

Google Colab

for R and Python

mLearn

mindConnect.cc

2023년 3월 24일

목차

- 1 시작
 - 첫 대면
 - 런타임과 세션
 - 코랩 문서 작성법

- 2 파일, 구글 드라이브, 깃헙
 - 파이썬에서
 - R에서
 - [파일 - 노트 열기]
 - 깃헙

첫 대면

구글 코랩(*Google Colab*) = 쥬피터 노트북의 온라인 버전, 구글이 무료로 제공

<https://colab.research.google.com>

화면에 ‘Colaboratory에 오신 것을 환영합니다’라는 제목의 코랩 홈 문서가 나타난다. 코랩 웹페이지 문서를 코랩 노트, 쥬피터 노트 혹은 그냥 노트(note)라고 한다. 노트를 아래로 스크롤하면서 이따끔씩 클릭해 보면 화면의 특정 영역이 포커스 됨이 관찰된다. 이 영역들을 셀(cell)이라고 한다.

셀을 하나씩 보기로 한다.

- [Colab 시작 페이지] 나중에
- [Colab이란?] Colab = Colaboratory = Collaboration + Laboratory = 협업 실습실

파이썬(Python)과 R 코드를 작성하고 실행할 수 있음.

- [시작하기] 메모장 = 노트

셀에는 코드 셀과 텍스트 셀이 있다. 기존 텍스트 셀을 수정하고 Shift-Enter. ‘변경사항을 저장할 수 없음’이라는 메시지가 나타날 것임.

코드 셀 = 코드 영역 + 출력 영역

seconds_in_a_day 셀의 출력을 삭제하고 코드를 실행해 보라.
로그인이 필요할 것이다.

로그인, 연결, 런타임, 세션

구글 계정으로 로그인을 해야 하는 이유: ① 코드 쉘 실행, ② 구글 드라이브에 내가 작성한 노트, 혹은 남이 만들어 놓은 노트의 사본 저장
로그인 후에도 **연결**이 되어야 코드 쉘을 실행할 수 있다.

연결을 끊으려면 [런타임 - 런타임 연결 해제 및 삭제]

런타임(runtime)은 구글 코랩이 제공하는 가상 머신.

세션(session)은 HTTP 세션을 뜻하며 하나의 런타임에서 여러 세션을 수용할 수 있다.

예를 들어 [추가 리소스] 쉘에서 [Markdown 가이드]를 클릭하면 새로운 탭에서 코랩 노트가 열린다. 여기서 ‘연결’하면 새로운 세션이 생성되며 이를 [런타임 - 세션 관리]로 가 보면 확인할 수 있다.
[데이터 로드: 드라이브 ..]에 대해서도 같은 작업을 해 보라.

코랩 문서 작성법

첫 노트를 만들기 전에 구글 드라이브 앱을 열어 놓고 내용을 확인한다.

새 노트를 만들려면 코랩 홈문서에서 [새 Colab 메모장 만들기]를 클릭해도 되고 다음의 웹주소(URL)를 방문해도 된다.

<https://colab.research.google.com/#create=true>

구글 드라이브 앱에 가서 **Colab Notebooks**라는 디렉토리가 생성되었음을 확인하고 그 디렉토리를 연다.

코랩에서 'Untitled0.ipynb'를 클릭하여 '첫노트.ipynb'로 바꾼다. 구글 드라이브로 가서 변화가 있는지 본다. 코랩에서 [파일 - 저장]하고 다시 구글 드라이브를 방문.

[Markdown 가이드], 또는 [여기](#)를 보고 헤딩(heading)과 $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ 수식을 포함한 노트를 만들어 본다. 헤딩이 목차에 반영됨을 확인한다.

파이썬에서 파일 다루기

코랩 화면의 맨 왼쪽 영역의 아이콘들을 이용하여 목차 패널을 닫고 파일 패널을 연다. 아까 저장했던 첫노트.ipynb는 보이지 않을 것이다. sample_data 디렉토리의 내용을 본다. 그리고 맨 위에 코드 썬을 하나 만들어 다음의 명령들을 하나씩 실행해 본다.

```
1 !ls
2 !ls -F
3 !ls sample_data -l
4 !pwd
5 !ls /content -F
6 !ls / -F
7 !cd /etc # doesn't work
```

파일 패널 윗 부분에 있는 첫 번째 아이콘을 클릭하여 아무 파일이나 업로드 한다. 이 파일은 런타임이 종료되면 사라진다. 업로드 된 파일은 다운로드, 삭제 등을 할 수 있다.

다른 세션, 예를 들어 [Markdown 가이드]에서 파일 패널을 열어 보면 첫노트.ipynb에서 업로드 했던 파일이 나타나지 않는다. [Markdown 가이드]에서 파일을 업로드 하면 이번에는 첫노트.ipynb에서는 보이지 않는다. 세션마다 파일 저장소가 따로 있다. 이 저장소를 **런타임 저장소** 혹은 **세션 저장소**라고 부르기로 한다. 이 저장소는 세션이 종료되면 사라진다.

항구적으로 파일을 저장하려면 구글 드라이브에 보관하면 된다. 코랩 화면 파일 패널의 3번째 아이콘을 클릭하여 구글 드라이브를 마운트 (mount)한다.

파일 패널에서 파일을 마우스로 드래그하여 위치를 옮길 수 있다.

구글 드라이브와의 연결을 끊으려면 다음의 코드를 실행한다.

```
drive.flush_and_unmount()
```


메뉴바의 [파일 - 다운로드]는 현재 화면에 보이는 코랩 노트를 다운 받을 때 사용한다. 이때 `.ipynb` 혹은 `.py`의 두 형식 중 하나를 선택할 수 있다. 각각의 형식으로 다운로드 해서 노트패드 등 텍스트 에디터로 열어 본다.

첫노트.ipynb 세션을 종료하고 구글 드라이브에서 삭제한다. 그리고 메뉴바의 [파일 - 노트 업로드]에서 다운 받아 놓았던 첫노트.ipynb를 업로드 하면 다시 사용할 수 있다.



코랩에서 파일의 업로드/다운로드에는 앞서 보았듯이 두 가지가 있다. 하나는 노트파일, 즉 `.ipynb` 파일에 대한 것이고 다른 하나는 비노트파일에 대한 것이다. 비노트파일이란 데이터 파일, 혹은 import 해서 사용할 `.py` 파일 등을 뜻한다.

비노트파일을 다루는 방법에 대하여는 [코랩 홈 - 추가 리소스 - 데이터 로드: 드라이브 ..]에 유용한 정보가 있다.

코랩에서 R 사용법

구글 코랩의 R에 대한 지원은 파이썬에 대한 지원보다는 약하다. 특히 구글 드라이브 마운트는 2021년까지는 가능했으나 현재 버전에서는 불가능한 것으로 보인다. googledrive라는 R-패키지가 존재하므로 장차 이 문제는 해결될 수도 있다.

코랩의 디폴트 런타임은 파이썬만 지원하므로 R을 사용하기 위하여는 다음 중 어느 한 방법을 사용해야 한다.

- 1 R로 작성된 노트 파일을 [파일 - 노트 업로드] 한다.
- 2 웹페이지에 있는 R로 작성된 노트 파일의 링크를 클릭하여 연다.
- 3 코랩에서 직접 R 노트를 작성한다.

실은 ①, ②, ③은 파이썬에 대해서도 마찬가지로 적용된다. 이들 중에 ③은 파이썬과 R에 대한 방법이 조금 다르다.

③의 방법을 사용해 보자. 다음의 URL로 접속한다.

<https://colab.research.google.com/#create=true&language=r>

노트 제목을 `first_R.ipynb`로 바꾸고 코드 셀에 `hist(rnorm(100))`을 넣고 실행한 다음 [파일 - 저장]한다. 구글 드라이브 앱에 가서 이 파일이 저장되었음을 확인한다.

런타임을 삭제하고 코랩 탭을 닫고, 구글 드라이브 앱에서 `first_R.ipynb`를 더블 클릭하면 코랩 화면이 열리며 노트 화면이 되살아남을 볼 수 있을 것이다.

②를 사용하는 경우는 남이 만들어 놓은 노트일 수 있다. 이럴 때는 [파일 - 저장]이 되지 않고 [파일 - Drive에 사본 저장]을 해야 할 것이다. 노트 제목에서 '의 사본'은 지우도록 한다.

①에 대해서는 설명이 별로 필요 없다. 노트 파일을 업로드 하고 필요한 작업을 한 다음 [파일 - 저장]하면 그만이다.

source()

자주 사용되는 함수, 또는 상수(constant)를 소스 파일에 저장해 두고 필요할 때마다 불러서 사용하는 것, 즉 import 하는 것은 대부분의 프로그래밍 언어에서 사용하는 기법이다.

```
# util.R

myblue = '#1C86EEFF' # dodgerblue2 (28,134,238)
my_square = function(x) {
  return(x^2)
}
```

위와 같은 파일을 텍스트 에디터로 작성하여 (비노트파일에 대한) 업로드를 한다. 그리고 다음의 노트 파일을 작성하여 실행한다.

```
source('util.R')
plot(c(1,2),c(1,2),col=myblue,pch=17,cex=3)
cat("11^2 =", my_square(11), "\n")
```

파일 - 노트 열기 (Ctrl O)

[파일 - 노트 열기]를 하면 나타나는 팝업 창은 아주 많이 사용하는 화면이다. 사용에 특별히 어려운 점은 없을 것이나 조금 혼동스러운 부분이 있을 수 있다.

- [예]는 코랩 사용법을 설명하는 유용한 문서들의 목록을 보여준다.
- [최근 사용]이 보여주는 것은 노트들의 목록에 불과하므로, 오른쪽 위에 있는 휴지통 아이콘을 클릭하여 삭제해도 노트는 (혹은 노트 파일은) 사라지지 않는다. 삭제 후의 목록을 보려면 일단 창을 닫고 다시 열어야 한다.
- [Google Drive]에서는 삭제되어 휴지통에 있는 노트도 보여주며 클릭해서 열어 볼 수 있다. 구글 드라이브에서 휴지통 비우기를 하면 비로소 [파일 - 노트 열기]에 나타나지 않게 된다.
- [GitHub]은 여러 사람들이 협업할 때 대단히 유용하다. 깃헙의 URL을 넣고 돋보기 아이콘을 클릭해서 사용하면 된다.

깃헙 파일 사용

노트 파일은 이미 경험했듯이 구글 드라이브를 마운트 하지 않고도 아무런 불편 없이 사용할 수 있다. 하지만 비노트 파일은 (런타임에서는 구글 드라이브 마운트가 되지 않으므로) 사용할 때마다 매번 런타임에 업로드 해야 한다.

이 문제는 깃헙에 파일을 저장함으로써 어느 정도 해결된다. (구글 드라이브에서도 여기서 소개하는 깃헙 방법과 유사한 해결책이 있을 것이다.)



[파일 - 노트 열기 - GitHub]에서 jhjeong314를 검색하고 저장소에서 jhjeong314/mytest를 선택한다. 그리고 그 밑에 있는 chisq.ipynb를 클릭한다.

파일 URL 조정

chisq.ipynb의 사본을 구글 드라이브에 저장하고 이름을 chisq_gdrive.ipynb로 바꿔 구별되게 한다. 깃헙 화면(chisq.ipynb)과 구글 드라이브 화면(chisq_gdrive.ipynb)을 각각의 탭에 둔다.

구글 드라이브 화면에서 [도구 - 모든 출력 지우기]를 하고 이어서 [런타임 - 모두 실행]을 하면 [2. 이미지 파일]은 잘 나오지만 [3. R 코드를 source()하여 사용]과 [4. 데이터 파일 read.csv()]에서는 다음과 같은 에러 메시지를 보게 된다.

```
"cannot open file 'util.R': No such file or directory"
```

[2. 이미지 파일] 셀에 있는 이미지 파일의 URL을 보고 이것을 참조하여 뒤의 두 셀의 URL을 알맞게 수정하면 해결된다.