データ解析マニュアル

-Macのパスワード: yamane.takahiro

-作業ディレクトリはMacの

/Users/yamane.takahiro/work/Random\_Forest/ActiGraph\_paper/release\_Hip/Experiments

-READMEというファイルも参考にしてください。

1. /Users/yamane.takahiro/work/Random\_Forestにおいて以下を実行

source env/bin/activate

0. Not-Processed/Combinedにある処理したいファイルを今のディレクトリにコピーする。

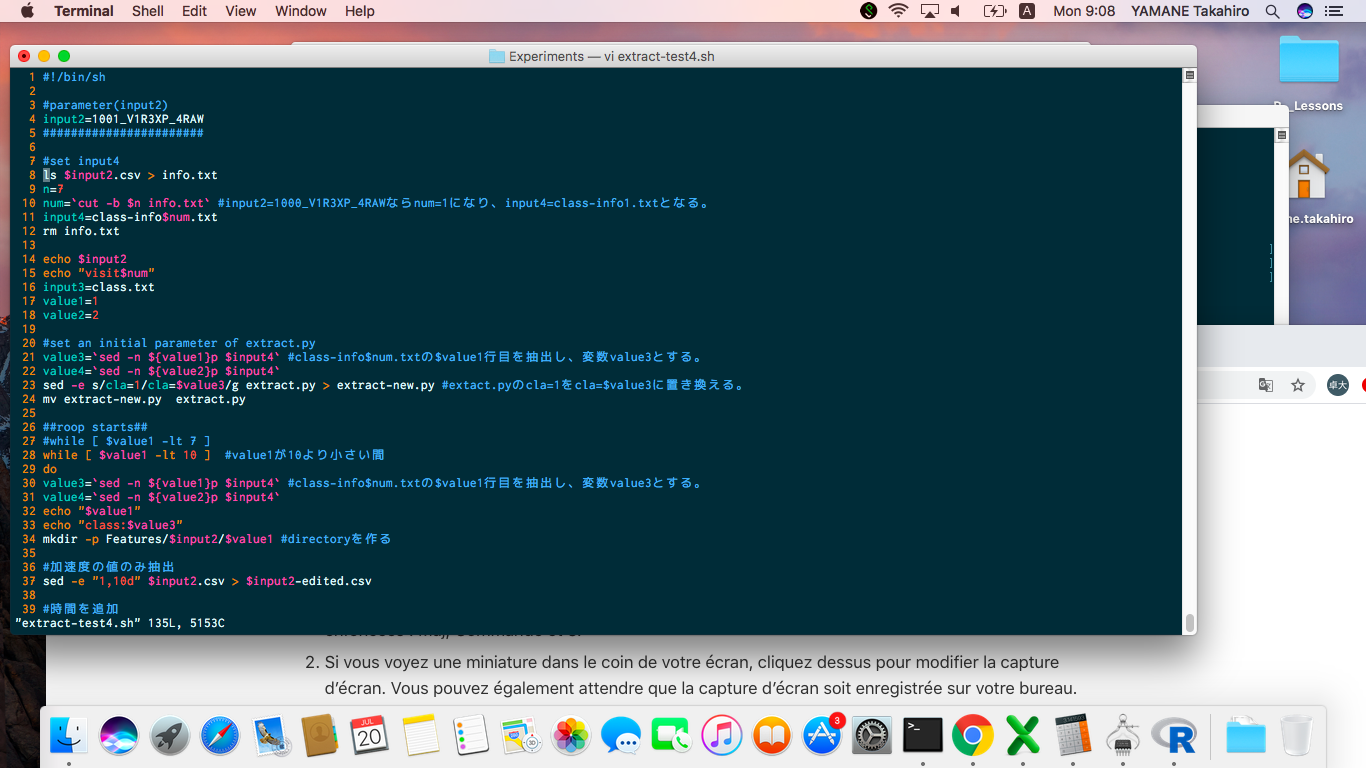
(例) cp Not-Processed/Combined/1000\_V1R3XP\_4RAW.csv .

1. extract-test4.shを編集する。

vi extract-test4.sh

4行目のinput2=1001\_V1R3XP\_4RAWを編集する。

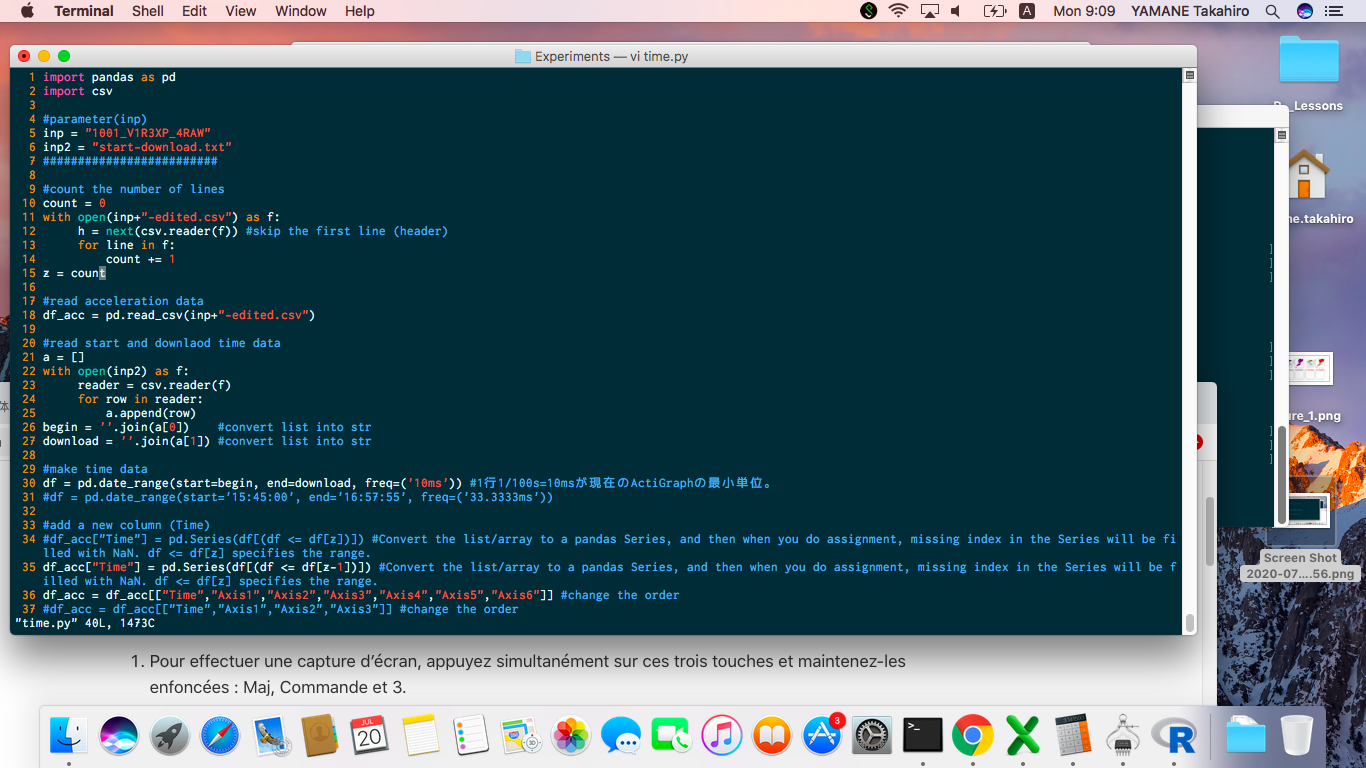
5行目のnumber=1001を編集する(写真には載っていない)。



2. time.pyを編集する。

vi time.py

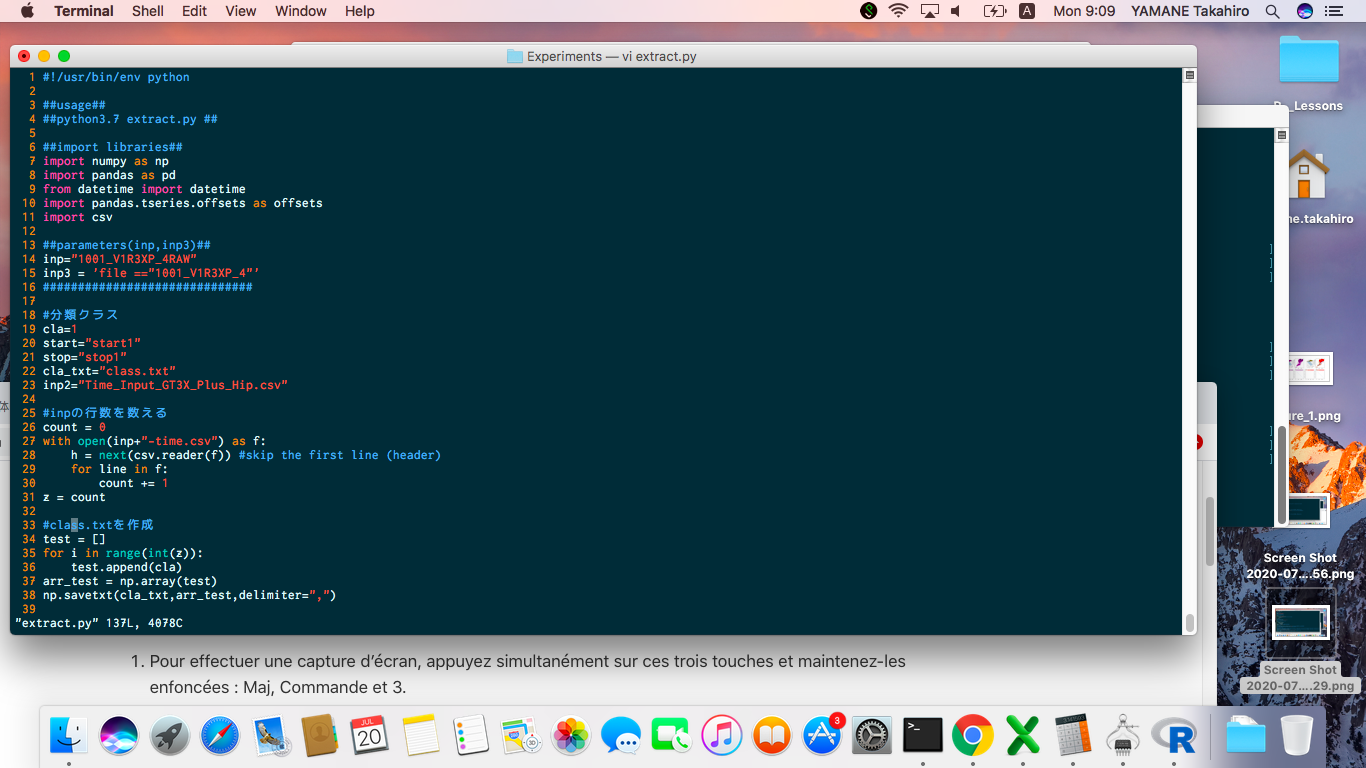
5行目のinp=”1001\_V1R3XP\_4RAW”を編集する。



3. extract.pyを編集する。

vi extract.py

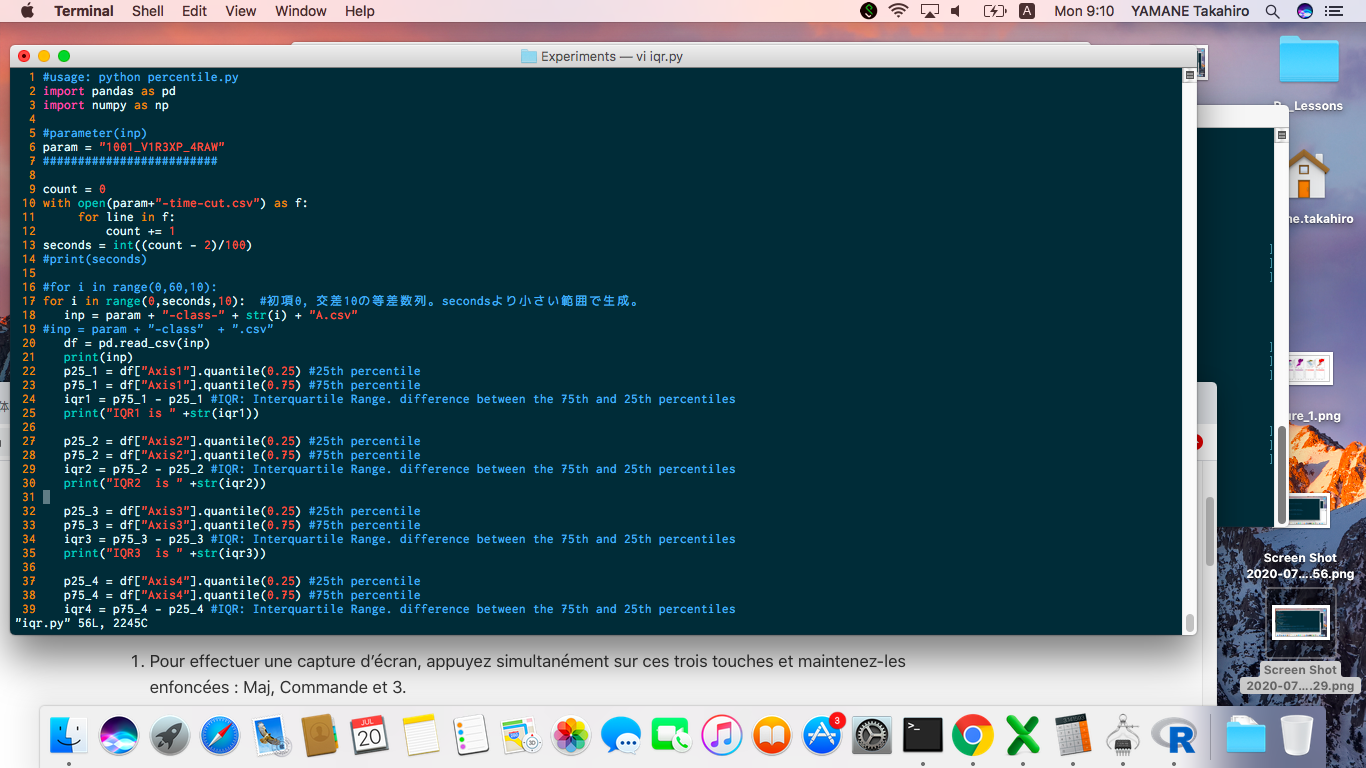
14行目のinp=”1001\_V1R3XP\_4RAW”と15行目のinp3 = ‘file ==”1001\_V1R3XP\_4”を編集する。



4. iqr.pyを編集する。

vi iqr.py

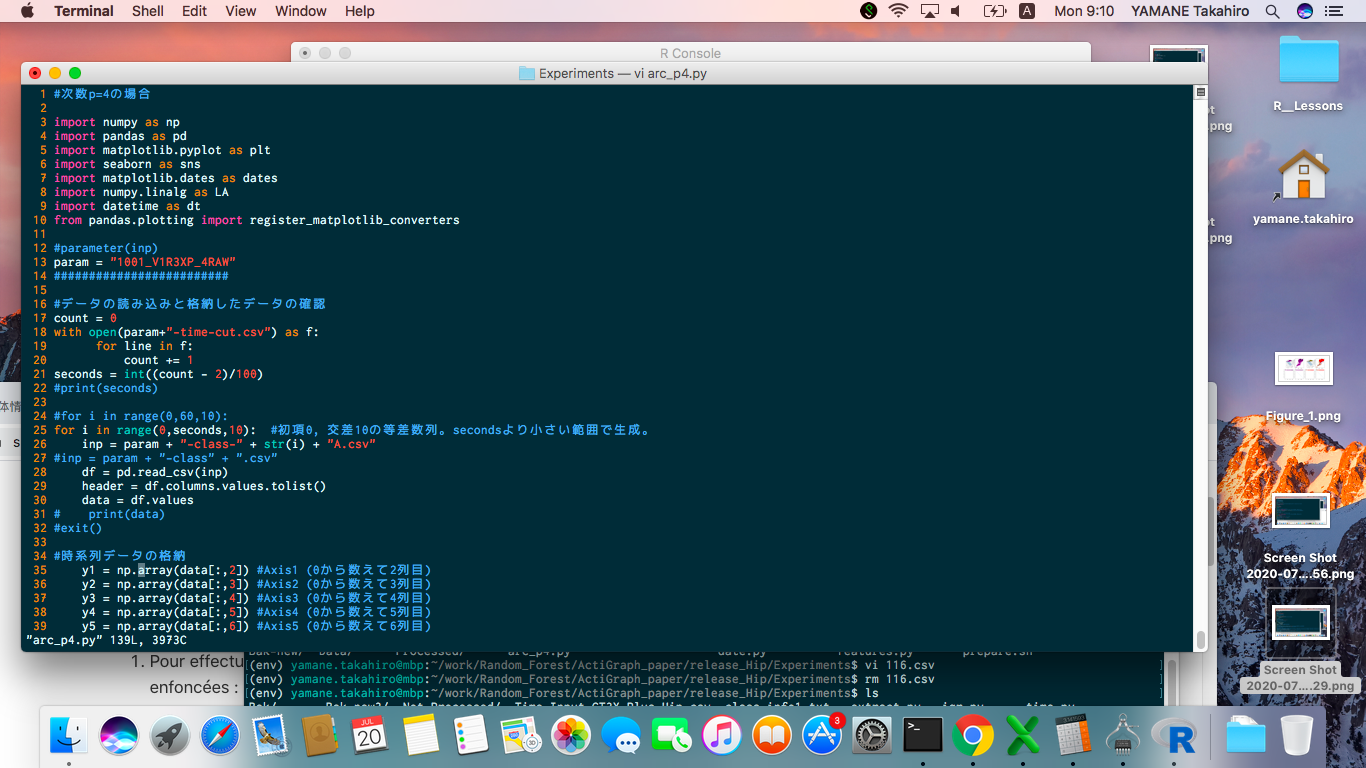
6行目のparam = “1001\_V1R3XP\_4RAW”を編集する。



5. arc\_p4.pyを編集する。

vi arc\_p4.py

13行目のparam = “1001\_V1R3XP\_4RAW”を編集する。



6. extract-test4.shを動かしてデータを切り出す。

sh extract-test4.sh

0-6を繰り返す。

7. choose.shを編集する。

vi choose.sh

4,5,6行目を編集する。

Value1には処理を開始する番号、value2には処理を終える番号に1を足した値を入れる。Choiceには各データからいくつ選択するかを入れる。

(例)1000~1010まで処理し、それぞれ6ブロック(1ブロックは10秒)ずつ

用いるのであれば

Value1=1000, value2=1011, choice=6となる。

8. choose.shを動かして特徴量を結合する。

sh choose.sh

features\_edited.csvという特徴量を計算したfileが出来る。

9. IDを振りなおして連番にする。1034-1037は抜けているので番号を振り直す必要がある。

cut -d “,” -f 1 features\_edited.csv > t

vi t

編集してIDを置換し、振り直す。例：1038→1034

cut -d “,” -f 2- features\_edited2.csv > t2

paste -d “,” t t2 > features\_edited-1011-1041.csv

このfeatures\_edited-1011-1041.csvを機械学習に用いる。

10. random-forest-onehot.pyを編集。

vi random-forest-onehot.py

17行目を編集して入力する特徴量ファイルを決める。

23, 37行目を編集して訓練データと確認データの範囲を決める。

\*確認データ、訓練データにはそれぞれ全ての行動が含まれるようにする。そうしないとinconsistent shapesというエラーが出る。

\*確認データ、訓練データに一人のデータがまたがらないようにする。

11. cross-validation.shを編集。

vi cross-validation.sh

4行目を編集してk-hold cross validationとする。

5,6行目を編集して最初の訓練データ、確認データの範囲を決める。

8,9行目を編集して2つめの訓練データ、確認データの範囲を決める。

10行目を編集してシフト幅を決める。

12. cross-validation.shを実行して機械学習を行い、行動を予測する。クロスバリデーションを行ってくれる。

sh cross-validation.sh

(+α)

sh importance.sh により各特徴量の重要度を評価できる。生成するimportance1e.csvなどを確認する。

-1000-1010まではsleep modeに入ってしまったのでエラーが生じることがある。

-1011-1033,1038-1045は問題なし。→1011-1041として連番にして処理。

-1,2,3,4,5,6,7,8,9,10の行動は混同行列では0,1,2,3,4,5,6,7,8,9となっている。特徴量のcsvでは1,10,2,3,4,5,6,7,8,9と並んでいるが混同行列では10は9に対応する。