AI基礎セミナー

第 1 回 カリキュラムと環境作成

改訂履歴

日付	担当者	内容
2021/05/08	M. Takeda	Git 公開
2021/06/14	M. Takeda	「(3) 開発環境の作成(Anaconda)」を「2021/06」時点のもので更新

目次

- (1) はじめに
- (2) カリキュラム
 - (2.1) セミナーの目標
 - (2.2) セミナーのカリキュラム
 - (2.3) 機械学習ライブラリ
 - (2.4) 言語
 - (2.5) OS
 - (2.6) 教材
- (3) 開発環境の作成 (Anaconda)
 - (3.1) インストール対象物
 - (3.2) インストール
- (4) 開発環境の作成 (Anaconda以外)
 - (4.1) Google Colaboratory
 - (4.2) Amazon SageMaker
- (参考) pip と conda

(参考.1) conda と pip について

(1) はじめに

- ・第1回では、カリキュラムと環境作成について解説します。
- ・第2回以降に掲載した Pythonコードは全て JupyterNotebook で動作確認済です。
- ・Pythonコードのローカルの開発環境として、「Anaconda」を紹介します。 サーバ上で公開している開発環境として、Google が提供している「Google Colaboratory」を紹介します。 後者は、参照するライブラリのバージョンの齟齬などを気にすることなく開発でき、 実装の検証がお手軽にできます。
- ・A I サービスは様々な企業がクラウドで提供しています。
 Amazon が提供しているAWS(Amazon Web Services) もその一つで、本セミナーでは少しだけ触れます。
 こうした環境では自作のA I サービスを開発・公開するだけでなく、
 出来合いのA I サービスも利用可能で、それほど作りこみをしなくても利用できるようになっています。

(2) カリキュラム

(2.1) セミナーの目標

- 本セミナーでは、以下の2点を目標とします:
 - (a) 機械学習の基礎的な理論が理解できるようになる。
 - (b) 機械学習の実装ができるようになる。

(2.2) セミナーのカリキュラム

- ・本セミナーでは、概ね、以下のような流れで進めます:
 - (a) 環境作成
 - (b) 機械学習の全貌
 - (c) Python の言語仕様
 - (d) 数学の基礎と実装
 - (e) 機械学習の理論と実装

(2.3) 機械学習ライブラリ

・機械学習ライブラリとして、以下の例のように様々なものが提供されています。

・Theano(テアァノ) : カナダ モントリオール大学

・Pylearn2 (パイラーンツー) : カナダ モントリオール大学 (「Theano」ベース)

・Caffe (カッフェ) : アメリカ UCLB

・Chainer (チェイナー) : 日本 (株) Preferred Networks (2015 OSSとして公開、2019開発終了)

TensorFlow (テンリルフロー) : アメリカ Google

・PyTorch (パイトーチ) : アメリカ Facebook (Chainerを参考に開発)

・CNTK : アメリカ Microsoft (CNTK は Microsoft Cognitive Toolkit)

本セミナーでは、

様々な機械学習ライブラリのうち、世界的にも広く支持されている TensorFlow を扱います。 これは、以下のように対応環境が多様で、様々な論文やサポートがあります。

• 対応OS : Windows、Linux, MacOS

・対応プログラミング言語 : C, C++, Python, Java, Go

・対応ハードウェア : CPU, GPU(Graphics Processing Unit),

TPU(Tensor processing unit)

(2.4) 言語

・TensorFlow は様々な言語に対応していますが、Python のAPIが最も完成されているため、 使用言語は Python とします。使用するライブラリの関係からバージョンは 3.7 以上とします。 (TensorFlow に限らず、多くの機械学習の本が Python で解説しています。)

(2.5) OS

・TensorFlow は様々なOSに対応していますが、本セミナーでは、 パソコンOSとして普及している Windows を使用OSとします。

(2.6) 教材

・本セミナーでは、主に以下の書籍を教材として使用します。

(教材1)

「Pythonで動かして学ぶ!あたらしい機械学習の教科書」 (2018年01月 翔泳社 伊藤真著)

この書籍では、AI分野で必須となる用語や手法が、わかりやすく紹介されています。 機械学習について、Pythonでの実装/関連する数学の基本/機械学習のアルゴリズム /TensorFlowを用いた実装、などについて、初学者向けに記された良書です。 まずは、この書籍で、AI分野に取り組む為の基礎を築くのが良いかと思います。 添付の Python ファイルで、実際に実装を見ながら動作確認できる、という利点があります。

(教材2)

「現場で使える! TensorFlow開発入門 Kerasによる深層学習モデル構築手法」 (2018年04月 翔泳社 太田満久、須藤広大、黒澤匠雅、小田大輔 共著)

この本は、TensorFlowの導入から、高レベルAPIであるKerasを利用した実践的な深層学習モデルまで解説した、エンジニア向けの入門書です。 主に画像関係のソリューションを扱っていて、 数式は殆ど使わない TensorFlow 初学者向の書籍でもあります。 この書籍で、機械学習のモデルがどのようなものかのイメージを築けるかと思います。

添付の Python ファイルで、実際に実装を見ながら動作確認できる、という利点があります。

(教材3)

「ITエンジニアのための強化学習理論入門」(2020年07月 技術評論社 中井悦司) 「現場で使える! Python 深層強化学習入門」(2019年07月 翔泳社 伊藤多一、他) 「強化学習」(2018年11月 第1版第11刷 森北出版 Richard S. Sutton and Andrew G. Barto)

これらの本は、強化学習で参考になります。

「ITエンジニアのための強化学習理論入門」は、実装を交えながらの解説例が多くて理解しやすいです。「現場で使える! Python 深層強化学習入門」は、理論と実装についての参考になります。「Sutton」本は強化学習入門者のバイブルと呼ばれていて、本資料の表記はこれに合わせています。

「Sutton」本の第2版は、以下のサイトでpdfファイルをダウンロードできます(英語):

ザイト⇒ https://web.stanford.edu/class/psych209/Readings/SuttonBartoIPRLBook2ndEd.pdf

【参照URL】

TensorFlow⇒ https://www.TensorFlow.org/
Chainer開発終了⇒ https://xtech.nikkei.com/atcl/nxt/column/18/00001/03341/

(3) 開発環境の作成 (Anaconda)

・Pythonコードのローカルの開発環境として「Anaconda」を紹介します。

(3.1) インストール対象物

(3.1.1) Python (パイソン)

・TensorFlow は様々な言語に対応していますが、Python のAPIが最も完成されているため、Python から使用することが一般的です。そのため Python の開発環境を構築します。 バージョンは、2021年6月時点での最新公開バージョンである 3.8 とします。

(3.1.2) Anaconda (アナコンタ・)

- ・Anaconda は、Anaconda 社(旧 Continuum Analytics 社)によって提供されている、 Python のディストリビューションです。
- ・Anaconda は、Python 本体に加え、科学技術、数学、エンジニアリング、データ分析など、よく利用される Python パッケージを一括でインストール可能にしたパッケージです。
- ・なお、Anaconda は商用目的にも利用可能です。

(3.1.3) Jupyter Notebook (ジュパイターノートプック)

- ・Python のエディッターとして「Jupyter Notebook」をインストールします。
- ・インテリセンス機能はないので少々不便ですが、実行確認が容易にできます。

(3.1.4) TensorFlow (テンソルフロー)

• TensorFlow は Google が無償で提供している機械学習ライブラリです。

(3.1.5) Keras (ケラス)

Kerasは、Pythonで書かれた、TensorFlow/CNTK/Theano上で実行可能な高水準のニューラルネットワークライブラリです。

元々は TensorFlow とは独立していましたが、2017年初めに TensorFlow と統合されました。 TensorFlow と統合された系統と、複数のバックエンドを選べるAPIという別の系統があります。

・TensorFlow を用いる開発者から見ると、Keras は TensorFlow の使い勝手を良くしたラッパーライブラリ でもあり、TensorFlow を統合した API仕様 となっています。

(3.1.6) その他のパッケージ

・上記以外に、科学技術計算用やグラフ表示用などの各種パッケージを、必要に応じてインストールします。 以下はその一例です。

パッケージ	説明
h5py	HDF5形式のファイルを取り扱うライブラリで、Kerasのモデルを保存する際に利用する
matplotlib	標準的な可視化ライブラリで、学習結果の可視化などで利用している
opencv	広く使われている画像処理ライブラリ
pandas	データ解析ライブラリ
pillow	標準的な画像処理ライブラリで、Kerasが内部的に利用している
scipy	科学技術計算ライブラリ

【参照URL·出典】

「現場で使える!TensorFlow開発入門 Kerasによる深層学習モデル構築手法」(2018年04月 翔泳社)

「https://keras.io/ja/」

「https://pythondatascience.plavox.info/pythonのインストール/pythonのインストール-windows」

Fhttps://deepblue-ts.co.jp/python/conda-comand-basic/_

https://anaconda.cloud/tutorials/getting-started-with-anaconda-individual-edition?source=download

(3.2) インストール

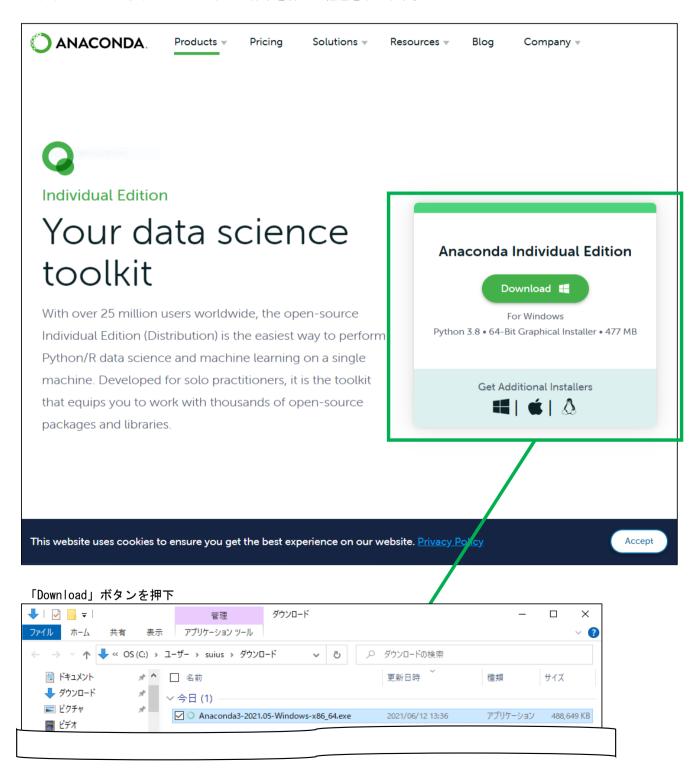
(※以下では「Anaconda Navigator」から仮想環境を指定の上でインストールする、という方法で説明しますが、「Anaconda Navigator」を使用しないコマンドラインを用いる方法等もあります。)

(3.2.1) Anaconda をインストール

・Anaconda は、以下のサイトからダウンロードしてインストールします。 インストールは、特に問題なければデフォルト値のままでよいでしょう。

「https://www.anaconda.com/products/individual」

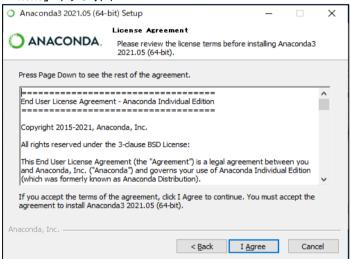
・以下、「https://www.python.jp/install/anaconda/windows/install.html」を参考にして、 ダウンロードからインストールまでの作業を行った経過を示します。



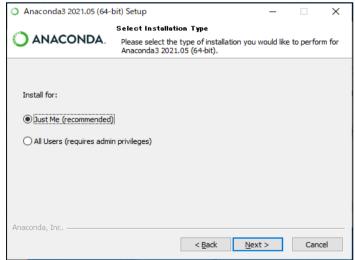
ダウンロードしたインストーラ「Anaconda3-2021.05-Windows-x86_64.exe」を実行。



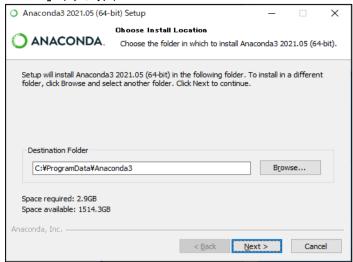
「Next」ボタン押下



「I Agree」ボタン押下



「Next」ボタン押下



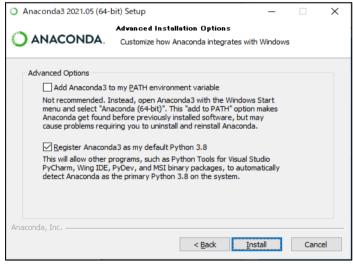
筆者の環境では、

「Destination Folder」として、デフォルトの「C:\ProgramingData\Anaconda3」は使用済だったので「Destination Folder」として、デフォルトの「C:\ProgramingData\Anaconda3_suiu」で指定しました。

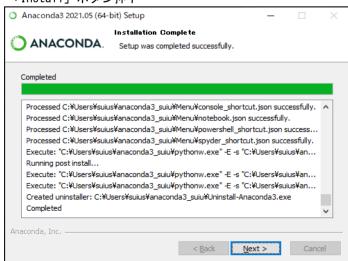
尚、インストール先のフォルダーパスは、英数字のみで構成されているものを指定します。

(日本語を含むフォルダーにはインストールしないように、注意してください!! そうしないと、例えば TensorBoard による可視化がうまく機能しません。)

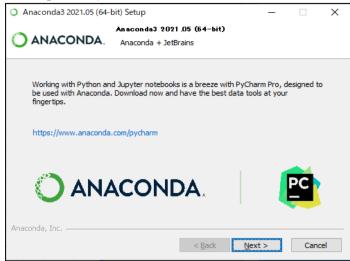
そのうえで「Next」ボタン押下



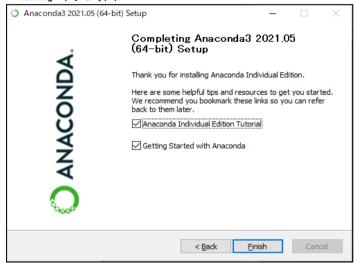
「Install」ボタン押下



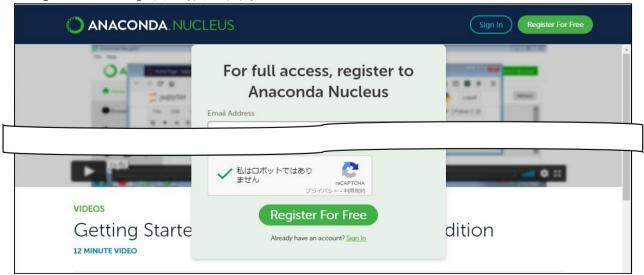
「Next」ボタン押下



「Next」ボタン押下



「Finish」ボタン押下でブラウザ上に「ANACONDA NUCLEUS」の画面が出てくるので必須項目を入力して「Register For Free」ボタン押下します。

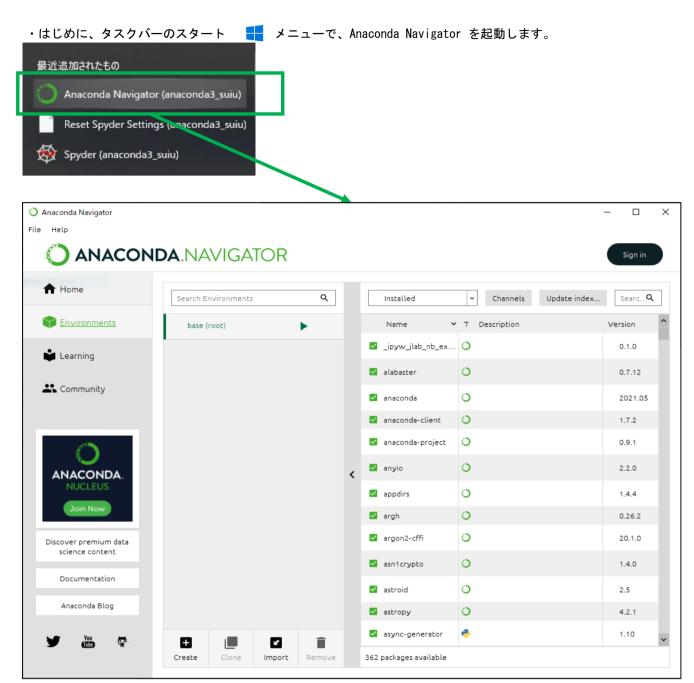


以上で、インストール作業は終了です。

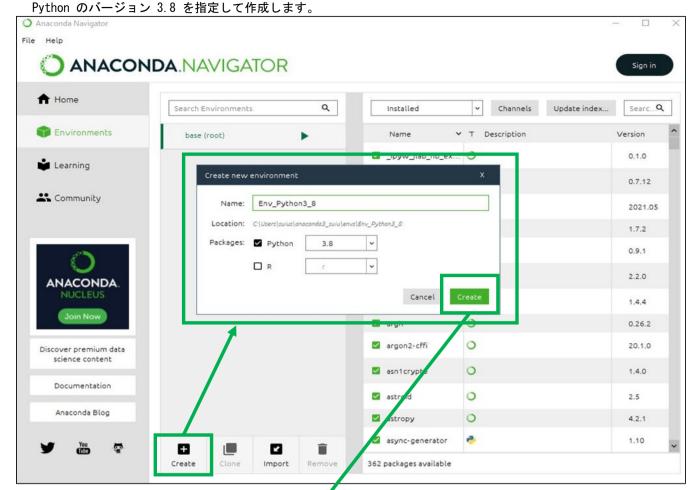
「スタート」メニューに登録した「Anaconda Navigator」が表示されているのが確認できます。

(3.2.2) Python のバージョンを指定して Anaconda 仮想環境を作成

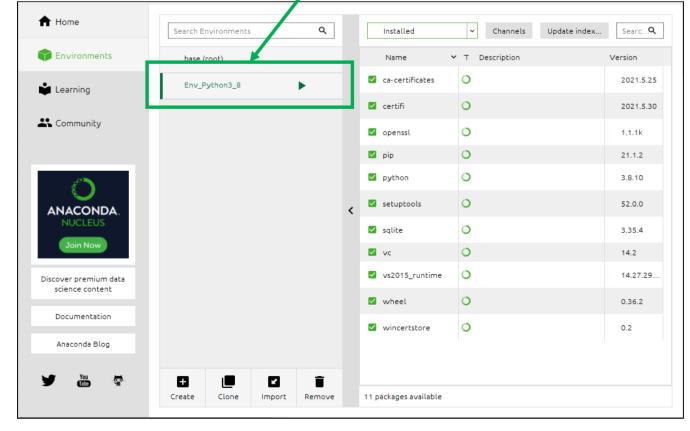
・Anaconda をインストールすると、「Anaconda Navigator」もインストールされます。
「Anaconda Navigator」から、バージョンを指定した開発環境(仮想環境)が、幾つでも作成できます。
その手順を以下に述べます。



- ・「Anaconda Navigator」から、「Environments -> Create」で仮想環境作成用のダイアログを表示し、 Python のバージョンを指定した上で、仮想環境を作成します。
- ・以下の例では、「Env_Python3_8」という名前の仮想環境を、

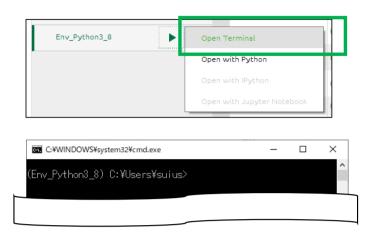


・上記例で、仮想環境「Env_Python3_8」を作成した直後の、インストール済みライブラリは 以下のようになっています。



(3.2.3) Anaconda 仮想環境でのインストール

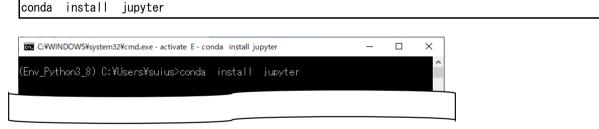
- 「Anaconda Navigator」から仮想環境を指定し、様々な環境をインストールします。 以下の手順で行います。
- (1st) 仮想環境「Env_Python3_8」で「Open Terminal」を指定し、コマンドプロンプトを立ち上げます。 これ以降のインストール作業は、この仮想環境のコマンドプロンプトで行います。



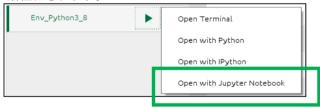
(3. 2. 4) Jupyter Notebook (ジュパイターノートプック)

・仮想環境のコマンドプロンプトから次のコマンドで、

Jupyter Notebook をインストールします。



インストールにより、仮想環境「Env_Python3_8」で「Open With Jupyter Notebook」のメニューが活性化されます。



(3. 2. 5) TensorFlow (テンソルフロー)

pip install

仮想環境のコマンドプロンプトから次のコマンドで、

tensorflow

TensorFlow をインストールします。

さらにオプション指定で、tensorflow のバージョンを指定をすることもできます。

CFWINDOWSFystem32Kcmd.exe-activate E-conda install jupyter

(Env_Python3_8) C:\footnote{\text{Users}\footnote{\text{Suius}\rightarrow{pip} install tensorflow}} \text{Collecting tensorflow} \text{Downloading tensorflow-2.5.0-cp38-cp38-win_amd64.whl (422.6 MB)} \text{Users}\footnote{\text{Users}\text{Vsers}\text{Users}\text{Vsers}\text{Users}\text{Vsers}\text{Users}\text{Vsers}\text{Users}\text{Vsers}\text{Users}\text{Vsers}\text{Users}\text{Vsers}

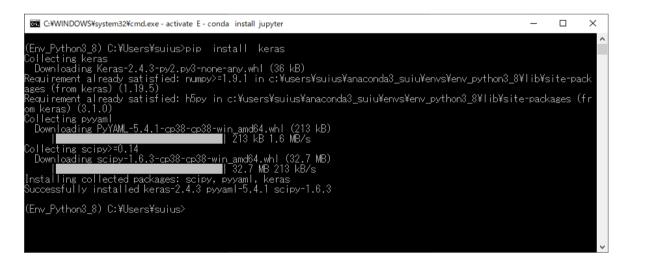
・場合によっては、インストール時に推奨メッセージが出ますので、 その推奨に従ってコマンドを発行します。

(3.2.6) Keras (ケラス)

・仮想環境のコマンドプロンプトから次のコマンドで

Keras をインストールします。

pip install keras



(3.2.7) その他のパッケージ

・仮想環境のコマンドプロンプトから次のコマンドで、 インストール済みのパッケージー覧を参照できます。

conda list [-n 仮想環境名] (凡例:[]内は省略可能)

以下は仮想環境 "Env_Python3_8" について、インストール済みのパッケージー覧を参照する例です。

conda list -n Env_Python3_8



(3.2.7.1) Anaconda で提供しているパッケージ

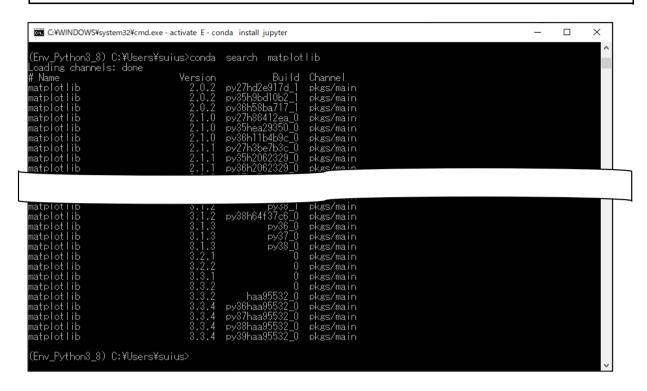
Anaconda で提供しているパッケージについて、

仮想環境のコマンドプロンプトから、次のコマンドで、パッケージのバージョン一覧を参照できます:

conda search パッケージ名

以下の例では、パッケージ "matplotlib" のバージョン一覧を参照します。

conda search matplotlib



・インストール済みでないパッケージのうち、Anaconda で提供しているパッケージについては、 仮想環境のコマンドプロンプトから次のコマンドで、

バージョンを指定してパッケージをインストールします。

・上記コマンドで、バージョンを指定しないでインストールする場合、 最新バージョンがインストールされます。 他のパッケージのバージョンとの整合性を考慮する場合は、バージョンを指定します。

・ここでは、以下の各コマンドで各パッケージを最新のものでインストールします。

```
conda install matplotlib
conda install pandas
conda install pillow
```

・インストール済のパッケージのバージョンが目的のものでない場合は、"conda uninstall パッケージ名" コマンドで、アンインストールした後で、バージョンを指定してパッケージを再インストールします。 以下は、パッケージ 'pandas' をバージョン '0.22.0' で再インストールする例です。

```
conda uninstall pandas
conda install pandas==0.22.0
```

(3.2.7.2) Anaconda で提供していないパッケージ

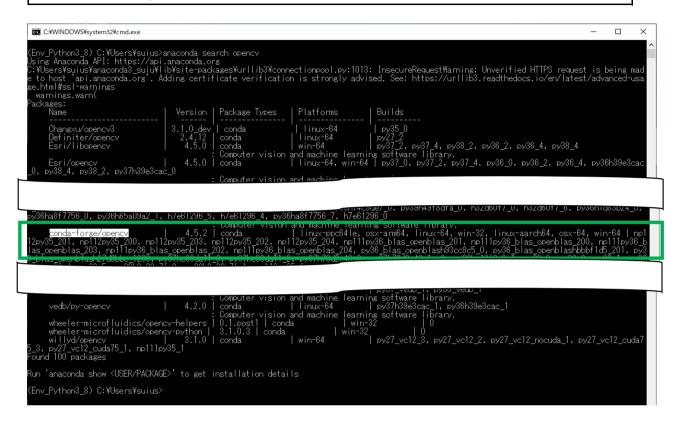
・Anaconda で提供していないパッケージの場合、

仮想環境のコマンドプロンプトから、次のコマンドでパッケージ提供しているチャネルがあるか探します。

anaconda search パッケージ名

以下は、パッケージ "opencv" を提供しているチャネルがあるか探す例です。

anaconda search opency



・パッケージを提供しているチャネルの検索で複数の検索結果がある場合、

バージョンと適用環境などから判断して、提供チャネルを選択します。

この場合もチャネルの優先順位など、様々な注意が必要です。

詳しくは、以下の公式ドキュメントを参照してください。

「https://conda.io/projects/conda/en/latest/user-guide/tasks/manage-channels.html]

・提供チャネルが決定したら、次のコマンドでインストールします。

最新バージョンをインストールする場合は、チャネルの指定は不要です。

|conda install [-c チャネル名] パッケージ名 (凡例:[]内は省略可能)

・以下は、パッケージ "opencv" の最新版をインストールする例です:

conda install opencv

C+WINDOWS+system32+cmd.exe - conda install opencv — □

(3.2.8) インストールしたパッケージ

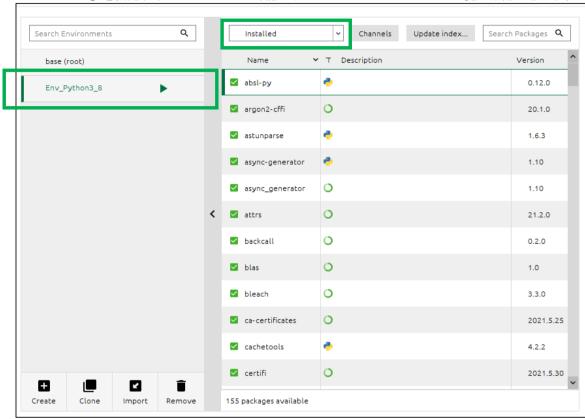
・(3.2.7) で既に述べたとおり、仮想環境のコマンドプロンプトから次のコマンドで、

インストール済みのパッケージー覧を参照できます。

conda list [-n 仮想環境名] (凡例:[]内は省略可能)

「Anaconda Navigator」で仮想環境の利用可能パッケージー覧から、 「Installed」を指定するとインストール済みのパッケージー覧が確認できます。

「Not installed」を指定するとインストール可能で未インストールのパッケージ一覧が確認できます。



・コマンドプロンプトを閉じるには、"exit" コマンドを実行します。

(Env_Python3_8) C:¥Users¥suius>exit

(3.2.9) Jupyter Notebook による動作確認

・作成した仮想環境で、「Open with Jupyter Notebook」のメニューを選択します。

以下は仮想環境「Env_Python3_8」での例です:



・するとブラウザが起動され、「Jupyter Notebook」の画面が表示されるので、

右上の「New」から「Python3」を選択します。



・すると「Jupyter Notebook」が表示されます。

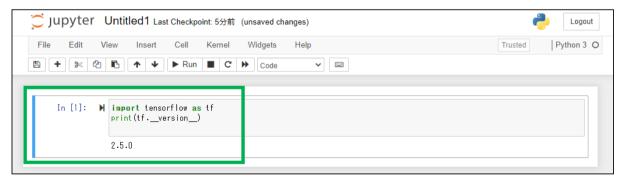
以下のコードを入力して、[Shift]+[Enter]キー(あるいは、[Ctrl]+[Enter]キー)を押すと、import Lt. Tangar [low のが がっとがまニュカナナ

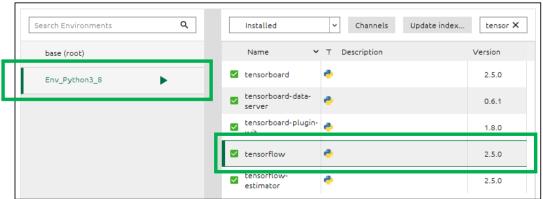
importした TensorFlow のバージョンが表示されます。

これが、「(3.2.5) TensorFlow (テンソルフロー)」でインストールしたバージョンと同じであれば、

「Jupyter Notebook」でバージョンの確認がとれたことになります。

import tensorflow as tf
print(tf.__version__)





(4) 開発環境の作成 (Anaconda以外)

・サーバー上に公開されている Pythonコードの開発環境として、「Google Colaboratory」と「AWS (特に "Amazon SageMaker")」を紹介します。

(4.1) Google Colaboratory

Pythonコードの開発環境として、Google は特別な設定なしで無料で使用できる「Google Colaboratory (グーグル コラボラトリ)」という環境をサーバー上に公開しています。
 以下の手順で用意します:

- (2) 「Google Colaboratory」へ手順(1)で作成したアカウントでログインします: 「Google Colaboratory (https://colab.research.google.com)」
- (3) ログインすると、以下のような画面が表示されるので、「ノートブックを新規作成」を選択すると、別のウィンドウで「Colab ノートブック」という、コードを記述して実行できる環境が開きます。

あるいは「Colaboratory へようこそ」画面で、「ファイル」メニューから「ノートブックを新規作成」でも同じことができます:





- (4) これより後は、Anaconda 環境で Jupyter Notebook を開いた時と同様な操作になります。 ライブラリのインストールなどの手間は不要です。
- (5) ご自分で作成した Anaconda 環境で実行した場合、機械学習のPCリソース占拠で PC環境が使用できなくなることがありますので、上記環境での試行をお勧めします。 尚、ファイルの保存や、接続時間などの制限がありますのでご注意下さい。

(4.2) Amazon SageMaker

・Amazon が提供しているAWS(Amazon Web Services) では、AIのサービスも提供しており、 機械学習サービス開発と公開や、公開されている出来合いのAPI利用が出来ます:

(自作サービスの開発と公開用)

Amazon SageMaker : 機械学習モデルの開発と提供

(出来合いのサービス利用・・・AWS提供)

Amazon CodeGuru : コードレビューを自動化する

Amazon Comprehend : 有用な情報を発見・分析する為の自然言語処理

Amazon Forecast : 過去の履歴から将来を予測する時系列データ予測サービス

Amazon Lex: 音声やテキストを使用して対話型インタフェースを構築するサービスAmazon Personalize: ユーザ向けにパーソナライズした推奨をするための機械学習サービス

Amazon Polly : テキストを音声に変換する、音声認識サービス

Amazon Rekognition : 画像・動画の認識サービス

Amazon Textract: PDFや画像からテキストを抽出するサービスAmazon Transcribe: 音声をテキストに変換する、音声認識サービス

Amazon Translate : テキストを言語間で翻訳するサービス

(出来合いのサービス利用・・・3rdベンダー提供)

3rdベンダー提供サービスをAWS上で利用することも出来ます。

【参照URL】

AWS機械学習サービス⇒https://aws.amazon.com/jp/machine-learning/

(参考.1) conda と pip について

インストール時に、pip と conda コマンドを混在して使用するにあたり、様々な混乱があるとのことなので、少し記事を集めてみました。 以下の記事は、下記URLからの要約です。

【参照URL】「https://code.i-harness.com/ja/q/1405a9c」

【参照URL】「https://teratail.com/questions/14133」

【参照URL】「https://pypi.org/」

【参照URL】「http://onoz000.hatenablog.com/entry/2018/02/11/142347」

- ・Condaは「Continuum Analytics」によって提供される、Anacondaのパッケージマネージャで、Anacondaの外部でも使用できます。Conda を使用して任意の言語のパッケージを管理できます。Condaは、言語に依存しない環境をネイティブに作成します。
- pipは、Python環境で標準のパッケージマネージャです。pipは「virtualenv」に依存して、Python環境のみを管理します。
- ・Anaconda下では基本的に「conda」を使ってパッケージをインストールするのですが、 一部のパッケージはAnaconda社のレポジトリからは提供されていません。 そのような場合にとるべきアプローチはいくつかあります。
 - 1. デフォルト以外のレポジトリ(チャネル)からインストール (例: 「conda install -c matsci pymatgen)
 - 2. 自分でconda用のパッケージを作る
 - 3. pipを使ってインストール
- ・このうち最後の「pipを使ってインストール」をすると、condaとpipのパッケージが混ざって厄介なことになります。condaから入れたパッケージはpipからも認識されるものの、
 - 1. 依存関係のバージョン違い
 - 2. condaとpipのパッケージ名の違い (例: pyqt (conda) vs. PyQt5 (pip))

等から予期せずcondaのパッケージが上書きされてしまうことがあります。
その結果、パッケージ1つのインストールでAnaconda環境が壊れてしまい、
Anacondaそのものを再インストールしない限り修復困難になってしまうことがあります。
また、condaがハードリンクを用いてパッケージを共有している関係から一つの環境で失敗してしまったが最後、他の仮想環境まで破壊されることもあります。

・「conda install パッケージ名」でパッケージが見つからなかった場合に、 安易にpipから入れるのは危険です。

リスクを減らすためには例えば次の様な手順を踏みます。

- 「anaconda search パッケージ名」でパッケージを提供しているチャネルがないか探します。 あれば「conda install -c チャネル名 パッケージ名」等の方法でインストールします。 (この場合もチャネルの優先順位など、様々な注意が必要。詳しくは公式ドキュメント参照 「https://conda.io/docs/user-guide/tasks/manage-channels.html」)。
- 2. pipから入れたい場合、まずPyPI(パイパイ, パイピイアイ)のサイト(「https://pypi.org/」)から 該当するパッケージを探し、依存関係を調べておきます。 依存するパッケージのうち、condaからインストール可能なものは予めインストールしておきます。
- 3. 依存関係を満たしたら「pip install --no-deps パッケージ名」でパッケージをインストールし、動作確認します。

- ・あるいは別の選択肢として、以下のような対応があります。
 - 1. pipからしか入れられないパッケージを入れたい場合、新しいcondaの環境を作る。 (「conda create -n env python」)その環境内では「conda install」は一切用いません。
 - 2. Anacondaを使うのをやめます。

Python公式サイトのPythonを使い、パッケージはpipで導入します。 仮想環境については「venv」や「virtualenv」を用います。