

## **Задания на олимпиаду «Миссия выполнима» (второй этап). Москва**

### **Задача №1. Машинное обучение.**

Дан файл Excel ([https://palchevsky.ru/uploads/Urovni2\\_1\\_1\\_new\\_1\\_2.xlsx](https://palchevsky.ru/uploads/Urovni2_1_1_new_1_2.xlsx)), в котором хранится больше миллиона значений с данными об уровнях воды по гидрологическим постам и автоматическим станциям Республики Башкортостан. Необходимо реализовать программу с интерактивным консольным меню, т.е. вывод списка действий по цифрам. При этом при нажатии на цифру у нас должно выполняться определенное действие. Каждый пункт меню должен быть отдельной функцией, либо классом, либо методом в классе. В файле Excel необходимо по каждому гидрологическому посту по всем параметрам (столбцам), за исключением столбца «Описание», реализовать восстановление данных любым из двух решений, предлагаемых библиотекой машинного обучения scikit-learn: KNeighborsClassifier или же SimpleImputer.

1. Восстановить пропущенные данные с помощью KNeighborsClassifier или SimpleImputer в файле Excel.

2. Сохранить файл Excel с пропущенными значениями в папку с проектом.

В качестве ответа необходимо прикрепить исходный код файлом, либо текстом.

### **Задача №2. Глубокое обучение.**

На основе всех данных задачи №1 реализовать прогнозирование уровней воды по каждому гидрологическому посту в нормализованном виде. Например, нормализованное значение – это 0,5643.

В качестве ответа необходимо прикрепить исходный код файлом, либо текстом.

### **Задача №3. Постобработка данных.**

Выполнить денормализацию прогнозных значений, полученных в ходе решения задачи №2. На выходе должны получиться «человеческие значения». Например, нормализованное значение – это 0,5643. После денормализации значение будет ~ 789.

В качестве ответа необходимо прикрепить исходный код файлом, либо текстом.

### **Задача №4. Глубокое обучение.**

Известно, что при анализе изображений используются сверточные слои. Основным элементов такого слоя выступает ядро свертки. При этом карта признаков рассчитывается по результатам движения ядра свертки, перекрывающего на каждом шаге небольшую локальную область изображения (см. рис. 1)

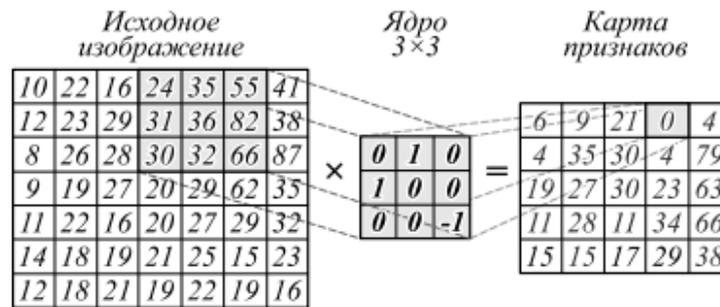


Рис. 1. Свертка

Реализуйте такой слой самостоятельно в виде функции `myConv()` с учетом следующего необходимого функционала:

1) На вход функция получает два обязательных параметра: массив изображения (`image`) и массив сверточного ядра (`kernel`).

2) Дополнительный параметр `padding = True` позволяет сохранить размеры исходного изображения при формировании карты признаков. По умолчанию `padding = False`.

3) Дополнительный параметр `bias = b` добавляет смещение к сумме произведений перекрывающихся пикселей на число `b`. По умолчанию `bias = 0`.

4) Дополнительный параметр `activation = some_function` позволяет рассчитывать значение признака не только как сумма со смещением, но и после обработки нелинейной функцией. Возможные варианты: `activation = 'sigmoid'` (применение функции сигмоиды для расчета признака), `activation = 'relu'` (применение функции ReLU для расчета признака). По умолчанию `activation = False`.

В качестве ответа необходимо прикрепить исходный файл с кодом.

### Задача №5. Машинное обучение.

Представлена статистика по выступлению футбольных команд в виде таблицы 1.

Таблица 1 Статистика футбольных команд

№	Место	Очки	Победы	Разница голов	Забитые голы
1	1	99	32	52	85
2	2	81	26	67	102
3	3	66	18	30	66
4	7	59	15	11	51
5	19	25	6	-44	35
6	20	20	5	-50	27

На основе данной статистики вам необходимо выполнить агломеративную кластеризацию команд по методу ближнего соседа. Для этого выполните следующие этапы:

1. Сократите число признаков до двух, выбрав два наименее коррелирующих признака.
2. Нормализуйте данные.

3. Проведите кластеризацию методом ближнего соседа, начиная рассматривать каждую команду как отдельный кластер и заканчивая объединением всех команд в один кластер.
4. Выделите промежуточный результат, когда у вас было 3 кластера. Интерпретируйте результаты такого разделения.

В качестве ответа необходимо прикрепить файл Excel с корреляционной матрицей, матрицами расстояний на каждом шаге обучения.

## **Задания на олимпиаду «Миссия выполнима» (второй этап). Регионы**

### **Задача №1. Машинное обучение.**

Дан файл Excel ([https://palchevsky.ru/uploads/Urovni2\\_1\\_1\\_new\\_1\\_2.xlsx](https://palchevsky.ru/uploads/Urovni2_1_1_new_1_2.xlsx)), в котором хранится больше миллиона значений с данными об уровнях воды по гидрологическим постам и автоматическим станциям Республики Башкортостан. Необходимо реализовать программу с интерактивным консольным меню, т.е. вывод списка действий по цифрам. При этом при нажатии на цифру у нас должно выполняться определенное действие. Каждый пункт меню должен быть отдельной функцией, либо классом, либо методом в классе. В файле Excel необходимо по каждому гидрологическому посту по всем параметрам (столбцам), за исключением столбца «Описание», реализовать восстановление данных любым из двух решений, предлагаемых библиотекой машинного обучения scikit-learn: KNeighborsClassifier или же SimpleImputer.

3. Восстановить пропущенные данные с помощью KNeighborsClassifier или SimpleImputer в файле Excel.

4. Сохранить файл Excel с пропущенными значениями в папку с проектом.

В качестве ответа необходимо прикрепить исходный код файлом, либо текстом.

### **Задача №2. Глубокое обучение.**

На основе всех данных задачи №1 реализовать прогнозирование уровней воды по каждому гидрологическому посту в нормализованном виде. Например, нормализованное значение – это 0,5643.

В качестве ответа необходимо прикрепить исходный код файлом, либо текстом.

### **Задача №3. Постобработка данных.**

Выполнить денормализацию прогнозных значений, полученных в ходе решения задачи №2. На выходе должны получиться «человеческие значения». Например, нормализованное значение – это 0,5643. После денормализации значение будет ~ 789.

В качестве ответа необходимо прикрепить исходный код файлом, либо текстом.

### **Задача №4. Глубокое обучение.**

Известно, что при анализе изображений используются сверточные слои. Основным элементов такого слоя выступает ядро свертки. При этом карта признаков рассчитывается по результатам движения ядра свертки, перекрывающего на каждом шаге небольшую локальную область изображения (см. рис. 1)

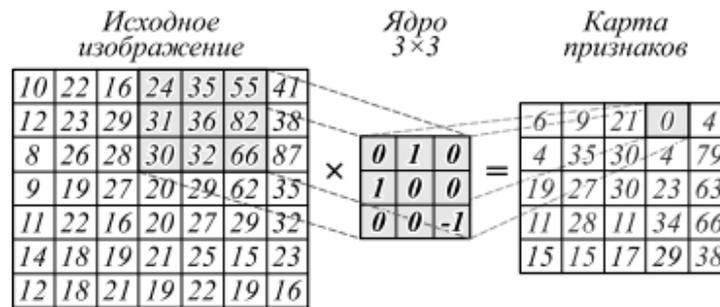


Рис. 1. Свертка

Реализуйте такой слой самостоятельно в виде функции `myConv()` с учетом следующего необходимого функционала:

- 1) На вход функция получает два обязательных параметра: массив изображения (`image`) и массив сверточного ядра (`kernel`).
- 2) Дополнительный параметр `padding = True` позволяет сохранить размеры исходного изображения при формировании карты признаков. По умолчанию `padding = False`.
- 3) Дополнительный параметр `bias = b` добавляет смещение к сумме произведений перекрывающихся пикселей на число `b`. По умолчанию `bias = 0`.
- 4) Дополнительный параметр `activation = some_function` позволяет рассчитывать значение признака не только как сумма со смещением, но и после обработки нелинейной функцией. Возможные варианты: `activation = 'sigmoid'` (применение функции сигмоиды для расчета признака), `activation = 'relu'` (применение функции ReLU для расчета признака). По умолчанию `activation = False`.

В качестве ответа необходимо прикрепить исходный файл с кодом.

### Задача №5. Машинное обучение.

Представлена статистика по выступлению футбольных команд в виде таблицы 1.

Таблица 1 Статистика футбольных команд

№	Место	Очки	Победы	Разница голов	Забитые голы
1	1	99	32	52	85
2	2	81	26	67	102
3	3	66	18	30	66
4	7	59	15	11	51
5	19	25	6	-44	35
6	20	20	5	-50	27

На основе данной статистики вам необходимо выполнить агломеративную кластеризацию команд по методу ближнего соседа. Для этого выполните следующие этапы:

5. Сократите число признаков до двух, выбрав два наименее коррелирующих признака.
6. Нормализуйте данные.

7. Проведите кластеризацию методом ближнего соседа, начиная рассматривать каждую команду как отдельный кластер и заканчивая объединением всех команд в один кластер.
8. Выделите промежуточный результат, когда у вас было 3 кластера. Интерпретируйте результаты такого разделения.

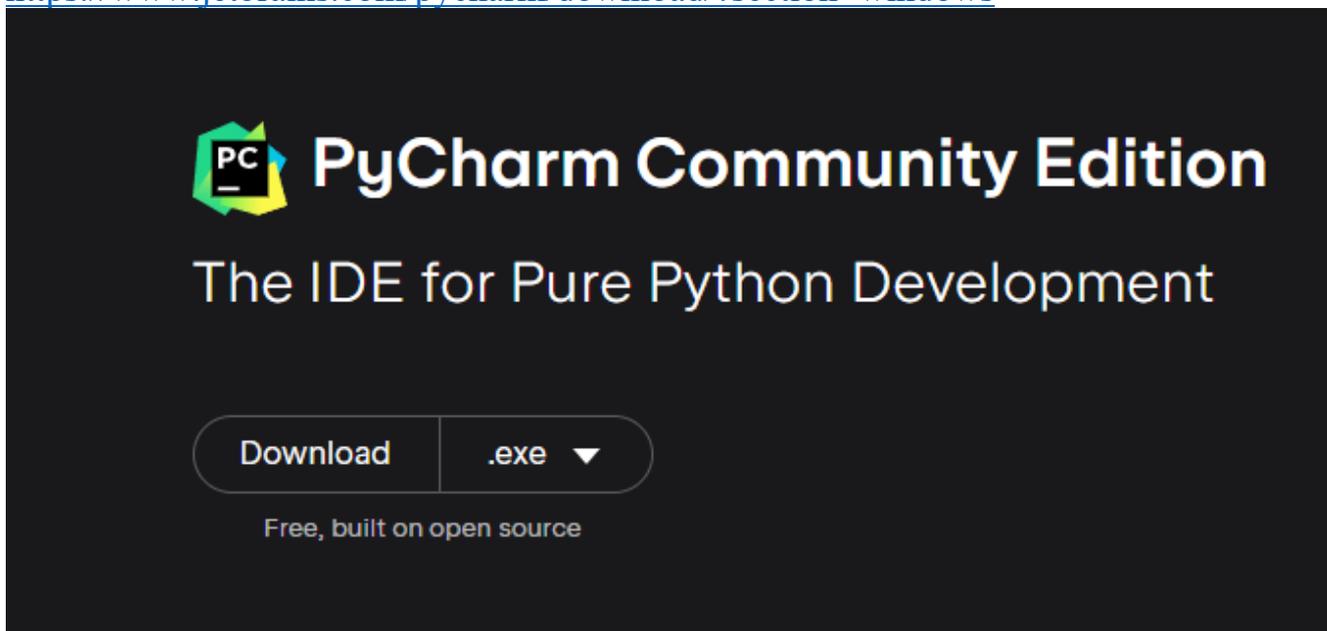
В качестве ответа необходимо прикрепить файл Excel с корреляционной матрицей, матрицами расстояний на каждом шаге обучения.

## Инструкция

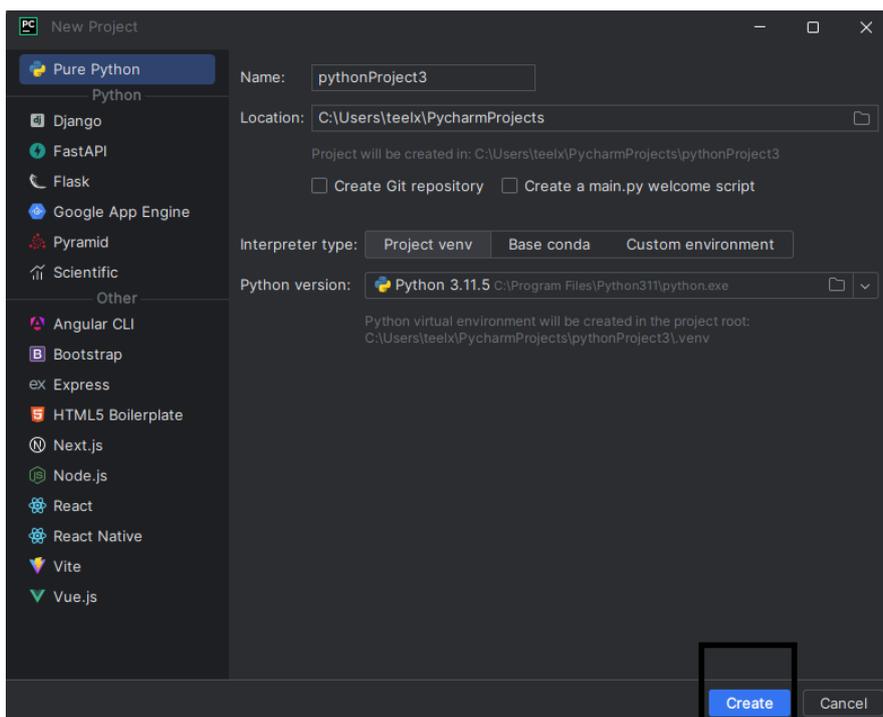
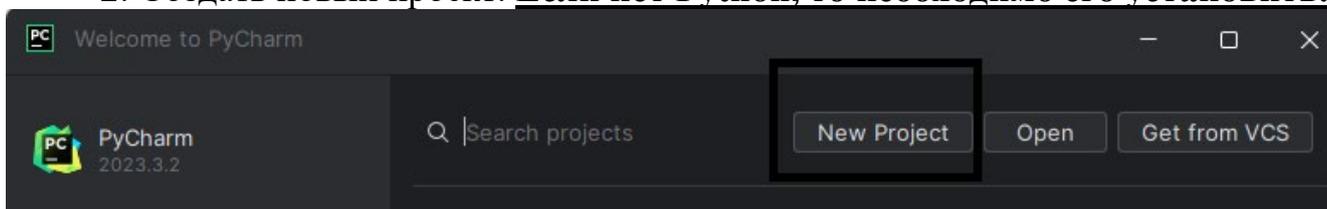
(настоятельно рекомендуется выполнить до начала олимпиады)

1. Скачать PyCharm Community по следующей ссылке:

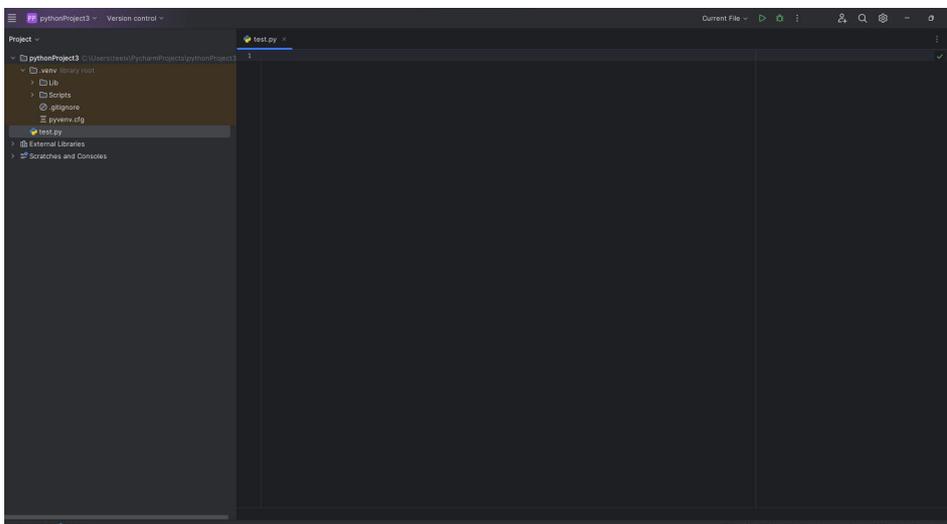
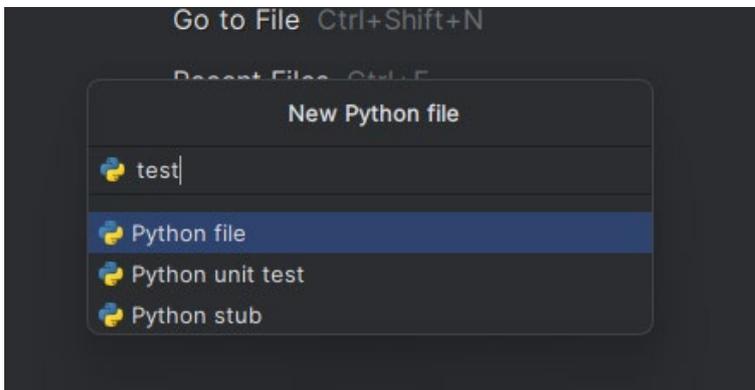
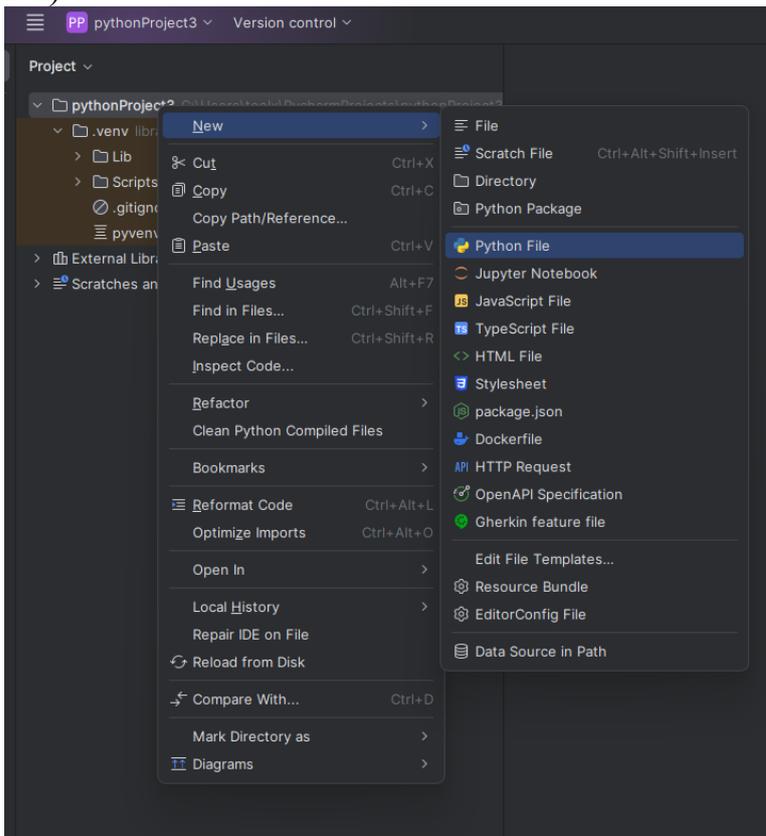
<https://www.jetbrains.com/pycharm/download/?section=windows>



2. Создать новый проект. **Если нет Python, то необходимо его установить.**

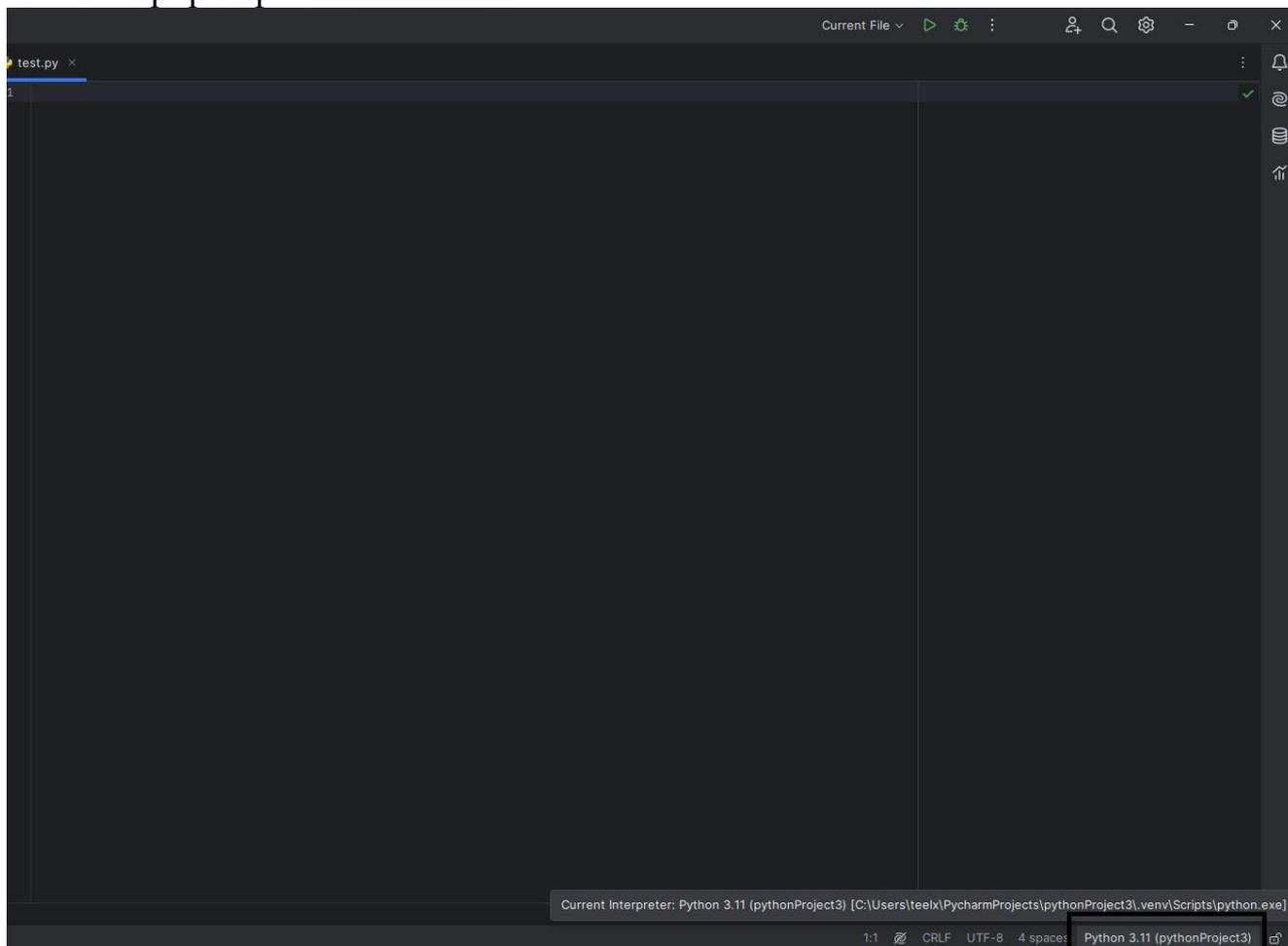


3. После создания проекта необходимо создать файл .py (если он не создан).

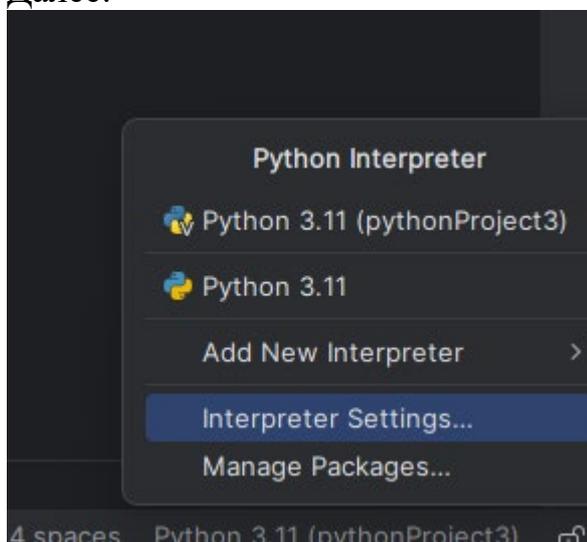


4. **Необходимо заранее установить следующие библиотеки:** scikit-learn, tensorflow, torch, pandas, opencv. Необходим выход в интернет. Перед установкой необходимо обновить pip до последней версии. Все библиотеки корректно работают с Python версии 3.11.5 (проверено).

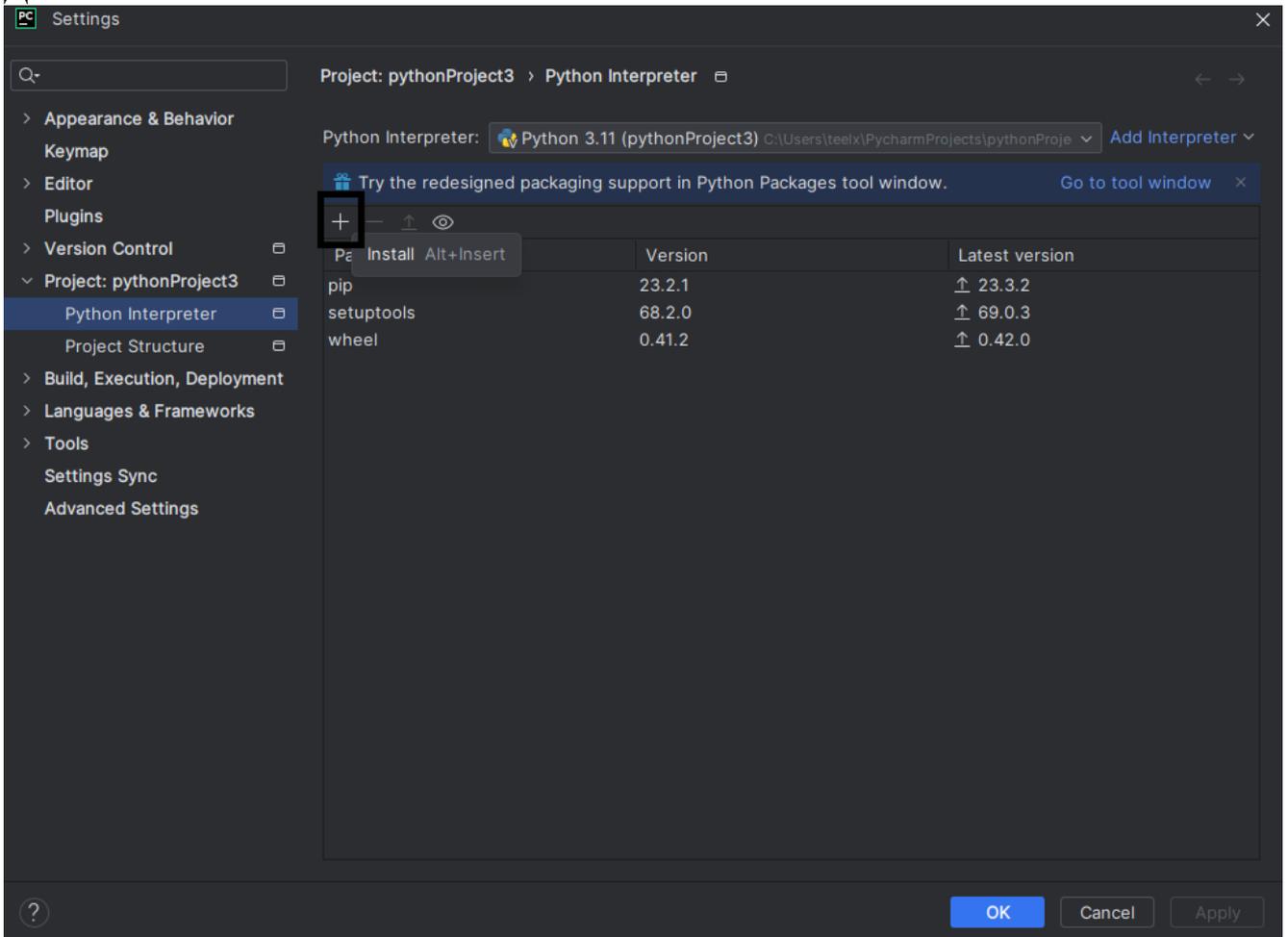
Для установки каждой библиотеки в готовом проекте необходимо выбрать наш интерпретер:



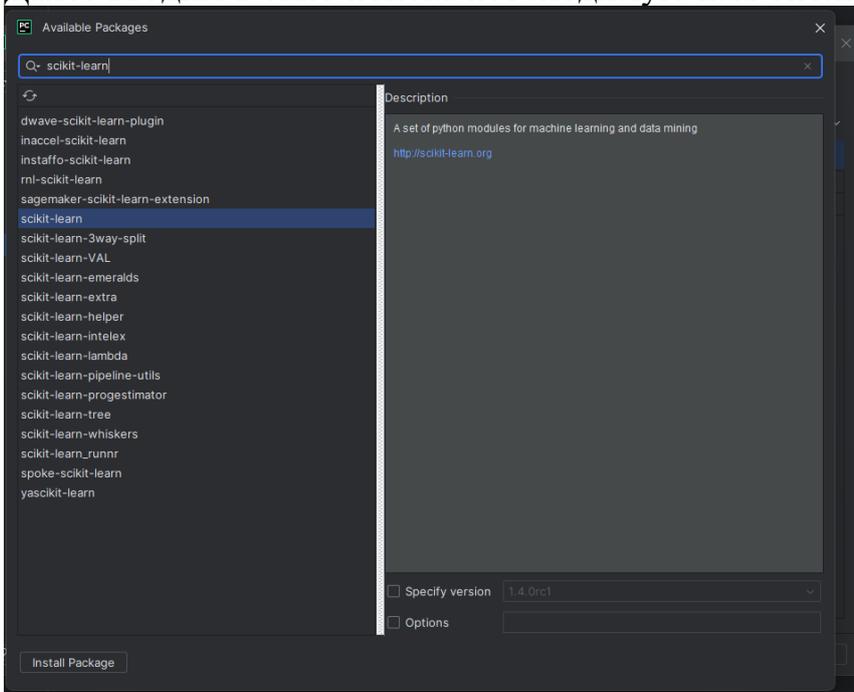
Далее:



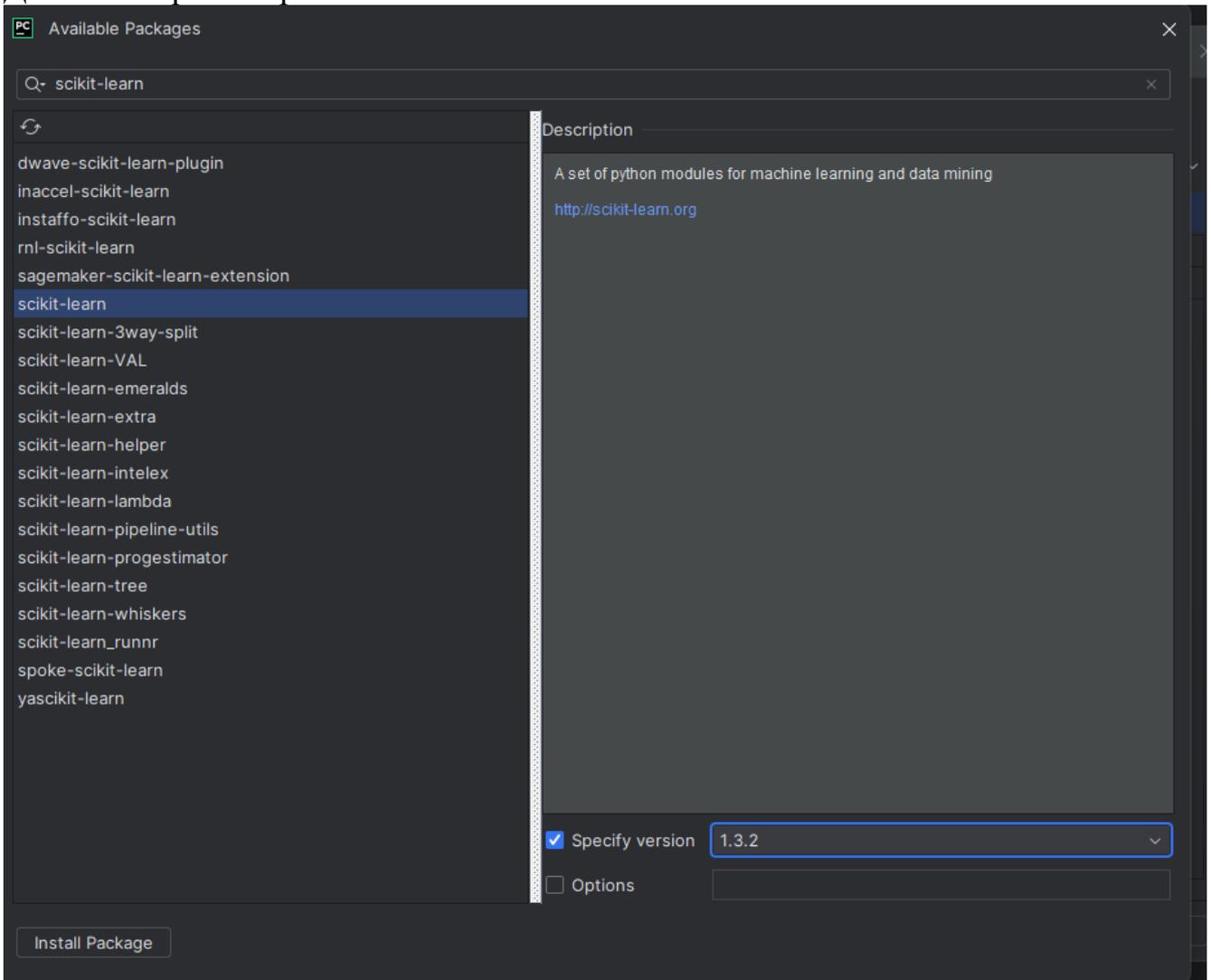
Далее:



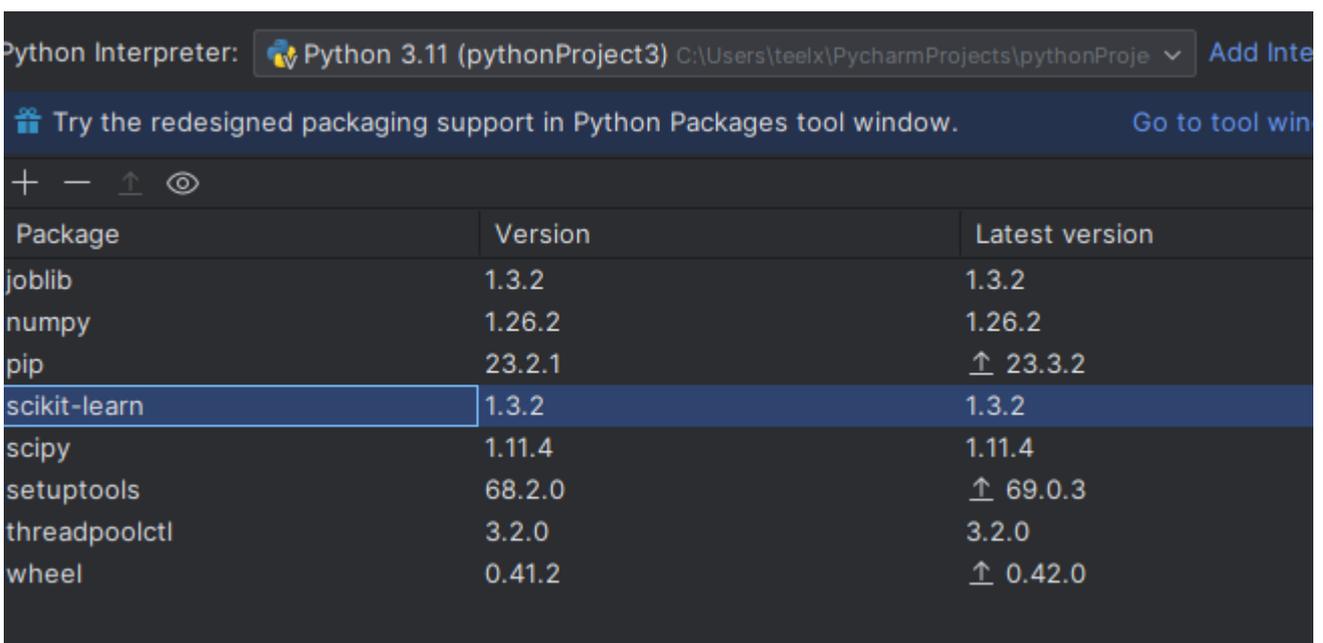
Далее вводим в окно поиска необходимую нам библиотеку:



Далее выбираем версию библиотеки:



И нажимаем «Install Package».



Готово. Библиотека установлена.