

Олимпиада ДЛС. Финал для студентов+

13 июня 2024, 18:57:37
старт: 30 марта 2024, 09:00:00

[Объявления жюри](#)[Завершить](#)

...

[Задачи](#) [Посылки](#) [Сообщения](#) [Ответы](#)

A. Метрики и оптимизация (10)

У Миши есть задача бинарной классификации на классы A и B , причём известно, что в тестовой выборке количества объектов различных классов совпадают. Начальник предложил Мише построить бинарный классификатор и пояснил, что требуемая метрика качества — F_1 -мера. Миша обучил модель, закодировал класс A единицей, а класс B — нулём. Пусть y — вектор правильных ответов на тестовой выборке, а y' — вектор ответов модели. Миша посчитал для полученных предсказаний F_1 -меру $X = F_1(y, y')$.

Затем пришёл начальник Миши и заявил, что метрика F_1 будет считаться по другой схеме, где классы кодируются наоборот: класс A — нулём, а класс B — единицей. Миша не хочет больше обучать модель, а просто инвертирует классы и считает метрику F_1 с инвертированными классами: $Y = F_1(1 - y, 1 - y')$. Найдите максимальное возможное значение $X - Y$.

- 10 A. Метрики и оптимизация (10)
- 10 B. Линейная регрессия с MAPE (10)
- 10 C. Действие свёртки на ковариацию (10)
- 10 D. Упрощённый attention (10)
- 10 E. Просмотры видео (20)
- X F. Удалить водяную марку (20)
- 7.73 G. Статус студента (20)

Формат вывода

Ответ дайте в виде несократимой дроби вида p / q . Например: "1/2".

2/3

X

[Отправить](#)i осталось 19 попыток[Следующая](#)

Время посылки	ID	Задача	Компилятор	Вердикт	Тип посылки	Время	Память	Тест	Баллы
30 марта 2024, 09:33:17	110934204	A	No-compiler	OK	-	-	-	-	10 отчёт

Олимпиада ДЛС. Финал для студентов+

[Задачи](#) [Посылки](#) [Сообщения](#) [Ответы](#)

В. Линейная регрессия с MAPE (10)

Вася решает задачу предсказания переменной y по переменной x . Вася хочет построить для решения этой задачи линейную модель вида $y'(x) = a \cdot x$. В качестве loss-функции Вася выбрал Mean Absolute Percentage Error:

$$\text{MAPE}(y, y') = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - y'_i|}{y_i}.$$

Датасет доступен в файле [data.csv](#). Помогите Васе найти максимальное a , минимизирующее MAPE.

Формат вывода

Ответ запишите в виде несократимой дроби p / q .

13 июня 2024, 18:58:20

старт: 30 мар 2024, 09:00:00

[Объявления жюри](#)[Завершить](#)

...

10 А. Метрики и оптимизация (10)

10 В. Линейная регрессия с MAPE (10)

10 С. Действие свёртки на ковариацию (10)

10 D. Упрощённый attention (10)

E. Просмотры видео (20)

× F. Удалить водяную знак (20)

7.73 G. Статус студента (20)

3/2039

[Отправить](#)

i осталось 19 попыток

[Предыдущая](#)[Следующая](#)

Время посылки	ID	Задача	Компилятор	Вердикт	Тип посылки	Время	Память	Тест	Баллы
30 мар 2024, 10:15:24	110936364	В	No-compiler	OK	-	-	-	-	10 отчёт

Олимпиада ДЛС. Финал для студентов+

13 июня 2024, 18:58:45

старт: 30 марта 2024, 09:00:00

Объявления жюри

Завершить

...

Задачи Посылки Сообщения Ответы

С. Действие свёртки на ковариацию (10)

Пусть \mathcal{L} — слой свёртки 2×2 с фильтром $\begin{pmatrix} 1 & -1 \\ 2 & 1 \end{pmatrix}$. Ей на вход подаётся случайная матрица X размера 3×3 , элементы которой — случайные величины, независимые в совокупности и имеющие распределение $\mathcal{N}(0, 1)$ каждая. Найдите матрицу ковариаций случайного вектора $\mathcal{L}(X)$ — результата применения свёрточного слоя к матрице X . В ответ запишите сумму элементов искомой матрицы ковариаций.

Формат вывода

Введите одно целое число — значение суммы всех компонент искомой матрицы ковариаций.

- 10 А. Метрики и оптимизация (10)
- 10 Б. Линейная регрессия с MAPE (10)
- 10 С. Действие свёртки на ковариацию (10)
- 10 Д. Упрощённый attention (10)
- 20 Е. Просмотры видео (20)
- × F. Удалить водяную знак (20)
- 7.73 G. Статус студента (20)

Примечания

Матрицу $\mathcal{L}(X)$ можно воспринимать как вектор размера 4, если записать её компоненты в столбец. Порядок записи не имеет значения, так как сумма элементов матрицы ковариаций от этого не изменится.

34	×
----	---

Отправить осталось 17 попыток

Предыдущая Следующая

Время посылки	ID	Задача	Компилятор	Вердикт	Тип посылки	Время	Память	Тест	Баллы
30 мар 2024, 10:46:31	110938110	C	No-compiler	OK	-	-	-	-	10 отчёт
30 мар 2024, 10:45:13	110938026	C	No-compiler	OK	-	-	-	-	10 отчёт
30 мар 2024, 10:38:24	110937643	C	No-compiler	WA	-	-	-	1	0 отчёт

Олимпиада ДЛС. Финал для студентов+

🕒 13 июня 2024, 18:59:09

старт: 30 мар 2024, 09:00:00

Объявления жюри

Завершить

...

Задачи Посылки Сообщения Ответы

D. Упрощённый attention (10)

Петя построил нейросеть, один из слоев которой — QKV self-attention. Так вышло, что сохранить веса этого слоя Петя забыл. Однако Петя помнит, что матрицы имели размер 2×2 , матрица W_k была единичная, а матрица W_q — симметричная. У Пети остались входные векторы в слой self-attention:

$$x_1 = (1, 0), \quad x_2 = (0, 2), \quad x_3 = (1, 2), \quad x_4 = (-1, -2)$$

и диагональные элементы матрицы внимания до нормализации и функции активации softmax: $a = (3, 4, 9, 9)$. Помогите Пете восстановить матрицу W_q . В ответ запишите четыре значения матрицы W_q через пробел: $q_{11}, q_{12}, q_{21}, q_{22}$.

Формат вывода

Ответ дайте в виде четырёх несократимых дробей вида p / q через пробел. Например: "1/1 -1/2 1/3 1/4".

3/1 1/2 1/2 1/1



Отправить

осталось 19 попыток

Предыдущая

Следующая

Время посыпки	ID	Задача	Компилятор	Вердикт	Тип посыпки	Время	Память	Тест	Баллы
30 мар 2024, 11:11:09	110939639	D	No-compiler	OK	-	-	-	-	10 отчёт

10 A. Метрики и оптимизация (10)

10 B. Линейная регрессия с MAPE (10)

10 C. Действие свёртки на ковариацию (10)

10 D. Упрощённый attention (10)

E. Просмотры видео (20)

F. Удалить водяную знак (20)

7.73 G. Статус студента (20)

Олимпиада ДЛС. Финал для студентов+

13 июня 2024, 18:59:25
старт: 30 мар 2024, 09:00:00

Объявления жюри

Завершить

...

Задачи Псылки Сообщения Ответы

E. Просмотры видео (20)

Таня интересуется созданием видео на YouTube. Она хочет научиться делать такие видео, чтобы они набирали как можно больше просмотров. Ей стало интересно, можно ли по тексту (субтитрам) видео предсказать, сколько просмотров наберет видео. Помогите Тане это выяснить!

Формат ввода

[Ссылка на данные](#)

Вам даны тренировочная и тестовая выборки текстов. Тренировочная выборка содержит следующие столбы:

- id — уникальный id видео;
- Subtitles — субтитры видео;
- PublishDate — дата публикации видео;
- Category — категория видео;
- ViewCount (целевая переменная) — сколько просмотров набрало видео.

- 10 A. Метрики и оптимизация (10)
- 10 B. Линейная регрессия с MAPE (10)
- 10 C. Действие свёртки на ковариацию (10)
- 10 D. Упрощённый attention (10)
- 10 E. Просмотры видео (20)
- 10 F. Удалить водяную знак (20)
- 7.73 G. Статус студента (20)

Формат вывода

Тестовая выборка содержит все те же столбцы, кроме целевой переменной. Ваша задача — обучить модель машинного обучения на тренировочных данных и предсказать величину *ViewCount* для тестовых данных. В качестве решения нужно отправить ответы модели на тестовой выборке. Пример файла ответа можно найти по [ссылке](#).

Примечания

Ваше решение будет оцениваться как задача классификации по метрике *MAE*.

Количество баллов за посылку равно

$$20 \cdot \max \left(0, \frac{\text{MAXMAE} - \text{MAE}}{\text{MAXMAE}} \right),$$

где *MAE* — значение метрики *MAE* для вашего ответа, *MAXMAE* = 16.425.

ВНИМАНИЕ. После окончания конкурса баллы будут линейно проштрафлены таким образом, что участник с наилучшим решением по этой задаче получит ровно 20 баллов.

Набрать здесь

1

Отправить i осталось 20 попыток

i Посылок нет

Олимпиада ДЛС. Финал для студентов+

Задачи Постылок Сообщения Ответы

F. Удалить ватермарку (20)

Ограничение времени	20 секунд
Ограничение памяти	6.3 ГБ
Ввод	
Выход	tests.log

- 10 А. Метрики и оптимизация (10)
- 10 В. Линейная регрессия с MAPE (10)
- 10 С. Действие свёртки на ковариацию (10)
- 10 D. Упрощённый attention (10)
- E. Просмотры видео (20)
- F. Удалить ватермарку (20)
- 7.73 G. Статус студента (20)

Вам предстоит реализовать модель машинного обучения, которая будет решать задачу удаления ватермарки с изображения. В рамках данной задачи необходимо построить модель, которая принимает на вход изображение X размера 128×128 в формате RGB и возвращает новое изображение M размера 128×128 в формате RGB, которое является восстановленной версией изображения X , с которого удалена ватермарка.

Формат ввода

Данные доступны по [ссылке](#).

Вам предоставляются следующие файлы:

- Файл `train_corrupted_images.npy` содержит `np.array` формы $[N_{train_images}, 128, 128, 3]$ с изображениями, содержащими ватермарки. Здесь $N_{train_images} = 4000$.
- Файл `train_true_images.npy` содержит `np.array` формы $[N_{train_images}, 128, 128, 3]$ с настоящими изображениями до добавления ватермарки.
- Файл `test_corrupted_images.npy` аналогичен `train_corrupted_images.npy` для тестовой выборки, $N_{test_images} = 2000$.
- Файл `baseline_generation.py`, который считывает данные в правильном формате, делает предсказания (в качестве предсказания в базовом решении возвращаются изначальные картинки) и формирует посылку в контест.

Внимание! Данные довольно объёмные, поэтому поставьте их качаться заранее.

Формат вывода

Файл `sample_submission_generation.npy` является примером посылки в контест. Файл содержит `np.array` того же формата, что и `train_true_images.npy` размера $[500, 128, 128, 3]$.

ВНИМАНИЕ. Из-за проблем с Яндекс.Конкурсом мы не можем принимать файлы .npy слишком большого размера, из-за этого некоторые решения не проходили проверку. Поэтому нам пришлось поменять чекер. Ваш файл .npy должен содержать предсказания для первых 500 картинок. Это можно сделать с помощью кода

```
test_true_images = test_true_images[:500]
with open("sample_submission_generation.npy", "wb") as f:
    ...np.save(f, test_true_images.astype('uint8'), allow_pickle=False)
```

Итоговый shape посылки должен быть $[500, 128, 128, 3]$.

ВНИМАНИЕ ЧАСТЬ 2. Ваш файл с посылкой должен называться в формате `some_submission_xyz.npy`. Замечены случаи, когда система отвергает файлы с названиями типа `submission (1).npy`.

Примечания

Ваше решение оценивается по метрике MAE (Mean Absolute Error). Описание метрики и пример её вычисления приведены в базовом решении. Итоговое количество баллов по задаче равно

$$points = \max \left(0, \frac{baselinescore - score}{baselinescore} \cdot 20 \right),$$

где $score$ — метрика MAE вашего решения, а $baselinescore = 4$.

ВНИМАНИЕ. После окончания контеста баллы будут линейно прошkalированы таким образом, что участник с наилучшим решением по этой задаче получит ровно 20 баллов.

Выбрать

Файл не выбран

Отправить

1 осталось 19 попыток

Предыдущая

Следующая

Время посылки	ID	Задача	Компилятор	Вердикт	Тип посылки	Время	Память	Тест	Баллы
30 мар 2024, 14:46:09	110957478	F	(make) yandexdataschool	WA	-	-	-	1	0

Олимпиада ДЛС.
Финал для студентов+

13 июня 2024, 19:03:50 | Объявления жюри | Завершить
старт: 30 мар 2024, 09:00:00

Задачи Псылки Сообщения Ответы

G. Статус студента (20)

Тася занимается организацией образовательных программ. Тася хочет лучше работать со студентами и лучше понимать, какие факторы влияют на то, успешно ли студент закончил обучение в ВУзе. Она решила построить модель машинного обучения, которая могла бы по информации о студенте предсказывать, закончит ли студент обучение в вузе. Помогите Тасе это сделать.

Формат ввода

[Ссылка на данные](#)
Вам даны тренировочная и тестовая выборки текстов. Тренировочная выборка содержит следующие столбцы:

- ID — уникальный идентификатор студента;
- Код_группы — идентификатор группы студента;
- Год_Поступления — год поступления абитуриента;
- Пол — пол абитуриента;
- Основания — основания поступления абитуриента;
- Изучаемый_Язык — иностранный язык, который изучался абитуриентом;
- Дата_Рождения — дата рождения студента;
- Уч_Заведение — наименование учебного заведения;
- ГдеХодитсяУЗ — локация учебного заведения;
- ГодОкончанияУЗ — год окончания учебного заведения;
- Пособие — получает ли абитуриент пособия;
- Страна_ПП — страна постоянного проживания;
- Регион_ПП — регион постоянного проживания;
- Город_ПП — город постоянного проживания;
- Общежитие — требуется ли общежитие абитуриенту;
- Наличие_Матери — поле, показывающие наличие матери;
- Наличие_Отца — поле, показывающие наличие отца;
- Страна_Родители — страна постоянного проживания родителей;
- Опекунство — является ли абитуриент опекаемым;
- Село — проживает ли абитуриент в сельской местности;
- Иностранец — является ли абитуриент иностранным студентом;
- КодФакультета — идентификатор факультета;
- СрБаллАттестата — средний балл аттестата или средний балл ЕГЭ;
- Статус (целевая переменная): статус студента. 1, если студент отчислен, не закончив обучение, 0, если спешно закончил обучение.

Формат вывода

Тестовая выборка содержит все те же столбцы, кроме целевой переменной. Ваша задача — обучить модель машинного обучения на тренировочных данных и предсказать величину *label* для тестовых данных. В качестве решения нужно отправить ответы модели на тестовой выборке. Пример файла ответа можно найти по [ссылке](#).

Примечания

Ваше решение будет оцениваться как задача классификации по метрике *F1*.
Количество баллов за посылку равно

$$\max(0, 40 \cdot (F1 - 0.5)),$$

где *F1* — значение метрики для вашего ответа.

ВНИМАНИЕ. После окончания конкурса баллы будут линейно прошkalированы таким образом, что участник с наилучшим решением по этой задаче получит ровно 20 баллов.

Набрать здесь | Отправить файл

1

Отправить | осталось 17 попыток

Предыдущая

Время посылки	ID	Задача	Компилятор	Вердикт	Тип посылки	Время	Память	Тест	Баллы
30 мар 2024, 13:23:16	110949693	G	No-compiler	OK	-	-	-	-	7.73 отчёт
30 мар 2024, 13:20:46	110949462	G	No-compiler	OK	-	-	-	-	7.83 отчёт
30 мар 2024, 13:19:23	110949363	G	No-compiler	OK	-	-	-	-	8.06 отчёт