

# Домашнее задание №5 по курсу "Машинное обучение"

Прокопенко Тимофей

## Задача 1.

### Решение:

В алгоритме k-fold параметр k влияет на часть выборки, используемую для валидации ( $1/k$ ).

Его нужно выбирать большим, если размер выборки невелик. В таком случае будет хорошее качество тренировки, валидация показывает нестабильные результаты, но усреднение даст приемлемое значение, которое можно принять в рассмотрение.

K выбирают маленьким, если в выборке много объектов. Тогда получается тренировка на репрезентативном количестве данных и хорошая валидация.

## Задача 2.

### Решение:

Тренировочной и валидационной выборки не хватает для решения практических задач, так как существует риск переобучения. Тестовые данные никогда не доступны для во время обучения, следовательно, модель не может приспособиться к решению задачи на конкретных экземплярах.

Также на валидационной выборке мы выбираем лучшую из обученных конкретным алгоритмом моделей, а на тестовой проверяем, насколько эта лучшая модель корректна для решения задачи в общем случае.

## Задача 3.

### Решение:

Недостатки:

1. Число всевозможных разбиений выборки на две части очень велико, таким образом алгоритм "leave-all-out" будет работать очень долго.
2. Все возможные разбиения на две выборки включают в себя очень маленькие тренировочные и валидационные наборы данных, то есть очень неадекватные оценки будут включены в итоговую оценку наравне с корректными значениями.

Преимущества:

1. Не зависит от k, таким образом мы каждый раз получаем оценки зависящие только от качества алгоритма обучения модели, а не от параметра.
2. Различный размер тренировочных выборок позволяет проверить склонность модели к переобучению.