# Provided MLP and LSTM architectures

## Description of MLP

* Input layer: 64 (8\*8 – board size)
* Two hidden layers: 128 and 128 Linear layers
* Output layer: 64 (8\*8 – board size)
* Optimizer: Adam
* Learning rate: 0.001
* Dropout: 0.1
* Epoch: 200 (Early stopping – 20)
* Batch size: 1000
* Len Samples: 1 (one to one)
* Number of parameters: 33088
* The best accuracy on DEV: 14.64%

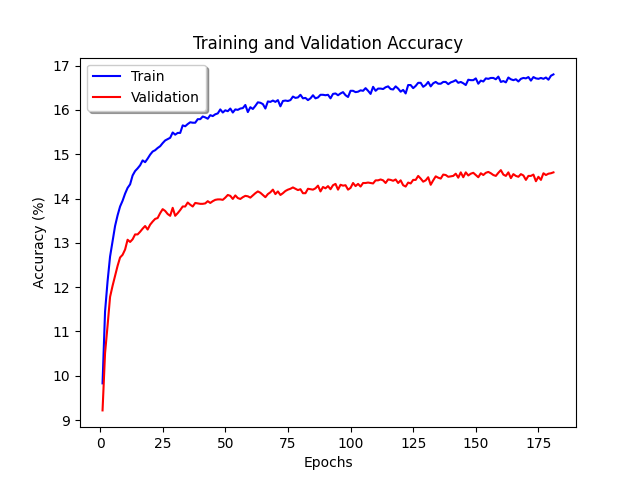
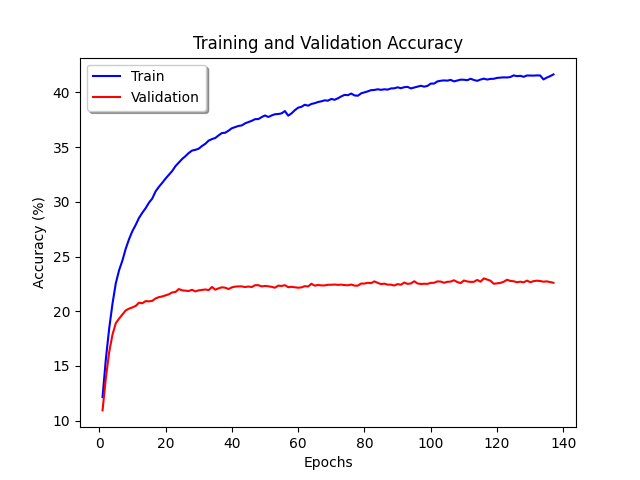
## Description of LSTM

* Input layer: 64 (8\*8 – board size)
* One hidden layer: 128 Linear layer
* Output layer: 64 (8\*8 – board size)
* Optimizer: Adam
* Learning rate: 0.005
* Dropout: 0.1
* Epoch: 200 (Early stopping – 20)
* Batch size: 1000
* Len Samples: 5 (sequence to one)
* Number of parameters: 107584
* The best accuracy on DEV: 23%

|  |  |  |
| --- | --- | --- |
| Architecture | Nombre de paramètres | Précision de Validation Dataset |
| MLP | 33088 | 14.64% |
| LSTM | 107584 | 23% |

Entre deux architectures, il y a une différence importante. MLP a besoin moins de temps de calcul comme le nombre de paramètres est égal à 33088. Ce qui est trois fois moins par rapport au nombre de paramètres de LSTM équivalent à 107584. Alors que LSTM est plus couteux à calculer, cela donne la meilleure précision : 23% contre 14.64% de MLP, ce qui montre que ce modèle est plus performant.

**MLP LSTM**

Dans les graphes suivants, on peut voir la précision de Train et Validation datasets en fonction de Epochs. La contraste entre deux architectures reste crucial. En MLP, la différence d’accuracy entre training dataset et validation dataset est moins dramatique qu’en LSTM. C’est-à-dire, le comportement d’overfitting est plus présent chez LSTM proposé. Donc, dès qu’il rencontre les nouvelles données, LSTM est moins adapté pour se comporter efficacement. Néanmoins, LSTM se compte mieux en général, comme la précision de training set a été élevé, plus de 40% en comparaison de 17% pour MLP.

En comparant, on obtient le tableau suivant :

|  |  |  |  |
| --- | --- | --- | --- |
| Architecture | Nombre de paramètres | Précision de Validation Dataset | Précision de Training Dataset |
| MLP | 33088 | 14.64% | 16.8% |
| LSTM | 107584 | 23% | 41.64% |

# MLP et LSTM optimizing

# Testing different optimizers

Pour tester et comparer les différents optimizers, on choisit les mêmes hyperparamètres que dans le modèle de base :

MLP :

* Input layer: 64 (8\*8 – board size)
* Two hidden layers: 128 and 128 Linear layers
* Output layer: 64 (8\*8 – board size)
* Learning rate: 0.001
* Dropout: 0.1
* Epoch: 200 (Early stopping – 20)
* Batch size: 1000
* Len Samples: 1 (one to one)
* Number of parameters: 33088

LSTM:

* Input layer: 64 (8\*8 – board size)
* One hidden layer: 128 Linear layer
* Output layer: 64 (8\*8 – board size)
* Learning rate: 0.005
* Dropout: 0.1
* Epoch: 200 (Early stopping – 20)
* Batch size: 1000
* Len Samples: 5 (sequence to one)
* Number of parameters: 107584

Et enfin on obtient le tableau suivant :

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Architecture | Precision de Adam | Precision de SGD | Precision de RMSprop | Precision de Adagrad | Precision de Adadelta |
| MLP | 14.64 | 1.58 |  |  | 1.76 |
| LSTM | 23 | 2.36 | 22.29 | 14.54 | 1.64 |

Ici, c’est la précision de DEV.

# Optimizing learning rate

Pour tester et comparer les différents learning rate, on choisit les mêmes hyperparamètres que dans le modèle de base :

MLP :

* Input layer: 64 (8\*8 – board size)
* Two hidden layers: 128 and 128 Linear layers
* Output layer: 64 (8\*8 – board size)
* Optimizer: Adam
* Dropout: 0.1
* Epoch: 200 (Early stopping – 20)
* Batch size: 1000
* Len Samples: 1 (one to one)
* Number of parameters: 33088

LSTM:

* Input layer: 64 (8\*8 – board size)
* One hidden layer: 128 Linear layer
* Output layer: 64 (8\*8 – board size)
* Optimizer: Adam
* Dropout: 0.1
* Epoch: 200 (Early stopping – 20)
* Batch size: 1000
* Len Samples: 5 (sequence to one)
* Number of parameters: 107584

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Architecture | Precision de 0.0001 | Precision de  0.001 | Precision de 0.005 | Precision de  0.01 | Precision de  0.1 | Precision de  1 |
| MLP | 13.13 | 14.64 |  | 7.69 | 2.12 |  |
| LSTM | 14.85 | 18.95 | 23 | 24.39 | 6.65 |  |

Ici, c’est la précision de DEV.

# The impact of different epochs and batch size

# Learning curves

# Result analyses

# Evaluation metric

# CNN

# New data

# Conclusion