

Краткий tutorial по использованию Google Colab и подготовке рабочей среды

В нашем курсе машинного обучения предполагается выполнение заданий с использованием языка Python3 и связанных с ним инструментов. Чтобы облегчить вам выполнение практических заданий, а также предоставить вам вычислительные ресурсы (если у вас слабый ноутбук, например), мы предлагаем вам выполнять некоторые части заданий в среде Google Colab.

- Сайт проекта - <https://colab.research.google.com/>
- Вам необходим доступ к Google Drive. Скорее всего, ваша почта *gse.cs.msu.ru* уже имеет этот доступ. Если доступа по каким-то причинам нет, то можно зарегистрироваться в обычной почте *gmail.com*. Google Colab умеет работать с Google Drive как с файловой системой или хранилищем, что будет использоваться при выполнении вами практических заданий
- Google Colab представляет из себя интерактивную среду выполнения кода на Python наподобие среды Jupyter Notebook/Jupyter Lab. Если вы знакомы с последними, то перейти на Google Colab вам не составит труда.
- В Google Colab можно писать в ячейках код, выполнять его, смотреть на результат, писать текст с разным форматированием, вставлять картинки и прочее
- Хороший tutorial по ознакомлению с Google Colab можно найти вот тут <https://colab.research.google.com/notebooks/intro.ipynb?hl=ru-ru> . Также не воспрещается искать дополнительную информацию в интернете :)

Здесь мы приведем функционал, который понадобится вам при выполнении заданий

Python Intro [CMC ML Course] final version.ipynb ☆

File Edit View Insert Runtime Tools Help All changes saved

Locate in Drive
Open in playground mode

New notebook
Open notebook ⌘/Ctrl+O
Upload notebook

Rename
Move
Move to trash

Save a copy in Drive
Save a copy as a GitHub Gist
Save a copy in GitHub

Save ⌘/Ctrl+S
Save and pin revision ⌘/Ctrl+M S
Revision history

Download
Print ⌘/Ctrl+P

Индикатор запуска среды выполнения → Connect

При начале выполнения какой-либо ячейки как правило автоматически запускается

Мы выдаем вам ссылку, которая доступна только (!!!) на чтение. Чтобы начать выполнять ноутбук, вам необходимо скачать его себе на Google Drive. Копия откроется в соседнем табе. В этой копии вы уже можете сохранять изменения и выполнять задание

Python Intro [CMC ML Course] final version.ipynb ☆

File Edit View Insert Runtime Tools Help

Files

bin
boot
content
datalab
dev
etc
home
lib
lib32
lib64
media

Здесь вам доступны файлы из образа рантайма. Как правило, дополнительно скачанные файлы, необходимые для задания (данные), будут попадать в папку content

Python Intro [CMC ML Course] final version.ipynb ☆

File Edit View Insert Runtime Tools Help

Locate in Drive
Open in playground mode

New notebook
Open notebook ⌘/Ctrl+O
Upload notebook

Rename
Сохранить ноутбук локально в формате *.ipynb

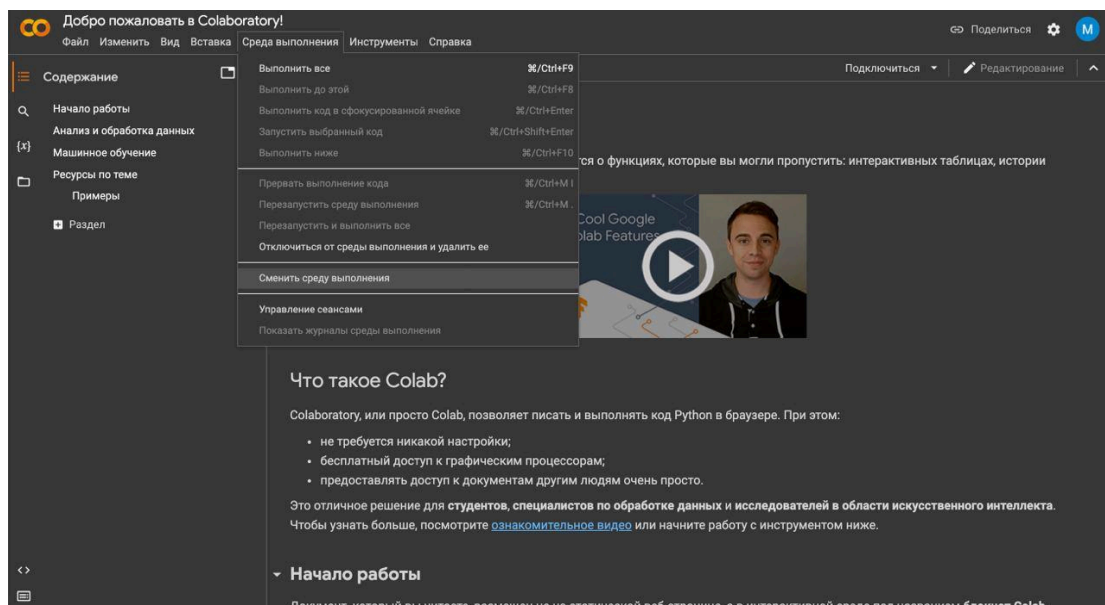
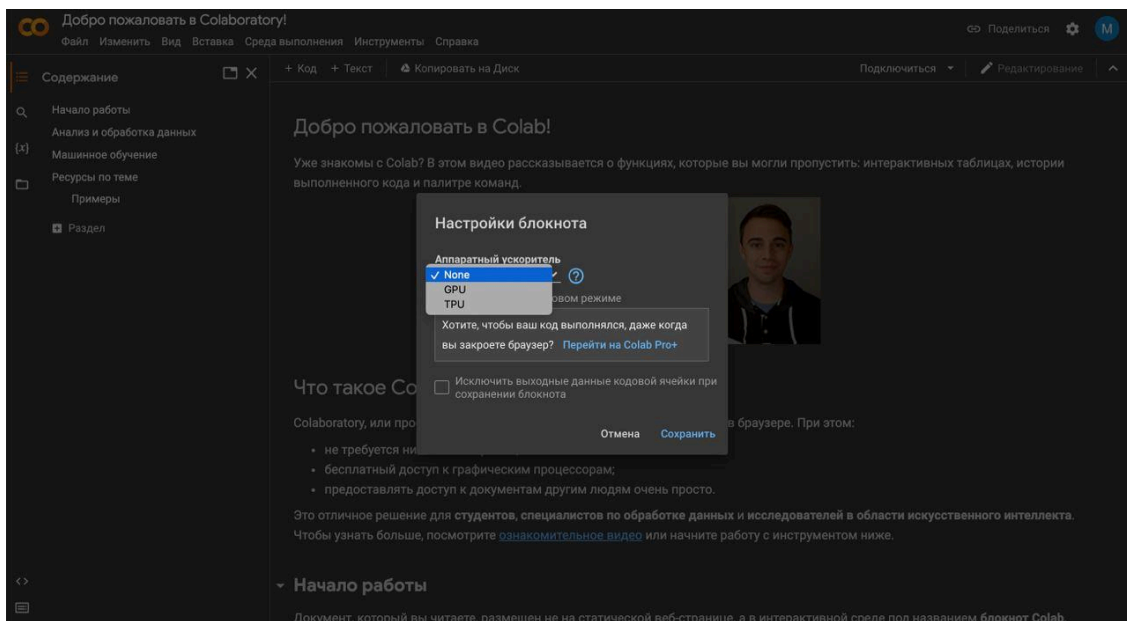
Save a copy in Drive
Save a copy as a GitHub Gist
Save a copy in GitHub

Save ⌘/Ctrl+S
Save and pin revision ⌘/Ctrl+M S
Revision history

Download
Print ⌘/Ctrl+P

Работа с GPU! (Актуально для курса по DL)

- Для обучения нейронных сетей и некоторых других алгоритмов (например, разных реализаций градиентных бустингов: xgboost, catboost, lightgbm) предпочтительнее использовать графические процессоры gpu. Их можно использовать, сменив runtime_type так, как показано ниже.
- Обратите внимание, что включить gpu в runtime_type недостаточно; нужно каким-то образом указать в коде (по-разному для каждого метода), что вы хотите использовать именно gpu. (например, для catboost – это параметр task_type = 'cuda', а для torch – torch.device("cuda").



Важное про использование runtime_type:

- При смене `runtime_type` ноутбук полностью перезапустится!

Важно!

- Google Colab имеет ограничения на используемую в сессии оперативную память и CPU, а также имеет ограничения на продолжительность выполнения одной сессии
- Хотя официально Google Colab не озвучивает ограничения <https://research.google.com/colaboratory/faq.html#usage-limits>, обычно часов через 12 активная запущенная сессия ноутбука прерывается <https://stackoverflow.com/questions/54057011/google-colab-session-timeout>
- Для нас это означает следующее: если вы хотите оставить ноутбук считаться (например на час) и пойти заниматься своими делами, то не закрывайте браузер и не переводите компьютер в спящий режим. Тогда ваши вычисления, скорее всего, не должны прерваться. Если все же вычисления прервались, можно попробовать сделать один из хаків, описанных тут <https://stackoverflow.com/questions/54057011/google-colab-session-timeout> или поискать еще варианты решений в интернете.
- После сохранения сделанной работы на локальный компьютер, откройте его на своем локальном Jupyter Notebook/Jupyter Lab/Visual Studio Code или пр. умеющим отображать *.ipynb приложением. Проверьте, что все требуемые выводы, картинки, код, вычисления отображаются -- без этого ваш ноутбук не может быть оценен корректно!
- В Google Colab вам предлагается делать только те части задания, которые идут на кросс-проверку (формата ipynb; проверяются вручную другими вашими однокурсниками). Задания типа юнит-тестирование или ML-решение предлагается делать локально, так как решения не требуют большого запаса вычислительных ресурсов. Также мы выдаем вам тесты и скрипт их запуска, на котором можно будет проверить ваше решение, что тоже гораздо удобнее делать на своем личном компьютере
- Если вы хотите делать задание формата ipynb также на своем личном компьютере, то вам предлагается установить виртуальное окружение с соответствующими версиями библиотек, которые вам выдали в начале курса (также их можно будет взять из проверяющей системы). Скорее всего, каких-то проблем с совместимостью в этом случае при выполнении практического задания быть не должно.