科目名：プログラミング技法II 担当教員名：新田　直子

学籍番号：01Z18032 名前：トゥチェク・トム

2019年06月06日

**平均，分散，標準偏差、場所間の距離、線形回帰**

課題２：平均，分散，標準偏差

　N（＝25）個のランダムな、－100以上100未満の値でリストを作った。平均は、合計を要素数で割ったものである。分散は、個々の値と平均との違い二乗、その合計をまた要素数で割ったものである。標準偏差は、分散の平方根となる。

　それらの方法でできた結果をモジュールstatisticsの関数の結果と一緒に出力した。平均と分散には少し違う結果があっても、あまり有意はないと思う。標準偏差の有意な違いの原因は、平方根を計算する関数である。

課題２：実行結果表

|  |  |
| --- | --- |
| 自分で作った関数 | モジュールstatisticsの関数 |
| Mean:　15.332881970289058 | 15.332881970289058 |
| Variance:　3320.9053586176756 | 3320.905358617675 |
| Standard deviation:　 57.627300000789496 | 58.81561370271286 |
|  |  |
| Mean:　9.354598797483284 | 9.354598797483286 |
| Variance:　3961.221435390986 | 3961.221435390985 |
| Standard deviation:　62.9382000009658 | 64.23606719384583 |
|  |  |
| Mean:　-4.300679735644522 | -4.300679735644522 |
| Variance:　2124.5622577029017 | 2124.5622577029017 |
| Standard deviation:　46.092900000406594 | 47.04344465605372 |

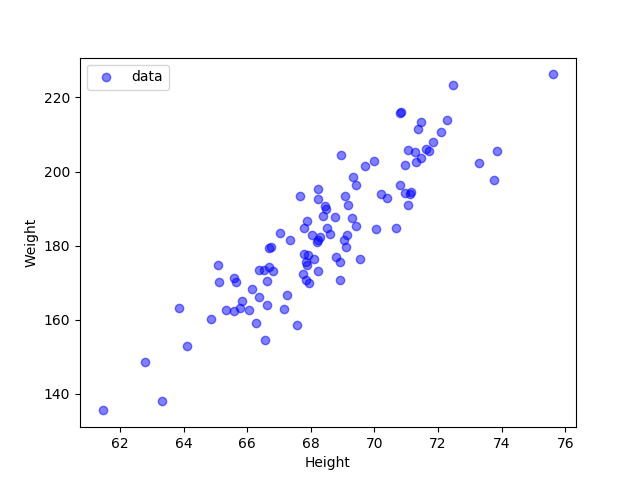
課題3: 場所間の距離

　距離を計算する関数は、式の通り、モジュールmathを使用しながら実装した。各地のデータをファイルから読み込み、辞書に変換した。まず、辞書を全部もラジアンに変換したが、それは無駄なので、後で選択された地点だけを変換した。その地点は名前の入力で選択される。入力された名前の一つも辞書にはない場合、エラーメッセージを出力する。両方の名前がある場合、距離を作った関数で計算し、出力する。

課題３：実行結果表

|  |  |  |
| --- | --- | --- |
| 入力１ | 入力２ | 出力 |
| Osaka | Tokyo | distance: 403.60244096985537 km |
| Toono | Nagoya | distance: 580.1441848848292 km |
| Suita | Toyonaka | distance: 4.840757283316017 km |
| Hokkaido | Okinawa | distance: 2208.1834787207986 km |
| Vienna |  | Error: Could not find Vienna |

課題４: 線形回帰

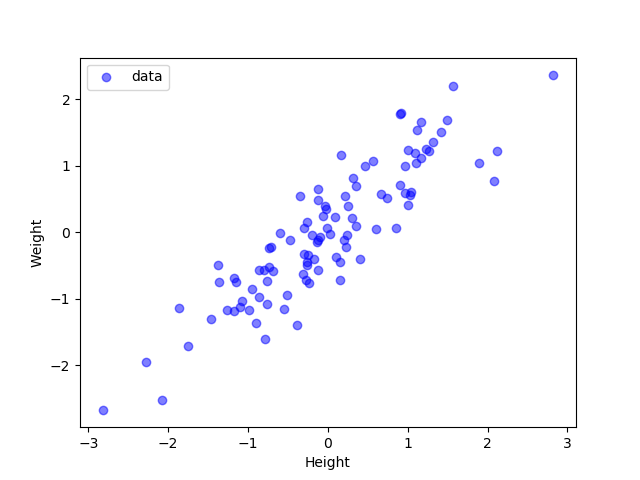


課題4-1

　データを読み込み、男性のデータをリストへ、zipとrandom.sampleで、N＝100個をランダムに抽出した。身長、体重をそれぞれのリストに変換し戻った。それらのリストで、プロットを出力する。

図 1　課題4-1の結果

課題4-2



　 標準化する関数を作成した。個々の値マイナス平均、標準偏差で割ると、標準化されたデータを得る。出力されたプロットは大体、前回と同じ見えるが、値が（0,1）の正規分布に従う。

図 2　課題4-2の結果

課題4-3

　MSEの関数を式の通り実装した。

|  |  |  |
| --- | --- | --- |
| a | b | 出力 |
| 2 | 2 | 5.794538806970911 |
| 0 | 0 | 0.9900000000000003 |
| 1 | 1 | 1.251836840620225 |
| 10 | 10 | 183.17735672606767 |
| -10 | -10 | 216.93241480351992 |

課題4-4

　df/daの関数はfdaという、df/dbの関数はfdbというになった。それでaとbの更新を、限界を超えること（max\_iterations = 1000）または、更新は十分低くなること（precision = 0.005）まで繰り返す。そうすると、関数からa、bとMSEのデータを得る。それぞれ、グラフに示す。

　初期値を変化すると、MSEの最初の値が強く変える。それは図3と図4に示したデータからわかる。また、学習率は繰り返しの数に関係がある。学習率は高くすると、もっと早く終わるが、結果の正確さが低くなる。

図 4　初期値は1、学習率は0.1のMSEグラフ

図 3　初期値は10、学習率は0.3のMSEグラフ

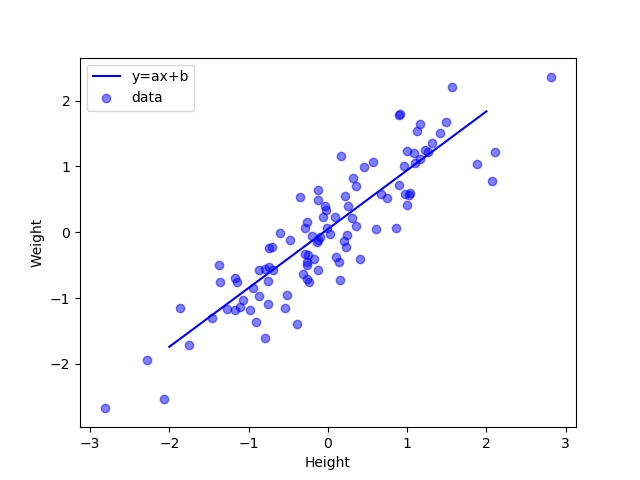
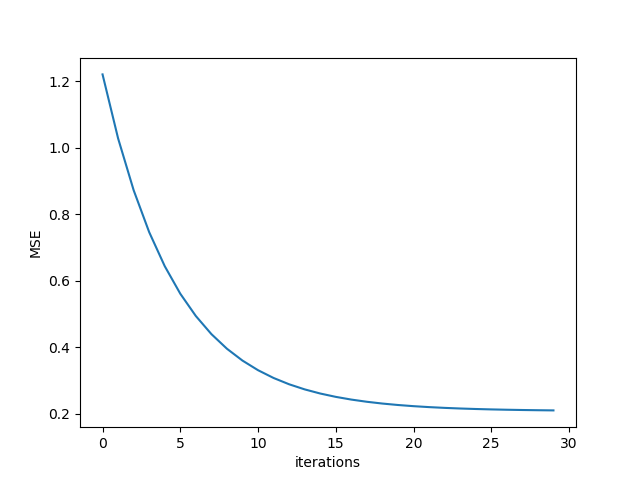
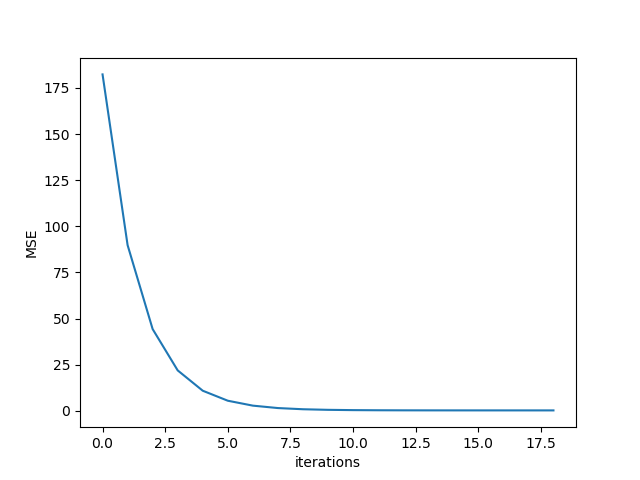
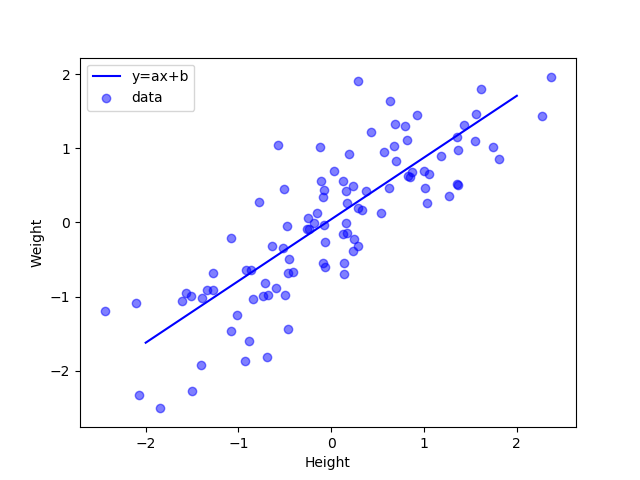


図 5　データとともにa、bを示すグラフ

課題4-5



　この課題は私にとって一番難しかった。データをNumpyの行列に変換するのは、どうするかわからなかったが、先生が手伝ってくださったおかげで、なんとかできた。それで、更新の関数を変化して、課題4－4と同じような結果が得る。

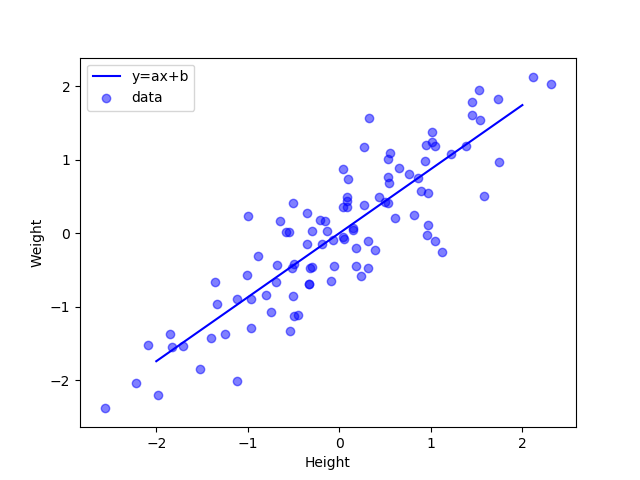


図 6　課題4－5の結果

課題4-6

　簡単な公式で、前回と同じような結果を得る。。

課題4-7

　np.polyfitでまた、同じような結果を得る。polyfitの関数には、普通の身長・体重の配列が必要だ。動作検証は下記、4－8のところでする。

図 7　課題4－6の結果

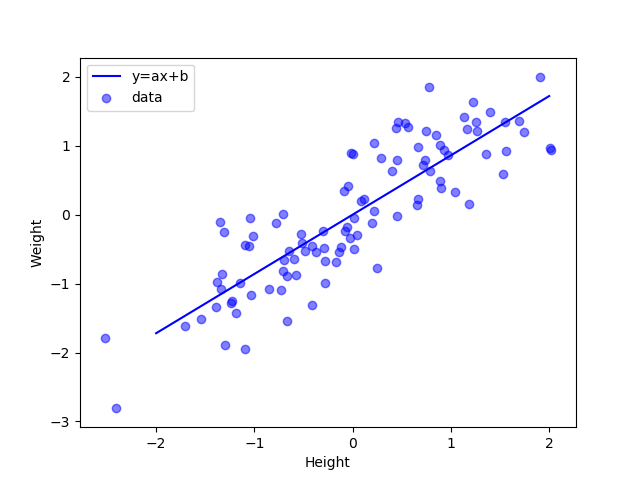


　　　　　　　図 8　課題4－7の結果

課題4-8

　それぞれの関数の結果を一緒に示すと（図9）、小さな違いが見えるが、それはあまり有意ではないと思う。それは、標準化後だ。標準化前、課題4－4と課題4－5の関数は無限に関する問題が起こるので、結果が出られない。それらのアルゴリズムには、標準化された値が必要であるようだ。

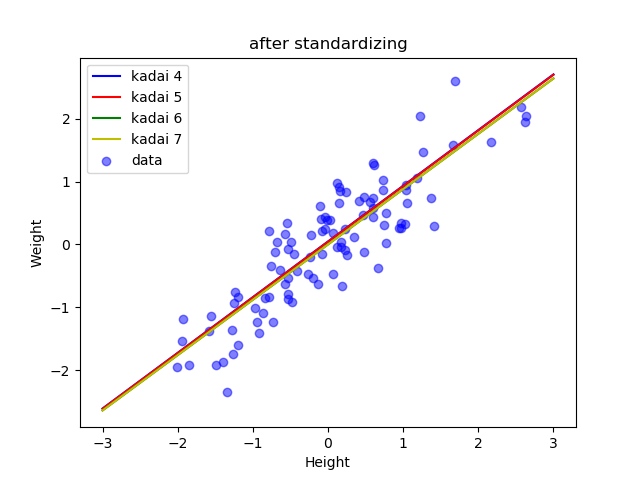


　　　　図 9　それぞれの関数の結果の比較（標準化後）

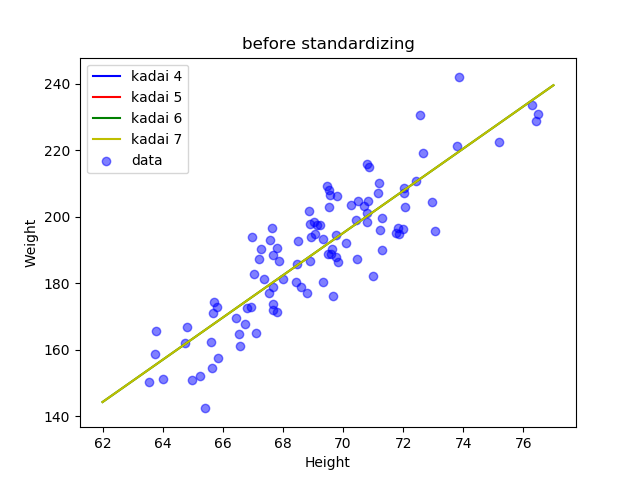


　　　　図 10　それぞれの関数の結果の比較（標準化前）