구글CoLab 이용하기

통계학에서 미분이 응용되는 방법

시작하기

Google CoLab이란?

🜔 Colaboratory에 오신 것을 환영합니다

Colaboratory는 설치가 필요 없고 완전히 클라우드에서 실행되는 무료 Jupyter 노트북 환경입니다. Colaboratory를 사용하면 브라우저를 통해 무료로 코드를 작 성 및 실행하고, 분석을 저장 및 공유하며, 강력한 컴퓨팅 리소스를 이용할 수 있습 니다. (구글 개발자 소개 글)

소개 Utube

https://youtu.be/inN8seMm7UI

Tensorflow 프로그래밍 개념

텐서플로우는 데이터 흐름 그래프를 이용하여 계산을 하는 오픈소스 라이브러리이 다. 구글 브레인 팀에서 개발을 하였고 지금 머신러닝 분야에서 활발히 사용

텐서는 여러 차원의 배열의 형태를 한 숫자값의 모음으로 구성되어 있다. 텐서의 rank는 차원의 수이다. 텐서플로우를 사용하면 차원 수가 아주 높은 텐서를 조작할 수 있습니다. 하지만 대부분은 다음과 같은 저차원 텐서 중 하나 이상을 사용하여 작 업하게 됩니다.: *스칼라는 0-d 배열(0번째 텐서)입니다. 예: \'Howdy\' 또는 5 *벡 터는 1-d 배열(1번째 텐서)입니다. 예: [2, 3, 5, 7, 11] 또는 [5] *행렬은 2-d 배열 (2번째 텐서)입니다. 예: [[3.1, 8.2, 5.9][4.3, -2.7, 6.5]]

TensorFlow 프로그래밍 과정 (구글 코랩 개발자 글)

텐서플로우 연산은 텐서를 만들고 없애고 조작합니다. 일반적인 텐서플로우 프로그 램에서 대부분의 코드 행은 연산입니다.

텐서플로우 그래프(또는 산출 그래프나 데이터플로 그래프)는 그래프 데이터 구조 입니다. 많은 텐서플로우 프로그램은 하나의 그래프로 구성되어 있지만, 텐서플로우 프로그램은 여러 그래프를 만들 수도 있습니다. 그래프의 노드는 연산이고; 그래프 의 엣지는 텐서입니다. 텐서는 그래프를 따라 흐르고, 각 노드에서 연산에 의해 조작 됩니다. 한 연산의 출력 텐서는 보통 다음 연산의 입력 텐서가 됩니다. 텐서플로우는 레이지 실행 모델을 구현하는데, 이는 연결된 노드의 필요에 따라 필요할 때만 노드 가 계산된다는 의미입니다.

텐서는 그래프에서 상수 또는 변수로 저장될 수 있습니다. 예상할 수 있듯이 상수는 값이 변하지 않는 텐서를 가지고, 변수는 값이 변할 수 있는 텐서를 가집니다. 하지 만 한 가지 예상하지 못한 점은, 상수와 변수가 그래프에서 또 다른 연산이라는 것입 니다. 상수는 항상 같은 텐서 값을 반환하는 연산이고, 변수는 할당된 텐서를 반환합 니다.

상수를 정의하려면 tf.constant 연산자를 사용하여 그 값을 전달합니다. 예를 들면 다음과 같습니다.

x = tf.constant([5.2])

유사하게 다음과 같은 변수를 만들 수 있습니다.

y = tf.Variable([5])

또는 변수를 먼저 만든 다음, 다음과 같은 값을 할당할 수 있습니다. 참고로 항상 기 본 값을 지정해야 합니다.

y = tf.Variable([0])y = y.assign([5]) 일부 상수 또는 변수를 정의하면 이를 tf.add와 같은 연산과 병합할 수 있습니다. tf.add 연산을 평가할 때 tf.constant 또는 tf.Variable 연산을 호출하여 값을 얻은 다음 그 값의 합으로 새 텐서를 반환합니다. 그래프는 반드시 텐서플로우 세션 내에서 실행되어야 합니다. 세션은 다음을 실행하 는 그래프의 상태를 가집니다. with tf.Session() as sess: initialization = tf.global_variables_initializer() print(y.eval()) tf.Variable을 사용할 때 위에서와 같이 세션 시작 시 tf.global_variables_initializer를 호출하여 명시적으로 초기화해야 합니다.



설치하기

준비 사항: 구글 계정 만들기, 크롬 Chrome 설치하기

크롬을 시작하면 우측 상단부분 🗰 누르면 구글 엡들이 리스트 되고 🍊 (구글 드라이브)가 나타난다. 이를 클릭하면 구글 드라이브로 나타난다.

Sand Luto... Course Mana... Oniversity Pol... Previous P ● ● ● M El × + ← → C ① ③ Google Old 검색하거나 URL을 입력하세요. ☆ ○ ○ ② Google 프레젠테이션 ▲ My Drive - Google... ③ 한남대학교 통계학과... ③ R-bloggers | R ne... > ④ ●<







R 사용하기

개념

Colaboratory는 주피터 노트북 환경을 재현하여 Python 사용이 가능하고 또한 R의 사용도 가능하다.

Activate R : CoLAB 환경에서 R 활성화

activate R magicimport rpy2%load_ext rpy2.ipython

[2] # activate R magic import rpy2 %load_ext rpy2.ipython

모든 R 코드는 %%R로 시작한다.

%%R secret<-sample(c(1:10),1) print(secret)

[5] %%R
secret<-sample(c(1:10),1)
print(secret)</pre>

[→ [1] 5

이미 전 코드 창에서 %%R로 실행했더라도 매번 코드 창에서 %%R 사용하 지 않으면 R 코드로 인식하지 못한다.

- x <- c("apples", "oranges", "bananas", "strawberries")
 for (i in 1:4) { print(x[i])}</pre>
- File <u>"<ipython-input-9-537c0e380967>"</u>, line 2
 for (i in 1:4) { print(x[i])}

SyntaxError: invalid syntax

SEARCH STACK OVERFLOW

코드 창마다 %%R로 시작해야 한다.

%%R

x <- c("apples", "oranges", "bananas", "strawberries")
for (i in 1:4) { print(x[i])}</pre>

[8] %%R

x <- c("apples", "oranges", "bananas", "strawberries")
for (i in 1:4) { print(x[i])}</pre>

- - [1] "oranges"
 - [1] "bananas"
 - [1] "strawberries"

구글드라이브 파일 사용

외부 파일 불러오기

http://wolfpack.hnu.ac.kr/Stat_Notes/adv_stat/LinearModel/data/SMSA.csv

A	В	с	D	E	F	G
city	Mortality	JanTemp	JulyTemp	RelHum	Rain	Education
Akron, OH	921.87	27	71	59	36	11.4
Albany-Schenectady-	997.87	23	72	57	35	11
Allentown, Bethlehem	962.35	29	74	54	44	9.8
Atlanta, GA	982.29	45	79	56	47	11.1
Baltimore, MD	1071.29	35	77	55	43	9.6

파이썬 pandas 이용

import pandas as pd url='<u>http://203.247.53.31/Stat_Notes/adv_stat/LinearModel/data/SMSA.csv'</u> smsa=pd.read_csv(url) smsa.head(3)

import pandas as pd

url='http://203.247.53.31/Stat_Notes/adv_stat/LinearModel/data/SMSA.csv'
smsa=pd.read_csv(url)
smsa.head(3)

C→		city	Mortality	JanTemp	JulyTemp	RelHum	Rain	Education
	0	Akron, OH	921.87	27	71	59	36	11.4
	1	Albany- Schenectady- Troy, NY	997.87	23	72	57	35	11.0
	2	Allentown, Bethlehem.	962.35	29	74	54	44	9.8

R 이용

%%R smsa<-read.csv('<u>http://203.247.53.31/Stat_Notes/adv_stat/LinearModel/</u> <u>data/SMSA.csv'</u>) head(smsa)

O ११R smsa<-read.csv('http://203.247.53.31/Stat_Notes/adv_stat/LinearModel/data/SMSA.csv')</pre> head(smsa) C→ city Mortality JanTemp JulyTemp RelHum Rain Education 1 Akron, OH 921.87 27 71 59 36 11.4 72 57 35 11.0 2 Albany-Schenectady-Troy, NY 997.87 23 3 Allentown, Bethlehem, PA-NJ 962.35 29 74 54 9.8 44 45 982.29 79 56 47 4 Atlanta, GA 11.1

물론 먼저 아래와 같이 R을 사용 설정해야 한다.



activate R magic import rpy2 %load ext rpy2.ipython



		Q Search Drive				
╋	New	My Drive 👻				
 • 	My Drive Colab Notebooks	CO CoLAB_시작하기.ipynb You edited today Name				
20 20	Computers Shared with me					
() ^	Recent					
¥	Starred 구글 드라	·이브 마운팅(연결하기)				
om g rive.n	oogle.colab import drive nount('/content/drive')	;				
rom o rive	google.colab import dra .mount <u>('/content/drive</u>	ive '.)				
o to	this URL in a browse	er: https://accounts.google.com/o/oa				



drive

My Drive

sample_data

SMSA.csv

sample_data

SMSA_COLAB.csv

폴더 위치 파악하기

사용할 데이터 파일이 있는 곳으로 cd 커맨드 사용하여 폴더 변경한다

/content/drive/My Drive

폴더 변경 결과 /content/drive/My Drive 폴더에 데이터가 있음을 확인

'Colab Notebooks' C→ SMSA COLAB.csv SMSA.csv

구글 드라이브 파일 파이썬, R로 불러오기 파일 이름을 폴더 이름과 함께 적어주면 된다. 파이썬 import pandas as pd smsa=pd.read csv('/content/drive/My Drive/SMSA.csv') smsa.head(3)city Mortality JanTemp JulyTemp RelHum Ra Akron, OH 921.87 27 71 59 0 Albany-1 Schenectady-997.87 23 57 72 R 활용 %%R smsa<-read.csv('/content/drive/My Drive/SMSA.csv')</pre> head(smsa)

1

C→				C	ity	Mortal	ity .	JanTemp	J	ulyTemp
_	1		Akro	on,	OH	921	.87	27	7	71
	2	Albany-Sche	enectady-Tro	ру,	NY	997	.87	23	3	72
	3	Allentown,	Bethlehem,	PA-	-NJ	962	.35	29)	74
	4		Atlant	:a,	GA	982	.29	45	5	79
	5		Baltimor	e,	MD	1071	.29	35	5	77
	6		Birmingha	am,	AL	1030	.38	45	5	80
		PopDensity	NonWhite	WC	рор	.house	inco	ome HCI	Pot	NOxPot
	1		• • • •							

	파일 구글 드라이	브에 파이썬(R) 만든 파일 저장하기					
[30]] from google.colab import drive drive.mount('/content/drive')						
C→	C→ Drive already mounted at /conte						
[34]	<pre>34] import pandas as pd smsa=pd.read_csv('/content/drive/My Drive/SMSA.csv') smsa.shape</pre>						
C→	(60, 15)						
[35]	5] smsa_colab=smsa[smsa['Education']>=12] smsa_colab.shape						
C→	(12, 15)						
My Drive 폴더에 파이썬에서 만든 파일 데이터를 저장하면 된다.							
[36] smsa_colab.to_csv('/content/drive/My Drive/SMSA코랩.csv')							
di	rive	pandas 데이터프레임					
 My Drive 		smsa_colab					

csv 데이터로 구글 드라이브 My

Colab Notebooks

SMSA_COLAB.csv

SMSA.csv

■ SMSA코랩.csv

10

CoLab 장점

구글 드라이브 계정만 있으면 어디서나 사용 가능

내 구글 드라이브 계정에 CoLab을 한 번만 설치하면 인터넷이 되는 어디서나 사용 가능하다.

파이썬, R 모두 사용 가능하다.

앞에서 설명한 것처럼 ryp2 모듈 불러들이면 된다. 물론 R 코드 사용하는 경우에는 %%R로 시작해야 한다는 것을 기억하기 바란다.

activate R magic import rpy2 %load_ext rpy2.ipython

• [물론 동일 프로그램에서도 R에서 만든 데이터, 오브젝트는 파이썬에서 사용하 지 못한다.]

구글 colab에서 만들어진 요약 통계량 데이터는 엑셀 파일로 본인의 구글 드 라이브에 저장 가능하다.

필요한 모듈은 어떻게 설치하나?

대부분의 모듈은 구글 코랩을 시작하면 가상 디렉터리에 존재한다. 대부분이지 모두 는 아니다.

주가 데이터 관련 모듈 'FinanceDataReader' 설치가 필요하다는 오류

from statsmodels.tsa.seasonal import seasonal_decompose import FinanceDataReader as fdr kospi = fdr.DataReader('KS11', '2018') kospi.head(3)

4 kospi.head(3)

ModuleNotFoundError: No module named 'FinanceDataReader'

!pip 이용하여 관련 모듈을 설치한다. -> 위의 프로그램 정상 실행됨

!pip install -U finance-datareader

Collecting finance-datareader

Downloading https://files.pythonhosted.org/packa Collecting requests-file (from finance-datareader) Downloading https://files.pythonhosted.org/packa Requirement already satisfied, skipping upgrade: p

^{[27] %%}R sample(c(1:10),1)