**Семинар 7. Бустинг. Стекинг.**

**Задание 7.2.1**

Выберите утверждения, верные для алгоритма *AdaBoost*:

* выбирает подвыборки из датасета
* выбирает подвыборки из датасета и меняет веса объектов ✅
* строит базовые модели параллельно
* строит базовые модели последовательно ✅
* уменьшает разброс (variance) ✅
* уменьшает разброс (variance) и смещение (bias)

**Задание 7.2.2**

Какие из перечисленных ниже утверждений верны для алгоритма *AdaBoost*?

* Использует мало параметров и прост в имплементации
* Чувствителен к выбросам ✅
* Не чувствителен к зашумленным данным
* Устойчив к переобучению (когда в данных мало шума) ✅
* Устойчив к переобучению (в случае сильно зашумленных данных)
* Необходимо адаптировать для случая несбалансированных классов ✅

**Задание 7.2.3**

Предположим, вы хотите применить *AdaBoost* к данным *D*, в которых *T* наблюдений. Вы разделяете данные пополам на обучающую и тестовую выборки. Теперь вы хотите увеличить количество наблюдений для обучения *T1, T2 … Tn*, где *T1* меньше *T2*, *T2* меньше *T3, …*, *Tn-1* меньше *Tn*.

Выберите верное утверждение.

* Разница между ошибкой на обучающей выборке и ошибкой на тестовой увеличивается с увеличением количества наблюдений
* Разница между ошибкой на обучающей выборке и ошибкой на тестовой уменьшается с увеличением количества наблюдений ✅
* Разница между ошибкой на обучающей выборке и ошибкой на тестовой не изменится
* Все варианты не подходят

**Стекинг:**

**Задание 7.5.1**

Как необходимо настраивать гиперпараметры для базовых алгоритмов в стекинге?

* На отложенной выборке
* При помощи кросс-валидации — на тех же фолдах, на которых проводится генерация признаков на первом этапе стекинга ✅
* На тренировочной выборке

**Задание 7.5.2**

На что влияет увеличение количество фолдов в стекинге?

* Увеличивается время обучения ✅
* Увеличивается количество используемых алгоритмов
* Увеличивается качество итоговой модели

**Задание 7.5.3**

Как применять стекинг для предсказаний (inference) на совершенно новых объектах?

* Помимо промежуточных моделей, обученных на фолдах, обучать ещё одну на всей промежуточной выборке и применять её для предсказаний как генератор признаков мета-модели ✅
* Сохранить все обученные на фолдах промежуточные модели и применить их на новых объектах с усреднением ответов

**Задание 7.5.4**

Какие из следующих утверждений верны для стекинга?

* Модель обучается на предсказаниях разных моделей машинного обучения ✅
* Логистическая регрессия будет работать лучше в качестве модели на второй стадии (как модель, агрегирующая результаты) по сравнению с другими методами классификации
* Модели на первой стадии обучаются на признаках исходного набора данных ✅

**Задание 7.5.5**

Какие из перечисленных ниже действий могут выполняться при стекинге?

* Разделить обучающую выборку на *k* фолдов ✅
* Обучить *k* моделей на каждом из *k-1* фолдов и получить предсказания для оставшегося фолда ✅
* Разделить тестовую выборку на *k* фолдов и получить предсказания для каждого из них с помощью разных алгоритмов

**Задание 7.5.6**

Вы используете стекинг с *n* методами машинного обучения и данными, разбитыми на *k* фолдов, для решения задачи бинарной классификации. Все базовые модели обучаются на всех признаках исходного датасета. Вы используете *k* фолдов для базовых моделей.

Какое утверждение верно для стекинга с одним уровнем (*m* базовых моделей и одна агрегирующая результаты модель)?

* После первой стадии у вас останется только *k* признаков ✅
* После первой стадии у вас останется только *m* признаков
* После первой стадии у вас останется только *k+m* признаков
* После первой стадии у вас останется только *k\*n* признаков
* Все утверждения неверны

**Задание 6.7.1**

В наборе данных 24 признака. Вы строите две модели: бэггинг над решающими деревьями и *RandomForest*. Пусть количество признаков, использующееся на одном разбиении для бэггинга над решающими деревьями равно А, для случайного леса — В.

Какое из следующих утверждений верно?

* А больше или равно В ✅
* А меньше В
* А больше В
* Нельзя сказать, так как на разных итерациях используется разное количество признаков

**Задание 6.7.2**

При решении задачи регрессии случайные леса имеют больший разброс (*variance*) в предсказаниях по сравнению с бустингом над решающими деревьями (при условии, что обе модели оптимизированы, то есть параметры к ним подобраны).

Верно ли это утверждение?

* Верно ✅
* Неверно
* Невозможно сказать, так как ответ зависит от данных

**Задание 6.7.3**

В реальной жизни часто приходится работать с большими наборами данных. Для этого важно знать, какие модели могут обучаться быстрее благодаря возможности распараллелить вычисления (например, в предыдущих уроках вы уже выставляли параметр n\_jonbs=-1 для этих целей).

В каких из моделей, перечисленных ниже, можно сделать вычисления (полностью или частично) параллельными?

* Random Forest ✅
* Gradient Boosted Trees
* XGBOOST (Extreme Gradient Boosting) ✅

**Задание 6.7.4**

Какие из приведенных ниже алгоритмов не используют темп обучения (*learning rate*) в качестве гиперпараметра?

* GradientBoosting
* ExtraTrees
* RandomForest ✅
* AdaBoost

**Задание 6.7.5**

Попробуйте вспомнить ансамблевые алгоритмы, которые вы изучили в этом и предыдущем модулях. Выберите их из списка ниже:

* Decision Tree
* RandomForest ✅
* GradientBoosting ✅
* AdaBoost ✅
* Linear Regression
* Logistic Regression
* Bootstrap

**Задание 6.7.6**

Выберите основные преимущества стекинга.

* Более устойчивая к выбросам модель ✅
* Более точное предсказание ✅
* Требуется меньшее время на обучение по сравнению с классическими моделями машинного обучения.