機械学習 課題 1

高林秀
三宅研究室 博士前期課程 1 年
V-CampusID : 23vr008n

June 9, 2023

Abstract

本稿は本年度必修授業の機械学習の第 1 回レポートの答案用紙である。
本稿は、第 1 回授業〜第 4 回授業までの範囲を対象とし、各回で課された課題に対する解答を記載する。
答案の問題番号は各章のタイトルに記載している。
各問に対する解答は本稿に、コードなどの実行結果は別途 GoogleColaboratory のノートブックに記載した巻末の付録から参照できる。

1 第 1 回授業 : 4/11 宿題 1

1.1 問題文

二つのデータ点の近さを測る定量的指標の一つとして距離がある。いま、次の三つのデータ点があるとする

$$\text{データ点 } 1 : (5, 2, 5.8) \tag{1}$$
$$\text{データ点 } 2 : (7, 10, 1, 12) \tag{2}$$
$$\text{データ点 } 3 : (3, 2, 6, 3) \tag{3}$$

上の各データ点は 4 種類の計測値で与えられている（例えば「緯度、経度、水深、温度」のような感じ）。このデータ点同士の間の近さをユークリッド距離で計算し、互いの距離が近いペアの順番を答えよ。データが近いほど距離が小さいことに注意。

1.2 解答

データ点の近さは以下の順番である。

1. データ店 1 とデータ点 3、距離 : 5.4
2. データ店 1 とデータ点 2、距離 : 10.0
3. データ店 2 とデータ点 3、距離 : 13.6

ユークリッド距離は以下の式で計算できる。

$d(x, y)$: データ点 x とデータ点 y の距離として、

$$d(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

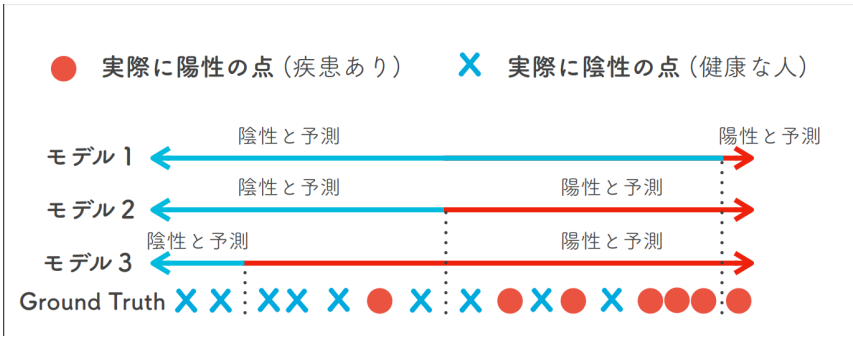
※ x_i, y_iはそれぞれデータ点 x, y の i 番目の要素を表す

各点間の距離は以下のように計算できる

$$d(1,3) = \sqrt{(5-3)^2 + (2-2)^2 + (5.8-6)^2 + (0-3)^2} = 5.4$$
$$d(1,2) = \sqrt{(5-7)^2 + (2-10)^2 + (5.8-1)^2 + (0-12)^2} = 10.0$$
$$d(2,3) = \sqrt{(7-3)^2 + (10-2)^2 + (1-6)^2 + (12-3)^2} = 13.6$$

2　第 2 回授業 : 4/18 宿題 2

2.1　問題文



1. 上の三つの分類モデルそれぞれに対し、混同行列、Precision、Recall を計算しなさい。
2. モデル 1 →モデル 2 →モデル 3 の順に、Precision は増加や減少の傾向にあるでしょうか？その傾向はなぜ生じるでしょう？また、Recall についても傾向はどうなっていますか？Precision の傾向と対比して議論して見てください。

2.2　解答: 問 1

まず、各モデルの混同行列と Precision、Recall を計算した結果は以下のようになる。

2.2.1 モデル１

		予測	
		陰性	陽性
実	陰性	1	0
データ	陽性	6	9

Table 1: モデル 1 の混同行列

- Precision : 1.0
- Recall : 0.14

2.2.2 モデル 2

		予測	
		陰性	陽性
実	陰性	6	3
データ	陽性	1	6

Table 2: モデル 2 の混同行列

- Precision : 0.66
- Recall : 0.85

2.2.3 モデル 3

		予測	
		陰性	陽性
実	陰性	7	7
データ	陽性	0	2

- Precision : 0.5
- Recall : 1.0

2.3 解答: 問２

計算結果から、Precision（適合率）は減少の傾向にあるといえる。Precision は、陽性と予測したデータ数のうち、実際に陽性だったデータ数の割合を示す。計算

式で示すと以下の通り。

$$Precision = \frac{TP}{TP + FP}$$

※ TP は真陽性、FP は偽陽性を表す

この式より、Precision は分母の、モデルが陽性と予測した数が少なければ値が大きくなる傾向にあることがわかる。モデル 1 では陽性と予測されたデータ数は 1 で、モデル 2 では 9、モデル 3 では 14 と徐々に陽性予測のデータ数が増えており、Precision の式の分母が大きくなっているので、モデル 1〜モデル 3 にかけて Precision は減少していると考えられる。

また、Recall（再現率）は増加の傾向にあるといえる。Recall は、実際に陽性だったデータのうち、モデルが正しく陽性と予測したデータ数の割合を示す。計算式で示すと以下の通り。

$$Recall = \frac{TP}{TP + FN}$$

※ TP は真陽性、FN は偽陰性を表す

この式より、Recall は分母の、実際は陽性だが予測では陰性だった割合が少ないほど Recall は大きくなる傾向にあることがわかる。実際は陽性だったデータのうち、モデルが陰性と予測したデータの割合がモデル 1〜モデル 3 にかけて減少したことから、Recall は増加していると考えられる。

3 第 3 回授業： 4/25 宿題 1

3.1 問題文

X の数値スケールが大きく (例えば 0〜1000000)、Y のスケールが小さい (例えば 0〜1) データに対して、スケーラを使わないで回帰をしたとする。

1. 勾配降下法で学習率を大きくしても小さくしても問題が起こることを説明せよ。
2. パラメータ a, b ごとに学習率を変えて勾配降下法をすれば、スケーラを使わなくても 1. の問題は抑制できるか？具体的な誤差関数の形を使って議論せよ。

3.2 解答

4 第 4 回授業：

5 付録

- GoogleColab ノートブック：
- 提出用 GoogleDrive フォルダ：