

Практическая работа № 2

Машинное обучение. Общая постановка задачи.

Оглавление

Машинное обучение. Общая постановка задачи.....	1
Цель работы.....	1
Задачи работы	1
Перечень обеспечивающих средств.....	2
Общие теоретические сведения	2
Обновление форка репозитория	2
Основные определения из лекции №1	4
Переобучение модели	4
Недообучение модели.....	4
Задание	5
Требования к отчету	7
Литература	7

Цель работы

- Получить практические навыки построения обучающей, валидационной и тестовой выборок.
- На практическом примере разобрать различия между параметрами и гиперпараметрами модели машинного обучения.
- На практических примерах научиться определять переобучение и недообучение модели машинного обучения.

Задачи работы

1. Изучить способ обновления своего форка из оригинального репозитория.
2. Реализовать и сравнить два способа формирования обучающей выборки.
3. Определить параметры и гиперпараметры приведённой модели машинного обучения.
4. По данным графикам ошибок на обучающей и тестовой выборках определить, наблюдается ли эффект переобучения или недообучения.

Перечень обеспечивающих средств

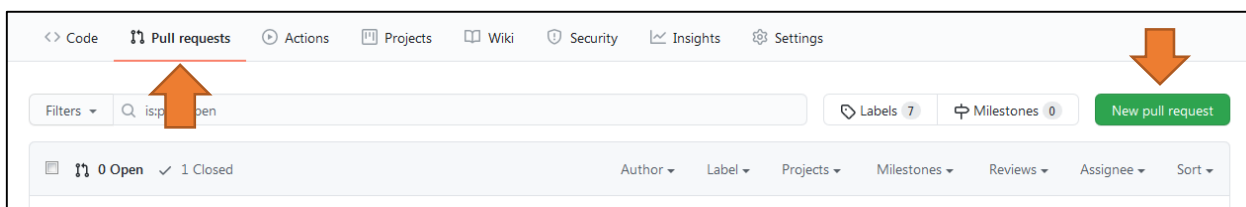
1. ПК.
2. Учебно-методическая литература.
3. Задания для самостоятельного выполнения.

Общие теоретические сведения

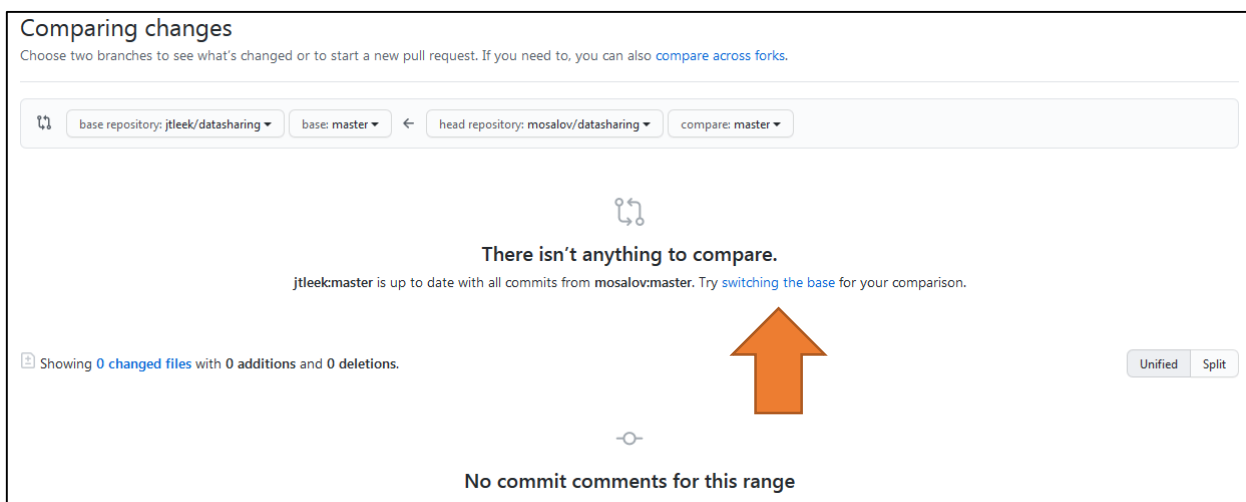
Обновление форка репозитория

Для того, чтобы обновить свой форк, т.е. получить все изменения, сделанные в оригинальном репозитории, необходимо выполнить следующие шаги в веб-интерфейсе GitHub:

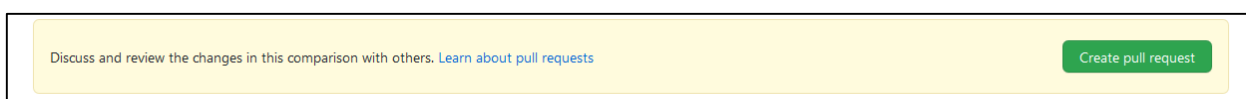
1. Откройте свой форк, перейдите на вкладку «Pull requests», нажмите кнопку «New pull request»:



2. В открывшемся окне «Comparing changes» найдите и нажмите ссылку «switching the base»:



3. Если страница не загрузится с первой попытки, просто перезагрузите её.
4. Нажмите кнопку «Create pull request»:



5. Укажите название пул-реквеста в поле «Title», например – «Обновление репозитория», и нажмите кнопку «Create pull request»:

Open a pull request

Create a new pull request by comparing changes across two branches. If you need to, you can also [compare across forks](#).

base repository: mosalov/practice_assignment


base: master

←

head repository: derekfranks/practice_assignment








compare: master

✓ **Able to merge.** These branches can be automatically merged.



Title


Write Preview

H B I  <>     @  


Leave a comment


Attach files by dragging & dropping, selecting or pasting them.


Create pull request

 Remember, contributions to this repository should follow our [GitHub Community Guidelines](#).

6. Внизу страницы нажмите кнопку «Merge pull request»:




 Continuous integration has not been set up
[GitHub Actions](#) and [several other apps](#) can be used to automatically catch bugs and enforce style.

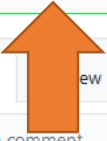






 **This branch has no conflicts with the base branch**
Merging can be performed automatically.

Merge pull request

You can also [open this in GitHub Desktop](#) or [view command line instructions](#).




Write Preview


H B I  <>     @  

Leave a comment

Attach files by dragging & dropping, selecting or pasting them.

 Close pull request

Comment

 Remember, contributions to this repository should follow our [GitHub Community Guidelines](#).

7. И затем кнопку «Confrim merge», которая появится на том же месте.

8. Ваш форк обновлён до состояния оригинального репозитория.

Основные определения из лекции №1

Данные – оцифрованная информация, которая может содержать ошибки, пустые значения и т.п.

(Математическая) модель – математически сформулированная зависимость между частями данных.

Алгоритм – последовательность действий, позволяющая на основании данных строить модели определенного типа.

Параметры модели – величины, описывающие устройство и работу модели, вычисляются в процессе обучения.

Гиперпараметры модели – величины, описывающие устройство и работу алгоритма построения модели, задаются до начала обучения.

Переобучение модели

Переобученная модель не отличает шум от полезных данных.

Метрики производительности модели на обучающей выборке будут иметь хорошие значения, однако на валидационной и тестовой выборках – плохие.

Методы противодействия:

- Увеличение количества данных.
- Упрощение модели.
- Уменьшение времени обучения.

Недообучение модели

Недообученная модель не может обобщить информацию из обучающей выборки.

Метрики производительности модели будут плохими на всех выборках: обучающей, валидационной, тестовой.

Методы противодействия:

- Усложнение модели.
- Увеличение размерности данных.
- Уменьшение шума в данных.
- Увеличение времени обучения.

Задание

Пояснение

Для сохранения результатов данной работы вам понадобится два файла: doc/docx – для текста и ipynb – для кода. Назовите их одинаково: «Фамилия – задание 2».

Часть 1

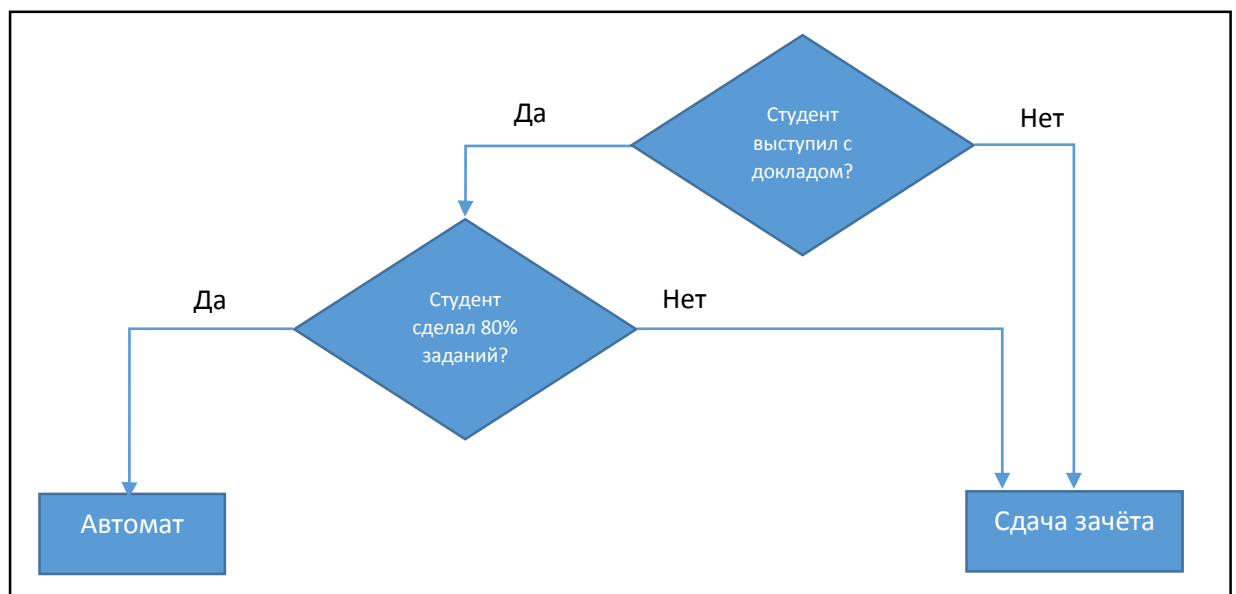
- Обновите свой репозиторий, созданный в практической работе №1, из оригинального репозитория:
<https://github.com/mosalov/Notebook For AI Main>.

Часть 2

- Откройте свой репозиторий в Binder (<https://mybinder.org/>).
- Откройте файл «task2.ipynb».
- Используйте свою фамилию для инициализации генератора случайных чисел, используя код в файле в качестве примера.
- Создайте две выборки, используя приведённый в файле код.
- Напишите свой код в соответствии с инструкциями, сохраните код в ipynb-файле. Необходимые пояснения опишите в своём docx/doc-файле.

Часть 3

- Проанализируйте данную простую модель машинного обучения. На ваш взгляд, что в данной модели является параметрами, а что гиперпараметрами? Опишите в своём docx/doc-файле.



Часть 4

- Проанализируйте графики ошибок, опишите наблюдаемую проблему, предложите меры по её устранению. На графиках синяя кривая – ошибка на обучающей выборке, зелёная кривая – ошибка на валидационной выборке. Опишите в своём docx/doc-файле.

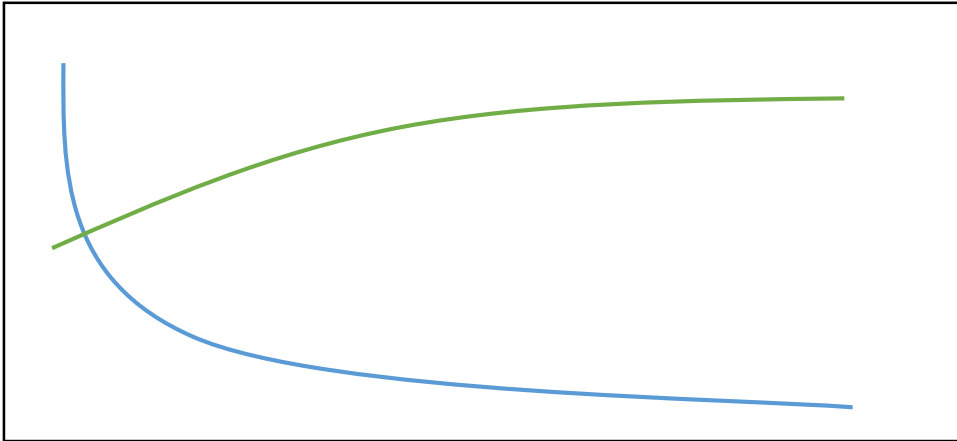


График №1.

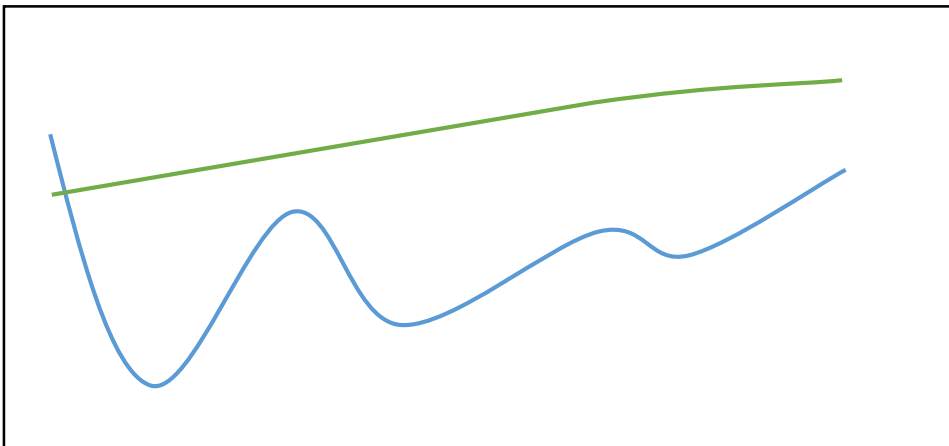


График №2.

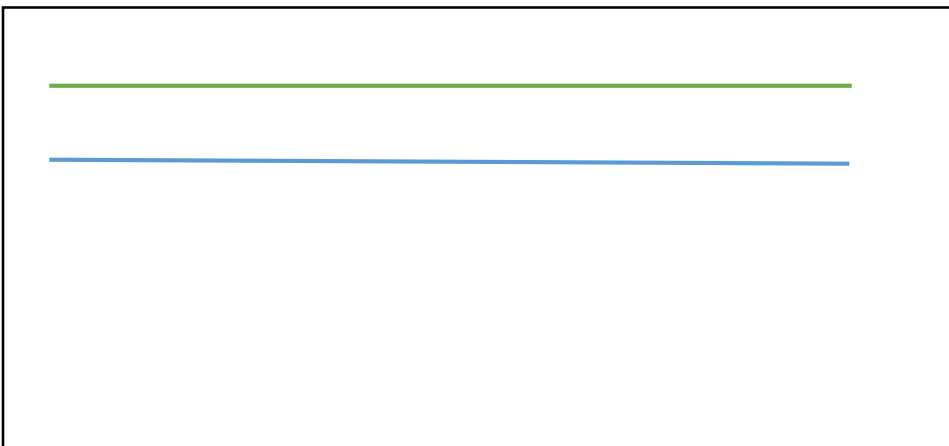


График №3.

Требования к отчету

Оба файла (doc/docs и ipynb) загрузите в свой репозиторий, созданный в практическом задании №1 по пути: «Notebook_For_AI_Main/2020 Осенний семестр/Практическое задание 2/» и сделайте пул-реквест.

Литература

- https://ru.wikipedia.org/wiki/Машинное_обучение
- <https://neerc.ifmo.ru/wiki/index.php?title=Переобучение>
- <https://wiki.loginom.ru/articles/hyperparameters.html>