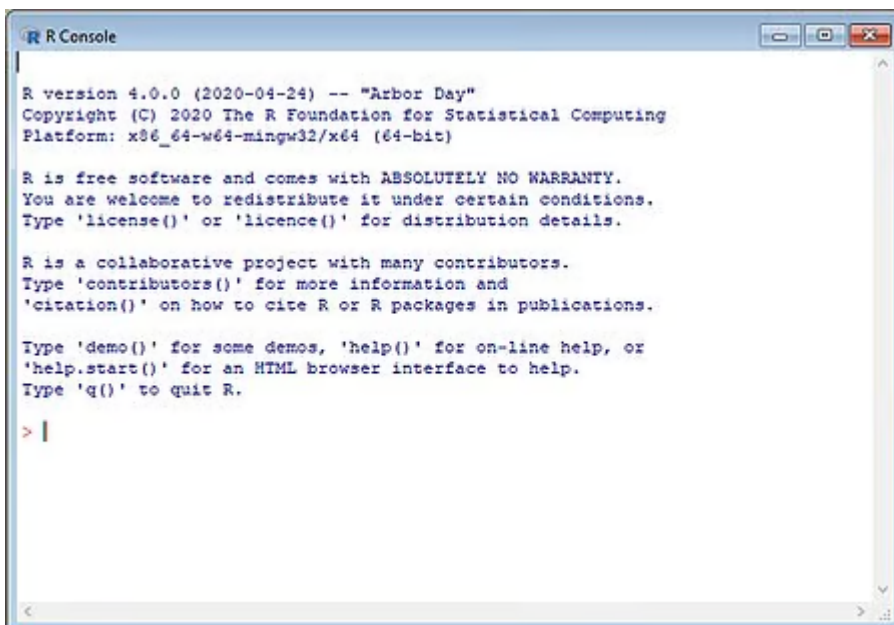


R против Python: 11 ключевых различий

30.06.2023

Когда речь заходит о науке о данных и разработке машинного обучения, выделяются два очень популярных языка программирования – R и Python. Использование любого из этих языков поможет командам специалистов по науке о данных легко исследовать, анализировать, визуализировать, учиться и работать с данными. Однако для выбора языка необходимо знать основные различия между ними. В этой статье рассматриваются и объясняются эти различия.

R: обзор



```
R Console

R version 4.0.0 (2020-04-24) -- "Arbor Day"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |
```

R – это язык программирования для статистических вычислений и создания графиков. Впервые он был выпущен в 1995 году. Роберт Джентльмен и Росс Унака разработали его в Оклендском университете в Новой Зеландии. R был разработан как практическая реализация языка программирования S. Хотя R – это язык программирования, он распространяется как нечто гораздо большее. Он включает в себя среду программирования – R Studio. Эта среда позволяет создавать визуализации, управлять данными

и писать код R с подсветкой синтаксиса и поддержкой отладки. R имеет открытый исходный код и работает на всех основных операционных системах, включая Windows, MacOS и Linux. Чтобы продемонстрировать, как выглядит пример программы, написанной на языке программирования R, вот пример программы, которая генерирует набор случайных значений, следующих нормальному распределению.

```
# Set the seed for reproducibility
set.seed(123)

# Generate 10,000 random numbers following the normal
distribution

random_numbers <- rnorm(10000)
```

Python: Обзор



Python – один из самых популярных языков в мире. Это универсальный язык программирования общего назначения, который поддерживает множество парадигм программирования. Он широко используется для машинного обучения, веб-разработки и написания коротких скриптов для автоматизации. Python прост в изучении и используется в качестве первого языка большинством

людей, изучающих код. Гвидо ван Россум разработал язык Python, который впервые был выпущен в 1991 году. Несмотря на свой возраст, он обновлялся и модернизировался на протяжении многих лет; наиболее заметным является выпуск Python версии 3 – это огромный скачок по сравнению с Python версии 2.

Как и R, Python является открытым исходным кодом и имеет переводчик для всех основных операционных систем. Кроме того, существуют различные браузерные среды для работы с кодом, такие как Jupyter Notebooks. Python является универсальным языком, но имеет пакеты и библиотеки для науки о данных и машинного обучения. Вот пример программы, которая использует пакет NumPy для создания 10000 случайных чисел, соответствующих нормальному распределению.

```
import numpy as np

# Set the seed for reproducibility
np.random.seed(123)

# Generate 10,000 random numbers following the normal
distribution

random_numbers = np.random.randn(10000)
```

Голова к голове: R против Python

Как видите, языки похожи, но имеют некоторые фундаментальные различия. Решая, какой язык изучать или использовать для своего следующего проекта, обратите внимание на некоторые ключевые различия.

Цели

R – это язык программирования, созданный для того, чтобы обеспечить простой способ анализа данных и создания визуализации. Его использование в основном ограничено статистикой, наукой о данных и машинным обучением. С другой стороны, Python – это язык общего назначения, созданный для

того, чтобы быть элегантным и простым. Поэтому он широко используется в области искусственного интеллекта и веб-разработки.

Пользователи

R разработан для использования статистиками. В результате он предлагает простой интерфейс с функциями, необходимыми для работы со статистикой. Python широко используется различными разработчиками всех уровней опыта. Благодаря своей универсальности, он используется не одной профессией, а разными профессиями, такими как инженеры-программисты, аналитики и ученые.

Анализ данных и визуализация

Одним из самых популярных применений Python является наука о данных – анализ и визуализация данных. Однако Python не поддерживает это приложение изначально и вместо этого полагается на библиотеки и пакеты, такие как Pandas и Matplotlib. В отличие от Python, который является универсальным, R создан для использования в науке о данных. Поэтому в R есть функции для анализа данных и построения графиков. Помимо встроенных функций, существуют пакеты, которые можно использовать для дальнейшего анализа.

Производительность

За простоту и абстракции R приходится расплачиваться производительностью. Он не очень быстрый. В программировании быть медленным означает работать на несколько миллисекунд медленнее, чем в других языках. Поэтому медлительность R может быть едва заметна для пользователей. Как и R, Python также предоставляет высокоуровневые абстракции и упрощения, что делает его более гибким, но более медленным. Для повышения производительности программ на Python такие библиотеки, как NumPy и PyTorch, используют C++ или C под капотом и предоставляют привязки в Python. Это позволяет Python работать

быстрее.

Использование в машинном обучении

Помимо визуализации и анализа, R можно использовать для построения моделей машинного обучения с помощью алгоритмов регрессии, классификации и кластеризации. В нем есть такие пакеты, как Caret, lattice, DataExplorer и Dalex. Python – самый популярный язык, используемый в машинном обучении. Он имеет богатую экосистему библиотек, таких как Tensorflow, PyTorch и Jax. Он также наиболее широко поддерживается платформами инфраструктуры машинного обучения.

Кривая обучения

R был разработан, чтобы быть простым. Однако его синтаксис может занять некоторое время, чтобы привыкнуть, и, по моему опыту, он немного сложнее, чем Python. Python – один из самых дружелюбных к новичкам языков программирования, созданный для того, чтобы быть элегантным и простым. Он имеет более мягкую кривую обучения по сравнению с R.

Сообщество

У R большое сообщество – хотя и меньшее, чем у Python. Но оно все равно значительное, поскольку в нем более 10 000 пакетов, которые вы можете использовать. В нем также есть множество образовательных ресурсов, на которых можно учиться, поскольку язык по-прежнему невероятно популярен. Это еще одна область, где Python сияет. У него, вероятно, самое большое сообщество среди всех языков программирования в мире. В нем есть пакеты с открытым исходным кодом практически для всего, включая машинное обучение и обработку данных. Существует множество бесплатных и платных образовательных ресурсов для изучения Python и его многочисленных инструментов, созданных сообществом.

Популярность

Python – один из самых популярных языков программирования в мире, соперничающий с JavaScript и Java. Он также является одним из самых любимых языков благодаря своей простоте и универсальности. R, конечно, не так популярен, как Python, но он все же достаточно популярен, чтобы не быть непонятным языком программирования, которым никто не пользуется. Напротив, его используют крупные технологические компании, и значительное количество вакансий требуют знания R.

Гибкость/ универсальность

R в первую очередь предназначен для статистических вычислений. Поэтому он не используется ни для чего, выходящего за рамки этой области применения. С другой стороны, Python универсален. Он широко используется для различных целей, и наука о данных – лишь одна из многих.

IDE

Python распространяется исключительно как интерпретатор. Выбор IDE зависит от пользователя. Двумя наиболее популярными IDE являются VsCode и PyCharm – их сравнение приведено в этом посте. Кроме того, вы можете использовать Jupyter Notebooks для интерактивного запуска Python в ячейках. Основной IDE, используемой для R, является R Studio, которая доступна для MacOS, Windows и Linux.

Общие приложения

R, статистический язык, обычно используется для машинного обучения, статистики и аналитики, а также визуализации данных. В то время как Python, язык общего назначения, имеет широкий спектр применения, включая машинное обучение, веб-разработку, анализ данных и численные вычисления.

Популярные компании, использующие язык

И у R, и у Python есть крупные технологические пользователи. Среди них Google, Meta, Accenture и Microsoft для R. Среди пользователей Python – Tesla, Google, Meta и Spotify для Python.



Заключительные слова

Python и R – сильные языки, которые, вероятно, будут использоваться в обозримом будущем. Изучение любого из них – это хорошая инвестиция, так как это даст вам навыки работы с данными и получения выводов в этом мире, который все больше управляется данными.