

# 知能プログラミング演習 I 第 4 回レポート

2025 年 7 月 4 日 学籍番号 氏名

## 1 実験設定

本実験では、リカレントニューラルネットワーク（RNN）を用いた音声認識タスクにおいて、異なるハイパーパラメータ設定による性能比較を行った。

### 1.1 ネットワーク構造

- 入力層：77 次元（音響特徴量）
- 中間層（隠れ層）：1 層の RNN 層
- 出力層：10 次元（数字 0-9 の分類）

### 1.2 実験パラメータ

以下のパラメータ設定で実験を実施した：

- 中間層ユニット数：64, 128
- 最適化手法：SGD, Adam
- 活性化関数：Sigmoid 関数
- 学習率：0.001
- エポック数：30
- 訓練データ：発話番号 1-7
- テストデータ：発話番号 8-9, 0

### 1.3 データセット

Lyon decimation 128 を用いた音声データセットを使用し、スケーリング係数  $1 \times 10^3$  を適用した。

## 2 結果

### 2.1 誤差関数の推移

図 1 から図 4 に各設定における訓練誤差とテスト誤差の推移を示す。

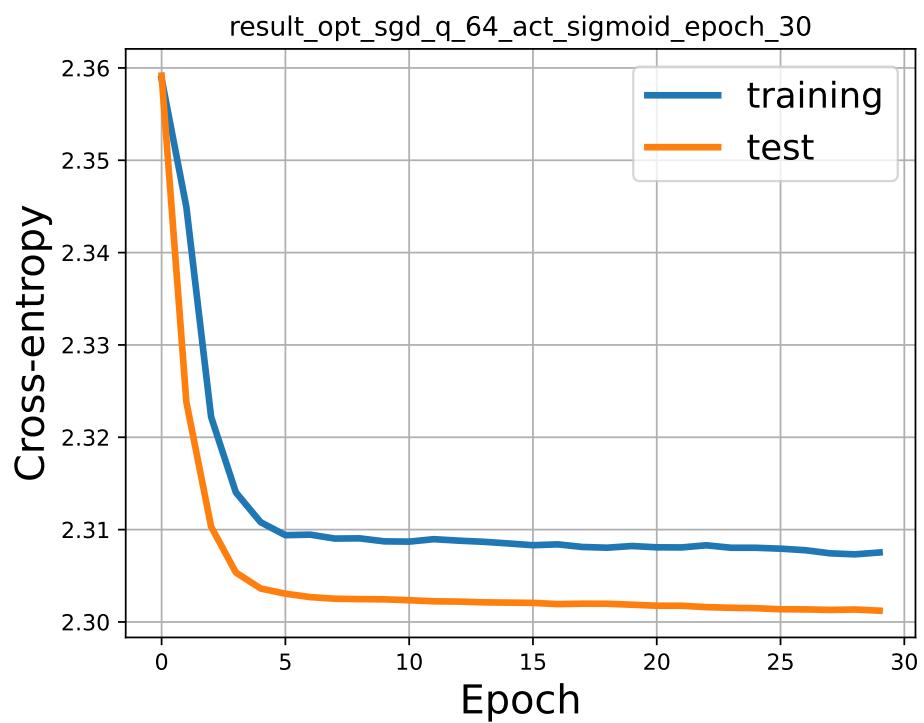


図 1: SGD, 64 ユニットにおける誤差推移

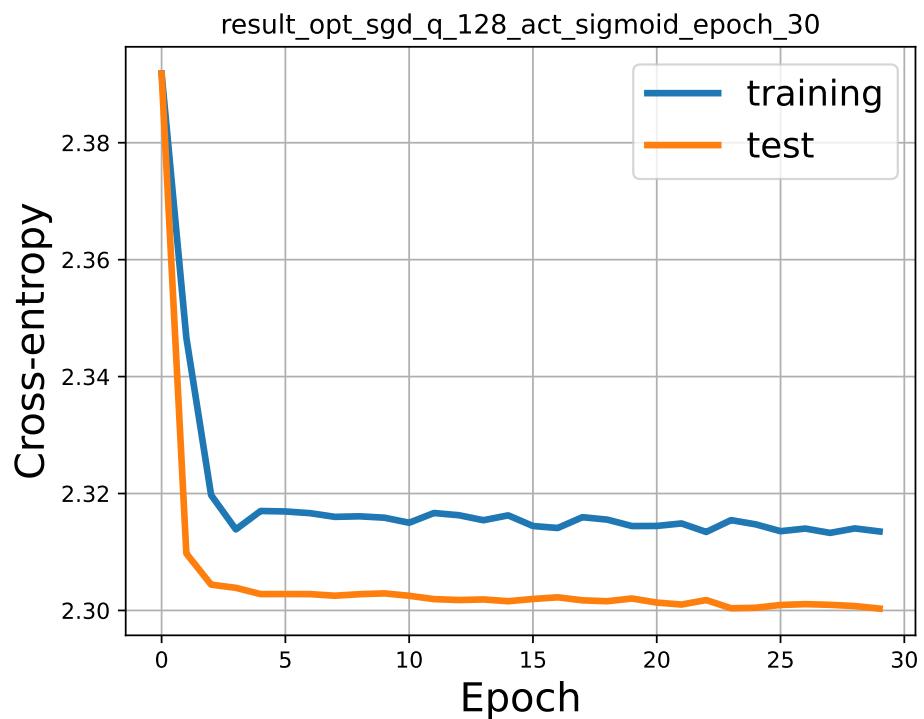


図 2: SGD, 128 ユニットにおける誤差推移

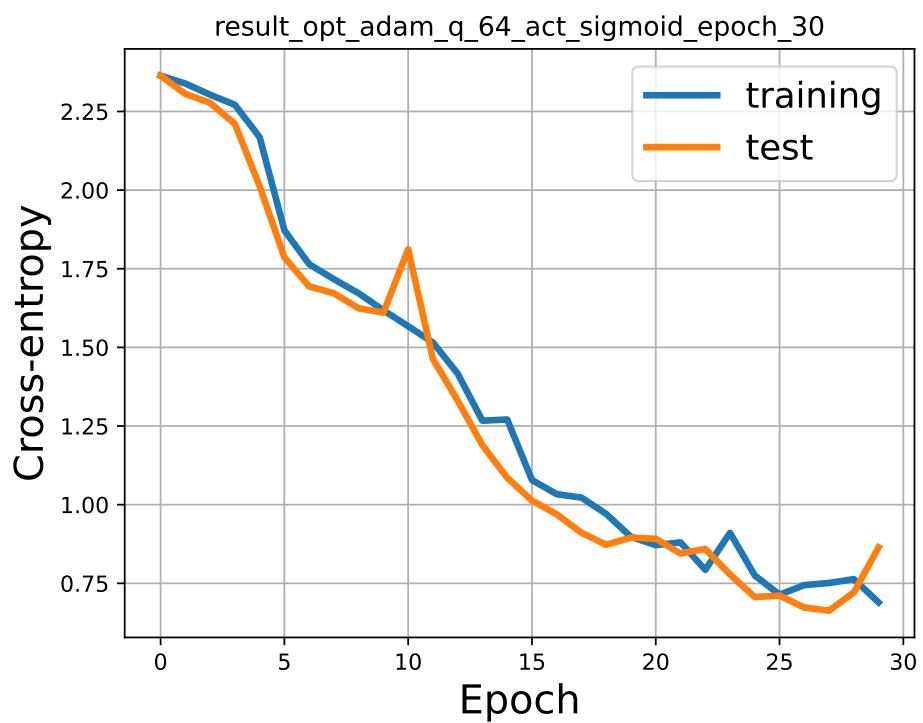


図 3: Adam, 64 ユニットにおける誤差推移

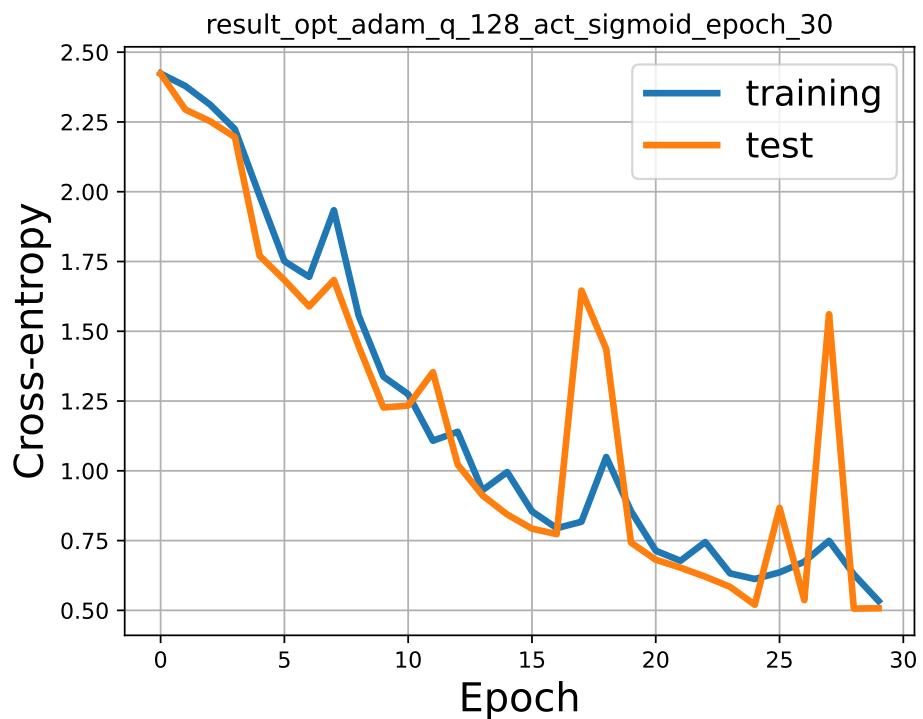


図 4: Adam, 128 ユニットにおける誤差推移

## 2.2 混同行列

図 5 から図 8 に各設定における混同行列を示す。

result_opt_sgd_q_64_act_sigmoid_epoch_30									
True	0	1	2	3	4	5	6	7	8
0	0	0	0	0	0	15	0	0	0
1	0	0	0	0	0	15	0	0	0
2	0	0	0	0	0	15	0	0	0
3	0	0	0	0	0	15	0	0	0
4	0	0	0	0	0	15	0	0	0
5	0	0	0	0	0	15	0	0	0
6	0	0	0	0	0	15	0	0	0
7	0	0	0	0	0	15	0	0	0
8	0	0	0	0	0	15	0	0	0
9	0	0	0	0	0	15	0	0	0

図 5: SGD, 64 ユニットにおける混同行列

result_opt_sgd_q_128_act_sigmoid_epoch_30										
True	0	1	2	3	4	5	6	7	8	9
Predict	0	0	0	0	0	15	0	0	0	0
0	0	0	0	0	0	15	0	0	0	0
1	0	0	0	0	0	15	0	0	0	0
2	0	0	0	0	0	15	0	0	0	0
3	0	0	0	0	0	15	0	0	0	0
4	0	0	0	0	0	15	0	0	0	0
5	0	0	0	0	0	15	0	0	0	0
6	0	0	0	0	0	15	0	0	0	0
7	0	0	0	0	0	15	0	0	0	0
8	0	0	0	0	0	15	0	0	0	0
9	0	0	0	0	0	15	0	0	0	0

図 6: SGD, 128 ユニットにおける混同行列

result_opt_adam_q_64_act_sigmoid_epoch_30										
True	0	1	2	3	4	5	6	7	8	9
Predict	0	15	0	0	0	0	0	0	0	0
0	0	15	0	0	0	0	0	0	0	0
1	1	0	7	5	0	1	0	0	0	0
2	0	0	1	13	0	0	0	0	0	0
3	0	0	0	0	13	0	0	0	2	0
4	0	0	0	0	0	15	0	0	0	0
5	0	0	0	0	0	3	7	4	0	1
6	0	0	0	0	0	0	0	15	0	0
7	0	0	0	0	1	2	1	7	4	0
8	0	0	0	0	0	0	0	13	0	2
9	0	0	0	0	0	0	0	0	10	1

図 7: Adam, 64 ユニットにおける混同行列

result_opt_adam_q_128_act_sigmoid_epoch_30									
True	0	1	2	3	4	5	6	7	8
0	15	0	0	0	0	0	0	0	0
1	0	15	0	0	0	0	0	0	0
2	0	6	9	0	0	0	0	0	0
3	0	0	0	15	0	0	0	0	0
4	0	0	0	0	15	0	0	0	0
5	0	0	0	0	1	13	0	1	0
6	0	0	0	0	0	0	15	0	0
7	0	0	0	0	0	0	0	15	0
8	0	0	0	0	0	0	13	0	2
9	0	0	0	0	0	1	0	9	0
	0	1	2	3	4	5	6	7	8
Predict									9

図 8: Adam, 128 ユニットにおける混同行列

### 3 考察

#### 3.1 最適化手法の比較

最適化手法として SGD を用いた場合（図 1, 2）、誤差が十分に収束せず、学習が停滞している様子が確認できる。混同行列（図 5, 6）からも、ほとんどのテストデータが一つのクラスに誤分類されており、有効な学習が行われていないことがわかる。一方、Adam を用いた場合（図 3, 4）では、訓練誤差・テスト誤差ともにエポック数の増加に伴い滑らかに減少し、学習が正常に進行している。これは、RNN の複雑な誤差曲面において、固定学習率の SGD では局所解に陥りやすいのに対し、パラメータごとに学習率を適応的に調整する Adam がはるかに有効であることを示唆している。

#### 3.2 隠れ層ユニット数の影響

Adam を用いた結果においてユニット数 64（図 3）と 128（図 4）を比較すると、128 ユニットの方が最終的な訓練誤差、テスト誤差ともに低い値に収束しており、より高い性能を示した。混同行列（図 7, 8）を見ても、128 ユニットの方が正解数（対角成分の合計）が多く、分類精度が向上している。これは、ユニット数を増やすことでモデルの表現力が高まり、より複雑なデータの特徴を捉えられたためと考えられる。ただし、ユニット数を過剰に増やすと過学習のリスクも高まるため、モデルの複雑さと汎化性能はトレードオフの関係にある。

#### 3.3 分類性能の評価

最も性能が良かった Adam・128 ユニットの混同行列（図 8）を用いて詳細な分類性能を評価する。テストデータ 150 サンプルに対し、正しく分類されたのは 119 サンプルであり、全体の正解率は約 79.3% であった。誤分類の例を見ると、数字の「9」を「7」と、「2」を「1」と、「8」を「6」と誤認識するケースが見られる。これは、発音における音響的特徴が類似しているため、モデルが混同しやすいことを示しているはずだが、個人的な感覚からするとあまり似た発音ではないので不思議だった。全体としては、Adam を用いることで多くの数字を高精度で分類できていることが対角成分の数値から確認できる。

#### 3.4 学習の安定性

誤差の推移グラフから、学習プロセスの安定性を評価できる。SGD を用いた場合、誤差の値が大きく振動し、学習が不安定である。これに対し、Adam を用いた場合は誤差が安定して単調に減少しており、学習プロセスが非常に安定していることがわかる。特に、訓練誤差とテスト誤差の間に大きな乖離が見られず、両者が連動して減少している点（図 4）は、モデルが過学習に陥ることなく、未知のデータに対する汎化性能を保ちながら学習できていることを示している。この安定性は、Adam の持つ適応的な学習率調整機能が、RNN の学習における勾配の消失・爆発問題を緩和し、安定した勾配の流れを維持した結果と考えられる。