

Задания на заключительный этап  
олимпиады студентов "Я - Профессионал"  
по направлению "Искусственный интеллект"  
2020/2021 учебный год  
(КИМ для уровня "магистратура/специалитет")

1. Все задания заключительный этап Олимпиады делятся на два типа:
  - (a) (первого типа) задание на математические основы искусственного интеллекта, где требуется отправить решение в виде короткой программы на платформе Яндекс.Контест (таких заданий 5, и на решение их всех выделено 2 часа)
  - (b) (второго типа) задание, требующее решения конкретной практической задачи, связанной с применением методов искусственного интеллекта. (таких заданий 3, и на решение их всех выделено 4 часа)
2. Задания первого типа оцениваются исходя из 8 баллов. Баллы начисляются за успешное прохождение автоматических тестов программного кода, отправляемого участниками на платформу Яндекс.Контест (предварительные тесты не включаются).
3. Задания второго типа оцениваются исходя из 20 баллов. Каждая посылка решения участниками имеет какое-то количество очков. Очки линейно нормируются в баллы так, чтобы 20 баллов получила посылка решения участника с максимальным количеством очков. После отправки решения на Яндекс.Контест сразу автоматически будет выводиться количество баллов, вычисленное относительно идеального решения. По окончании времени выполнения заданий рейтинговые показатели будут пересчитаны в баллы относительно лучшего решения участников Олимпиады.
4. В случае подачи апелляции участником, она должна содержать номер/номера посылок, с оценкой которых участник не согласен.
5. Задания заключительного этапа олимпиады студентов "Я - Профессионал" по направлению "Искусственный интеллект" размещаются членами методической комиссии на платформе Яндекс.Контест <https://contest.yandex.ru>.

## 9 Обратное распространение ошибки. 8 баллов.

Задача предоставлена партнером Олимпиады - лабораторией интеллектуального транспорта МФТИ - НКБ ВС Центра когнитивного моделирования Московского физико-технического института.

Пусть дана вычислительная модель, показанная на рисунке 5, которую мы хотим обучить методом обратного распространения ошибки. Матрицы входных данных  $Q$  и  $V$  являются известными и заданными,  $W$  - матрица весов модели, которые изменются в ходе процесса обучения, известны ее начальные значения.  $T$  - вспомогательная матрица, вычисляемая в ходе работы модели.  $Y$  - выходная матрица или результат работы модели.

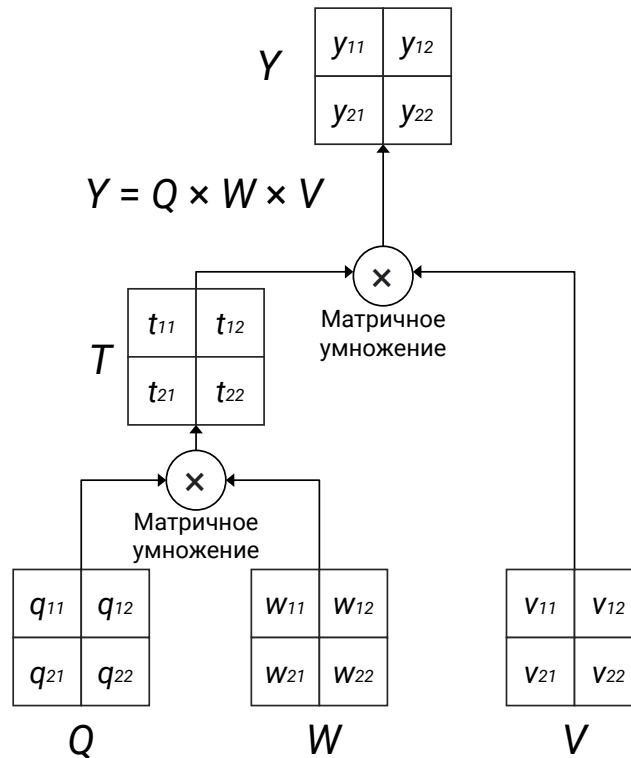


Рис. 5: Схематичное представление рассматриваемой модели

Указанная вычислительная модель является частью единой большой модели, для которой известно значение функционала качества (функции потерь)  $L$ .

Кроме того известны частные производные этого функционала качества  $L$  по значениям выходной матрицы модели  $Y$ :

$$\frac{\partial L}{\partial Y} = \begin{pmatrix} \frac{\partial L}{\partial y_{11}} & \frac{\partial L}{\partial y_{12}} \\ \frac{\partial L}{\partial y_{21}} & \frac{\partial L}{\partial y_{22}} \end{pmatrix}.$$

Требуется написать программу, которая, используя матрицы  $Q, V$ , а так же  $\frac{\partial L}{\partial Y}$  найдет частную производную функционала качества  $L$  по значениям матрицы весов модели  $W$ :

$$\frac{\partial L}{\partial W} = \begin{pmatrix} \frac{\partial L}{\partial w_{11}} & \frac{\partial L}{\partial w_{12}} \\ \frac{\partial L}{\partial w_{21}} & \frac{\partial L}{\partial w_{22}} \end{pmatrix} = ?$$

## 9.1 Формат ввода

Первая строка входных данных содержит четыре разделенных пробелом вещественных числа  $q_{11}$ ,  $q_{12}$ ,  $q_{21}$  и  $q_{22}$  ( $-1000.0 \leq q_{ij} \leq 1000.0$ ) — значения матрицы  $Q$ .

Вторая строка входных данных содержит четыре разделенных пробелом вещественных числа  $v_{11}$ ,  $v_{12}$ ,  $v_{21}$  и  $v_{22}$  ( $-1000.0 \leq v_{ij} \leq 1000.0$ ) — значения матрицы  $V$ .

Во третьей строке задаются четыре разделенных пробелом вещественных числа  $\frac{\partial L}{\partial y_{11}}$ ,  $\frac{\partial L}{\partial y_{12}}$ ,  $\frac{\partial L}{\partial y_{21}}$ ,  $\frac{\partial L}{\partial y_{22}}$  ( $-1000.0 \leq \frac{\partial L}{\partial y_{ij}} \leq 1000.0$ ) — значения частных производных функционала качества  $L$  по значениям выходной матрицы модели  $Y$ .

## 9.2 Формат вывода

Выведите четыре разделенных пробелом вещественных числа  $\frac{\partial L}{\partial w_{11}}$ ,  $\frac{\partial L}{\partial w_{12}}$ ,  $\frac{\partial L}{\partial w_{21}}$ ,  $\frac{\partial L}{\partial w_{22}}$  (с округлением до четвертого знака после запятой включительно) — значения частных производных функционала качества  $L$  по значениям матрицы весов  $W$ .

## 9.3 Комментарии

Рассматриваемая в задаче модель является упрощенным вариантом слоя Attention, широко распространенного и применяемого в глубоких нейронных сетях для задач анализа естественного языка и компьютерного зрения. Кроме того, обратное распространение ошибки через матричное произведение является важным для понимания процесса обучения нейронных сетей.

## 10 Загадочная строка. 8 баллов.

Археологи обнаружили подземелье с записями древних программистов. В нем находилась строка  $s$  бесконечной длины, а на стене была запись, из которой ученые сделали два вывода:

1. Каждый символ строки  $s$  является с вероятностью 50% нулем, а с вероятностью 50% единицей.
2. Рецепт бессмертия закодирован после первого вхождения последовательности бинарных символов  $w$ .

Перед тем, как начать рассматривать строку  $s$ , ученые решили найти математическое ожидание позиции первого вхождения слова  $w$ , в чем просят помощи у Вас!

### 10.1 Формат ввода

На единственной строке записана последовательность бинарных символов  $w$  ( $1 \leq |w| \leq 200$ ).

### 10.2 Формат вывода

Выведите математическое ожидание позиции первого вхождения  $w$  в строку  $s$  с точностью 6 знаков.

## 11 Обнаружение ботов. 8 баллов.

Вам дан фрагмент неориентированного графа социальной сети, вершинами которого являются пользователи, а ребро между двумя вершинами означает то, что соответствующие люди являются в данной социальной сети друзьями. Однако также в данной сети присутствуют боты, поведение которых отличается от поведения обычных пользователей.

Обычные пользователи в данной соцсети равномерно распределены на  $K$  групп по интересам. Внутри каждой группы вероятность дружбы между пользователями равна 0.95, в то время как между групп вероятность дружбы равна 0.05.

Боты же, стараясь мимикрировать под обычного пользователя, имеют в среднем столько же друзей, сколько и обычный пользователь, но друзья у него распределены равномерно по группам.

Дополнительно известно, что ботов в соцсети мало (не больше 5%).

Постарайтесь выследить всех ботов в соцсети!

### 11.1 Формат ввода

На первой строке записаны числа  $N$ ,  $M$  и  $K$  — количество вершин в социальном графе, количество ребер в социальном графе и количество групп соответственно ( $100 \leq N \leq 500$ ,  $1 \leq M \leq \frac{N(N-1)}{2}$ ,  $1 \leq K \leq 10$ ).

На последующих  $M$  строках записаны по два числа  $u_i$  и  $v_i$ , что означает что данные вершины соединены ребром ( $1 \leq u_i \neq v_i \leq N$ ).

### 11.2 Формат вывода

На первой строке выведите одно число  $Q$  — количество ботов в соцсети.

На второй строке выведите  $Q$  чисел, разделенных пробелами — номера вершин ботов в порядке возрастания.

### 11.3 Комментарии

Ваше решение будет оцениваться как усредненный по всем тестам F1-score для задачи классификации вершины, является ли она ботом.

## 12 Робот. 8 баллов.

Робот может перемещаться вдоль прямой, изначально стоит в координате 0. Ему приходит последовательность команд двух типов:

1. « $l$ » — сдвинуться на 1 влево
2. « $r$ » — сдвинуться на 1 вправо

Но из-за плохого соединения робот может некорректно распознавать команды, которые ему приходят. Например может принять « $l$ » за « $r$ » и наоборот, а может просто не распознать команду, тогда он с вероятностью  $p$  пойдет влево и с вероятностью  $(1 - p)$  вправо.

Найдите вероятность того, что робот все-таки доберется до той координаты, до которой должен был.

### 12.1 Формат ввода

На первой строке записана исходная последовательность команд  $a$ , состоящая из символов « $l$ » и « $r$ » ( $1 \leq |a| \leq 10^5$ ).

На второй строке записана полученная последовательность команд  $b$ , состоящая из символов « $l$ », « $r$ » и « $?$ » ( $1 \leq |b| \leq 10^5$ ).

На третьей строке записано число  $p$  с точностью до двух знаков ( $0 < p < 1$ ).

### 12.2 Формат вывода

Выведите одно число — вероятность того, что робот доберется до исходно заданной координаты, с точностью до 6 знаков.

## 13 Вирус. 8 баллов.

Задача предоставлена партнером Олимпиады - Российской ассоциацией искусственного интеллекта.

Вы исследуете новый компьютерный вирус, который распространяется по сети, которая задана связным неориентированным графом без петель. Известно, что вирус рационален. Он заражает узлы в порядке убывания их полезности.

Полезность узла  $i$  для заражения на шаге  $t$  задается функцией  $u(i, t) = w_1 x_1(i, t) + w_2 x_2(i, t)$ , где  $x_1$  — количество уже зараженных соседей узла  $i$ ,  $x_2$  — количество здоровых соседей узла  $i$ . Коэффициенты  $w_1, w_2$  принимают целые значения  $\{-1, 0, 1\}$ .

На шаге  $t = 0$  есть один зараженный узел. На шаге  $t$  вирус заражает узлы, для которых выполняются условия:

1.  $i = \arg \max_i u(i, t)$  — заражаются узлы с максимальной полезностью. Если таких узлов несколько, заражаются все они.
2.  $x_1 \geq 1$  — заражаются только узлы, смежные с уже зараженными.

Узлы остаются зараженными до конца распространения. Известна история распространения вируса на первых  $k$  шагах. Задание — вывести на печать траекторию распространения вируса до момента, когда распространение остановится.

Узлы сети пронумерованы от 0 до  $N - 1$ .

### 13.1 Формат ввода

В первой строке через пробел задано число узлов  $N$  ( $1 \leq N \leq 100$ ), число ребер  $M$  ( $N - 1 \leq M \leq \frac{N(N-1)}{2}$ ) и  $k$  ( $1 \leq k \leq N$ ) — длина известной истории распространения.

Далее в  $k$  строках задана история распространения вируса. Каждая строка содержит номера зараженных узлов.

В следующих строках задан список ребер сети в формате  $M$  строк из пар натуральных чисел « $u_i v_i$ » ( $0 \leq u_i < v_i < N$ ) через пробел.

### 13.2 Формат вывода

Программа должна вывести все подходящие комбинации  $w_1$  и  $w_2$ .

## 14 Ядерный реактор. 20 баллов

Задача предоставлена партнером олимпиады — АО "Концерн Росэнергоатом".

Ядерный реактор собран из шестигранных кассет с различным содержанием урана-235. Степень участия каждой кассеты в ядерной реакции, выработке энергии — относительная мощность кассеты  $Kq[i]$ , зависит от ее размножающих свойств и положения в активной зоне (номер  $[i]$  в массиве данных), см. рисунки 6, 7.

Размножающие свойства кассеты характеризуются коэффициентом размножения (отношение числа рожденных нейтронов к числу поглощенных в среде в единицу времени) бесконечной решетки, составленной из таких же одинаковых кассет и зависят от содержания урана-235 в кассете. Коэффициент размножения  $Kinf[i]$  рассчитывается по программе расчета кассеты, учитывающей уменьшение содержания урана-235 в процессе выработки энергии.

Определенные ранее  $Kinf[i]$  используются в программе расчета реактора. Основная цель — найти относительную мощность  $Kq[i]$  каждой из кассет. Теплоноситель — вода, прокачивается насосами через реактор и теплообменники, охлаждая реактор и перенося вырабатываемую реактором энергию к теплообменникам. Безопасность эксплуатации реактора зависит от предельных параметров теплоносителя в кассете (закипания) при снижении расхода теплоносителя во время срабатывания аварийной защиты (ввода поглощающих стержней в активную зону) из-за отключения насосов, поэтому максимальная мощность кассеты  $\max_i \{Kq[i]\}$  (далее  $maxKq$ ) ограничена.

*Обоснование безопасности* — это проверка выражения  $maxKq$  строго меньше предела  $Kq$ .

В начале работы реактор в целом имеет избыточные размножающие свойства, что и делает возможным его работу в течение длительного времени. По мере выработки энергии концентрация урана-235 уменьшается, коэффициент размножения реактора падает ниже 1.0 и реактор останавливается.

Коэффициент размножения реактора в любой момент работы должен быть равен 1.0 (если больше, то реактор разгонится, если меньше, остановится). Для обеспечения коэффициента размножения равным 1.0 в реакторе применяется борная кислота, растворённая в воде. Изотоп бор-10 поглощает лишние нейтроны. По мере выработки энергии и снижения содержания урана-235 из реактора борную кислоту удаляют. Концентрация борной кислоты  $Cb$  в начале работы может служить параметром, определяющим длительность работы реактора.

При проектировании топливной загрузки реактора — выбора положения каждой из кассет в реакторе — необходимо обеспечить максимальную длительность работы (максимальную  $Cb$ ), при соблюдении *обоснования безопасности*.

Расчет реактора — сложная процедура, занимающая много времени. Предлагается на основе предварительно рассчитанных данных создать нейросетевую модель реактора для быстрой оценки  $maxKq$ ,  $Cb$  при заданных  $Kinf[i]$  для различных положений кассет в активной зоне. Это позволит рассматривать при проектировании активной зоны значительно большее количество возможных вариантов и найти лучший.

### 14.1 Формат ввода

Ссылки на данные:

train.csv — <https://disk.yandex.ru/d/FPwrZ-zqtirE6A>

test.csv — <https://disk.yandex.ru/d/i4We87sxe9TEzA>

Вам предоставляются следующие файлы, содержащие данные нескольких вариантов массивов  $Kinf[i]$  и соответствующие им  $maxKq$ ,  $Cb$ .

train.csv — данные для обучения и тестирования ваших моделей, где  $maxKq$ ,  $Cb$  не равны 0. test.csv — данные, на которых будут проверяться ваши модели, где  $maxKq$ ,  $Cb$  равны 0.



Данные представлены в виде таблицы, где в первой колонке **maxKq**, во второй — **Сб** и колонки с номерами 3-55 — **Kinf[i]**

### 14.2 Формат вывода

Ответ принимается в формате csv в виде файла test\_out.csv, полностью идентичного по содержанию файлу test.csv, но с заполненными (не равными нулю) 1 и 2 колонками — **maxKq**, **Сб**.

Полученные результаты будут проверяться путем сопоставления с файлом test\_fill.csv (не доступен в процессе разработки модели).

### 14.3 Комментарии

Описание *обоснования безопасности* упрощенно. Данные для обучения и тестовые данные получены по двух групповой диффузионной программе переноса нейтронов. Расчет проводился для 1/6 части активной зоны (53 кассеты), т.к. активная зона симметрична в угле 60 градусов. Общее число кассет в реакторе 313 шт. — 1 центральная и 6 секторов по 52 кассеты. Нумерация в массиве **Kinf[]** соответствует нумерации ячеек активной зоны по рядам в 60 градусной симметрии. Данные о геометрии могут быть использованы при создании модели.

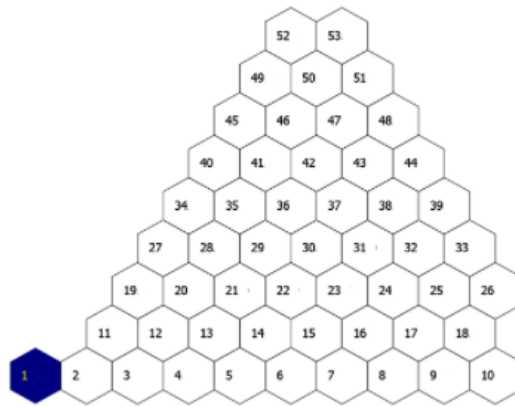


Рис. 6: Пример расположения кассет в 1/6 части активной зоны реактора

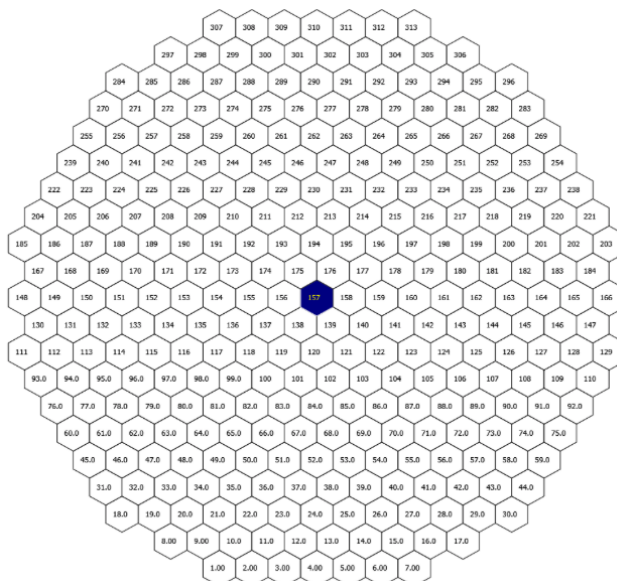


Рис. 7: Пример расположения кассет в активной зоне реактора

Баллы за посылку с *submission\_score* очками начисляются по следующей формуле:

$$\left[ 20 \cdot \frac{min\_score}{submission\_score} \right]$$

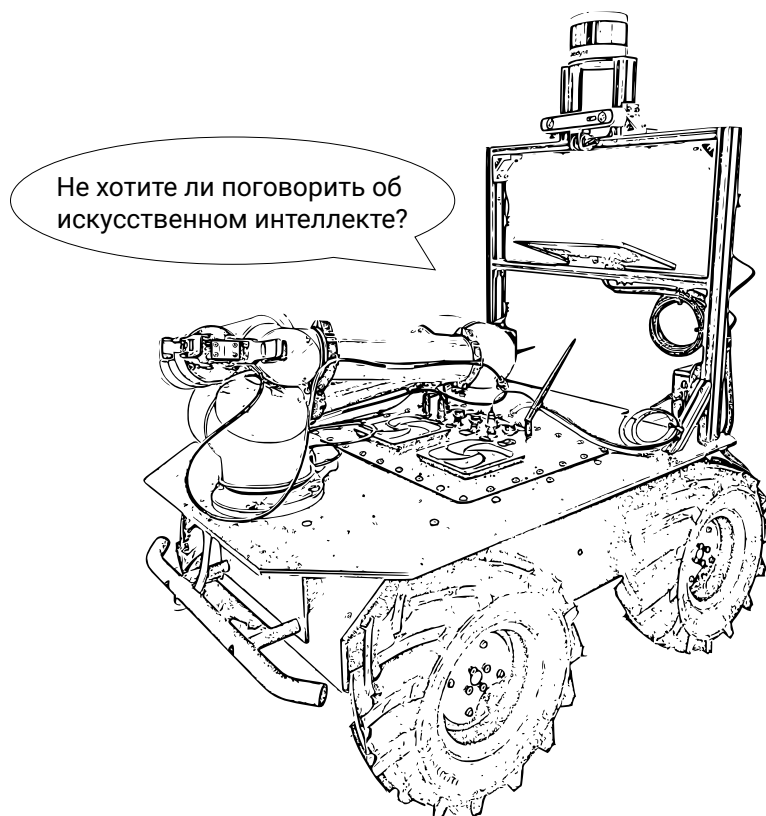
Где *min\_score* — максимальные очки по всем посылкам всех участников. В течение конкурса эта величина равна 0.1, а по окончании конкурса вычислится ее реальное значение, и баллы всех посылок будут пересчитаны.

Сам по себе *score* считается как усредненное евклидово расстояние между соответствующими точками в плоскости **maxKq-Cb**.

## 15 Обращение с подкреплением. 20 баллов

Задача предоставлена партнером олимпиады – Институтом проблем искусственного интеллекта ФИЦ ИУ РАН.

Робот Хаски выполняет важную задачу – оповещает студентов о летней школе по искусственному интеллекту. Согласно данным ему инструкциям, он передвигается между помещениями, одновременно предлагая поучаствовать в школе.



Робот настолько надоел всем своей навязчивой рекламой, что какие-то хулиганы стерли его память, по которой он выбирал свои действия. К счастью, после тщательного изучения, удалось восстановить часть *csv* файла, в котором хранилась история его действий в следующем виде: <состояние окружения>, <действие робота>, <вознаграждение>, <новое состояние окружения>. Вознаграждение – целое число  $-10 \leq r_t \leq 10$ , которое показывает реакцию людей на его действия. Состояние окружения – это некоторое внутреннее представление о мире, в котором он действует, задается строкой длины  $k = 5$ . Действие робота кодируется его номером  $0 \leq a \leq 4$ .

Ваша задача – восстановить память робота, назначив каждому состоянию и действию в этом состоянии вероятность, с которым роботу необходимо это действие выбирать, чтобы максимизировать его итоговое суммарное вознаграждение. Известно, что в каждом возможном состоянии у робота 5 возможных действий. Сумма вероятностей по всем действиям одного состояния должна равняться 1 или 100%, и не иметь отрицательных слагаемых. Итоговым результатом вашей работы будет суммарное вознаграждение, которое получит робот с использованием новой памяти. Известно, что во время каждого испытания памяти робот будет помещен в случайную комнату и сможет совершить ровно 100 действий, что соответствует длине одного эпизода.

Метрикой качества является нормированное среднее вознаграждение по  $N$  эпизодам взаимодействия робота со средой:

$$S = \left( \frac{\sum_{i=1}^N \sum_{t=1}^{100} r_t}{N} + 1000 \right) / 2000$$

## 15.1 Формат ввода

### [Ссылки на данные](#)

Вам предоставляется единственный файл *history.csv*, содержащий следующие колонки: состояние, действие, вознаграждение, следующее состояние.

## 15.2 Формат вывода

Вам нужно отправить единственный *csv* файл со следующими колонками: состояние, вероятность выбора каждого действия в процентах. Процент каждого действия должен быть целым неотрицательным числом. Сумма вероятностей по действиям одного состояния должна равняться 100%.

## 15.3 Комментарии

Баллы за посылку с *submission\_score* очками начисляются по следующей формуле:

$$\left[ 19 \cdot \frac{submission\_score - baseline\_score}{max\_score - baseline\_score} + 1 \right],$$

где *baseline\_score* - очки за базовое решение, *max\_score* — максимальные очки по всем посылкам всех участников. В течение конкурса эта величина равна 1.0, а по окончании конкурса вычислится ее реальное значение, и баллы всех посылок будут пересчитаны.

## 16 Распознай фейк. 20 баллов

Задача предоставлена партнером олимпиады — ПАО "Сбербанк".

Предлагается решить задачу бинарной классификации изображений для того, чтобы определить, на каких картинках изображены реальные люди, а какие сгенерированы нейронной сетью.

Для решения данной задачи предлагаются два набора данных: обучающий (train, размером 8 тысяч изображений) и тестовый (test, 12 тысяч изображений). Для каждого изображения из тестового набора нужно определить вероятность  $pred$  того, что это изображение - фейковое.

Основной метрикой является метрика accuracy (доля верно классифицированных изображений). Предсказанный для изображения класс в случае  $pred \geq 0.5$ , это класс 1 (фейковое), в случае  $pred < 0.5$  — это класс 0 (реальное изображение).

### 16.1 Формат ввода

[Ссылка на данные](#)

По ссылке выше лежат архив с изображениями, а также файл train.csv, в котором для изображений из обучающего набора train указаны истинные метки (где 1 — фейковое изображение, а 0 — настоящее).

### 16.2 Формат вывода

Результатом решения данной задачи должен быть файл submit.csv с двумя колонками — name и pred, где в первой будет название изображения из тестового набора test, а во второй — вероятность того, что это изображение является ненастоящим. Файл нужно отсортировать по колонке name.

### 16.3 Комментарии

Баллы за посылку с  $submission\_score$  очками начисляются по следующей формуле:

$$\left[ 19 \cdot \frac{submission\_score - baseline\_score}{max\_score - baseline\_score} + 1 \right],$$

где  $baseline\_score$  - очки за базовое решение,  $max\_score$  — максимальные очки по всем посылкам всех участников. В течение конкурса эта величина равна 1.0, а по окончании конкурса вычислится ее реальное значение, и баллы всех посылок будут пересчитаны.