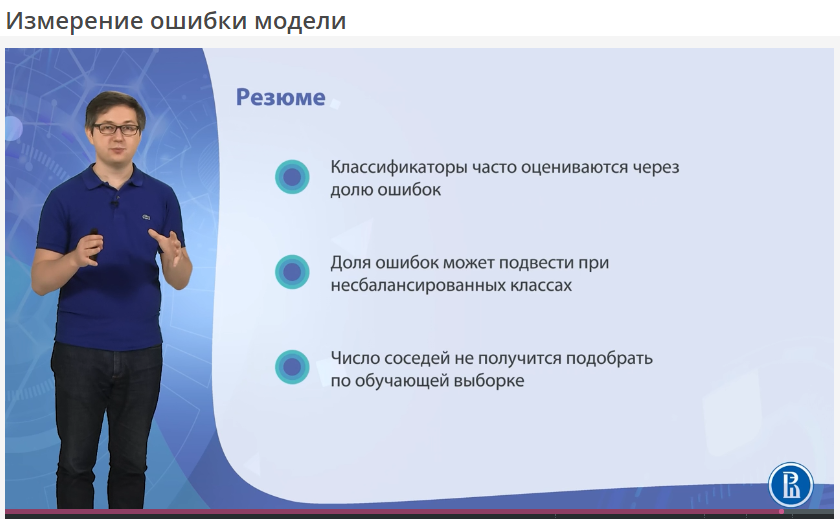


Если взять только 1 ближайшего соседа, то ошибок на обучающей выборке не будет



Модель с 20ю ближайшими соседями не пострадала из-за того, что отдельные синие или красные точки попали вглубь другого класса





Давайте проверим, насколько хорошо вы поняли материал прослушанной лекции.

2 из 3 баллов (не оценивается)

В выборке 5 объектов с ответами (-1, -1, -1, +1, +1). Модель выдала на этих объектах прогнозы (+1, +1, -1, -1, +1). Чему равна доля верных ответов?

0.8

0.4 верно

0.6

0.2

Почему доля верных ответов, равная 0.95, — это ещё не повод окончательно объявить, что задача успешно решена? Выберите все подходящие ответы.

Доля верных ответов — это в целом плохая мера качества модели, и её никогда не стоит использовать.

Могут быть задачи, где нужна более высокая доля верных ответов. верно

Выборка может быть несбалансированной, и в этом случае доля верных ответов 0.95 вполне может соответствовать константной модели. верно

В машинном обучении хорошей считается только та модель, которая даёт идеальное качество.

неверно

**Ответ**

Неверно:

У доли верных ответов есть свои недостатки, но всё-таки она очень понятна и может быть полезна, если не забывать о её ограничениях.

Всё-таки машинное обучение — это про построение приближенных решений. Мы решаем очень сложные задачи, и практически никогда не получается добиться нулевой ошибки.

Почему подбирать число соседей k в методе k ближайших соседей по обучающей выборке — плохая идея?

Лучшее значение k всегда будет равно единице, потому что тогда доля ошибок на обучающей выборке будет равна нулю. верно

Лучшее значение k всегда будет равно числу объектов в выборке, потому что тогда доля ошибок на обучающей выборке будет равна нулю.

Если подбирать k по всей обучающей выборке, то нет гарантий, что будет найдено правильное значение. Лучше делать это по небольшой выборке из 5-10 объектов.

верно

**Ответ**

Верно:

Да, потому что если взять объект из обучающей выборки, то ближайшим к нему будет он сам, и модель просто будет выдавать правильный ответ.