

Progress

Mizuno Yasuaki

October 24, 2022

1. アミノ酸配列の画像作成
2. 畳み込みニューラルネットワークによる学習
3. k 分割交差検証
4. まとめ

アミノ酸の画像作成

Table 1: アミノ酸配列のベクトル割り当て①

アミノ酸	x 成分	y 成分	アミノ酸	x 成分	y 成分
A	2.5	1.10	M	6.0	1.90
C	3.0	2.50	N	5.0	-3.50
D	2.5	-3.60	P	5.5	-1.90
E	5.0	-3.20	Q	6.0	-3.68
F	2.5	2.80	R	7.5	-5.10
G	0.5	-3.68	S	3.0	-0.50
H	6.0	-3.20	T	5.0	0.70
I	5.5	4.50	V	5.0	-0.46
L	5.5	3.80	Y	7.0	-1.30

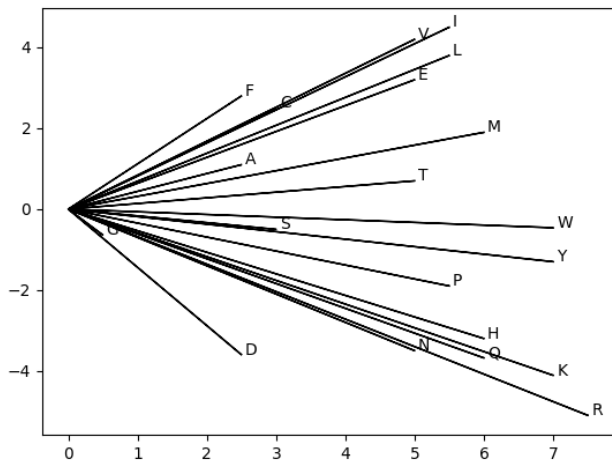


Figure 1: アミノ酸のベクトル割り当て①

Test Accuracy: 0.9507

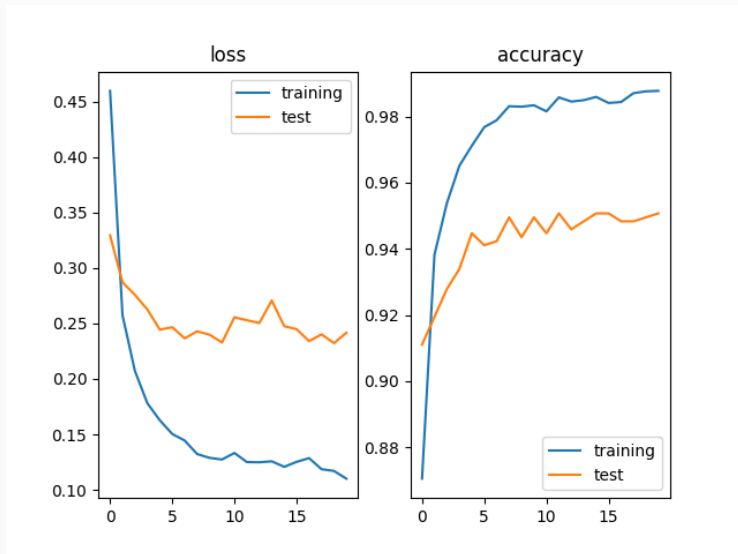
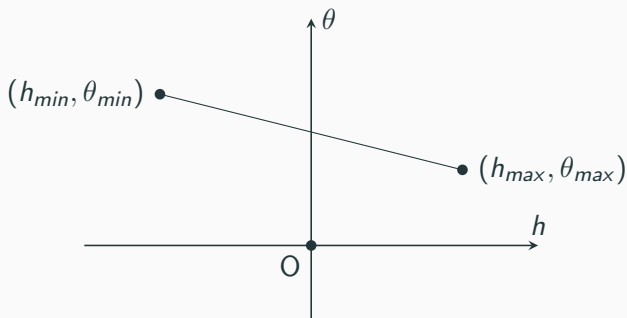


Figure 2: 學習結果

アミノ酸配列の画像作成

アミノ酸の疎水性度の最大値 h_{max} と最小値 h_{min} に対する角度をそれぞれ θ_{max} と θ_{min} と置く。



任意のアミノ酸の疎水性度を h_{amino} 、角度を θ_{amino} とすると

$$\theta_{amino} = \frac{\theta_{max} - \theta_{min}}{h_{max} - h_{min}} \times (h_{amino} - h_{max}) + \theta_{max} \quad (1)$$

となる。 $(h_{max}, \theta_{max}) = (4.50, 10)$ と $(h_{min}, \theta_{min}) = (-5.10, 170)$ をそれぞれ式 (1) に代入する。

$$\theta_{amino} = \frac{50}{3} \times h_{amino} + 85 \quad (2)$$

Table 2: アミノ酸の角度割り当て②

アミノ酸	疎水性度	θ	アミノ酸	疎水性度	θ
R	-5.10	170.0	G	-0.64	95.6
K	-4.11	153.5	S	-0.50	93.3
Q	3.68	23.6	W	-0.46	92.6
D	-3.60	145.0	A	1.10	66.6
N	-3.50	143.3	M	1.90	53.3
H	-3.20	138.3	C	2.50	43.3
E	-3.20	138.3	F	2.80	38.3
P	-1.90	116.6	L	3.80	21.6
Y	-1.30	106.6	V	4.20	14.9
T	0.70	73.3	I	4.50	10.0

疎水性度に同じ値があるので、アミノ酸のベクトル割り当て①のようにアミノ酸の大きさの情報を加える。アミノ酸の大きさを r_{amino} とすると、極座標を用いて

$$(r_{amino}, \theta_{amino}) \quad (3)$$

と表すことができる。

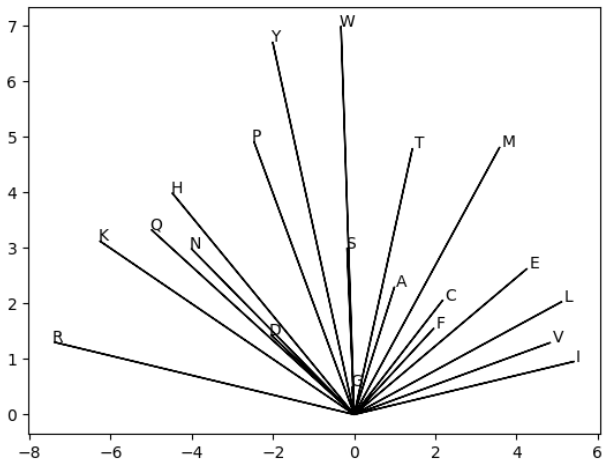


Figure 3: アミノ酸のベクトル割り当て②

アミノ酸配列の画像比較

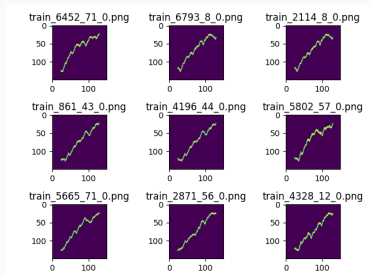


Figure 4:

ベクトル①を用いた画像

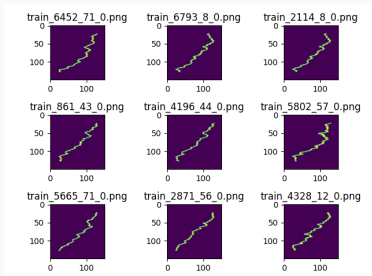


Figure 5:

ベクトル②を用いた画像

Test Accuracy: 0.9519



畳み込みニューラルネットワークによる学習

単純な畳み込みニューラルネットワークのモデル

- 畳み込み層
チャンネル数:32、フィルタサイズ:(3, 3)、活性化関数:Relu
- マックスプーリング層 プーリングサイズ:(2, 2)
- Flatten 層
- ドロップアウト層 ドロップアウト率:0.3
- 全結合層 ノード数:32、活性化関数:Relu
- 出力層 ノード数:5、活性化関数:softmax
- バッチサイズ:128
- エポック:10

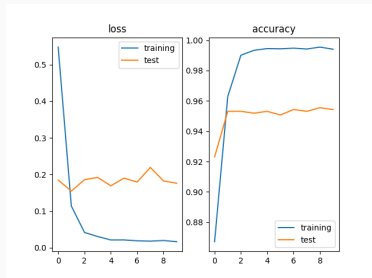


Figure 6:

アミノ酸配列画像①の学習結果

Accuracy: 0.9543

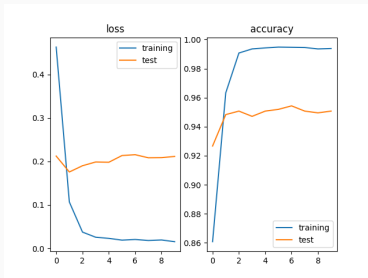


Figure 7:

アミノ酸配列画像②の学習結果

Accuracy: 0.9507

- 訓練データを同じサイズの k 個のサブセットに分ける
- $(k - 1)$ 個のサブセットで訓練し、残りのサブセットで評価する
- 最終的に k 個のスコアの平均



Figure 8: 3 分割交差検証

Table 3: 4 分割交差検証

フォールド	アミノ酸配列の画像①	アミノ酸配列の画像②
0	0.9483	0.9447
1	0.9350	0.9507
2	0.9531	0.9579
3	0.9555	0.9519
Average	0.9480	0.9513

- 画像やニューラルネットワークを変更してみたがあまり精度は変わらなかった
- 交差検証をテストケースでおこなう
- 以前読んだ論文で使用されていた Attention 層や RNN 層を用いて精度を向上させる