Report - US4-t4

Evaluation of the models with MLflow

MLflow is an open-source platform for managing the end-to-end machine learning lifecycle. It is designed to help organizations with various aspects of machine learning, including experimentation, reproducibility, and deployment. MLflow provides tools and components that can be used with any machine learning library and in any programming language.

In our case, we have used the platform for model **evaluation**, in order to compare performance metrics when using different hyperparameters as well as compare performance between models (sequential and factorizer).

Evaluation of the Sequential model:

In this case, we have decided to compare performance metrics modifying a combination of 3 hyperparameters (2 values for each). The hyperparameters that we have modified are:

1) Learning Rate (lr). 0.01 and 0.001
2) L2 : 0 and 0.5
3) Number of iterations (n_iter). 10 and 50

l2	0	0	0.5	0.5	0	0
lr	0.01	0.001	0.001	0.01	0.01	0.001
n_iter	10	10	10	10	50	50

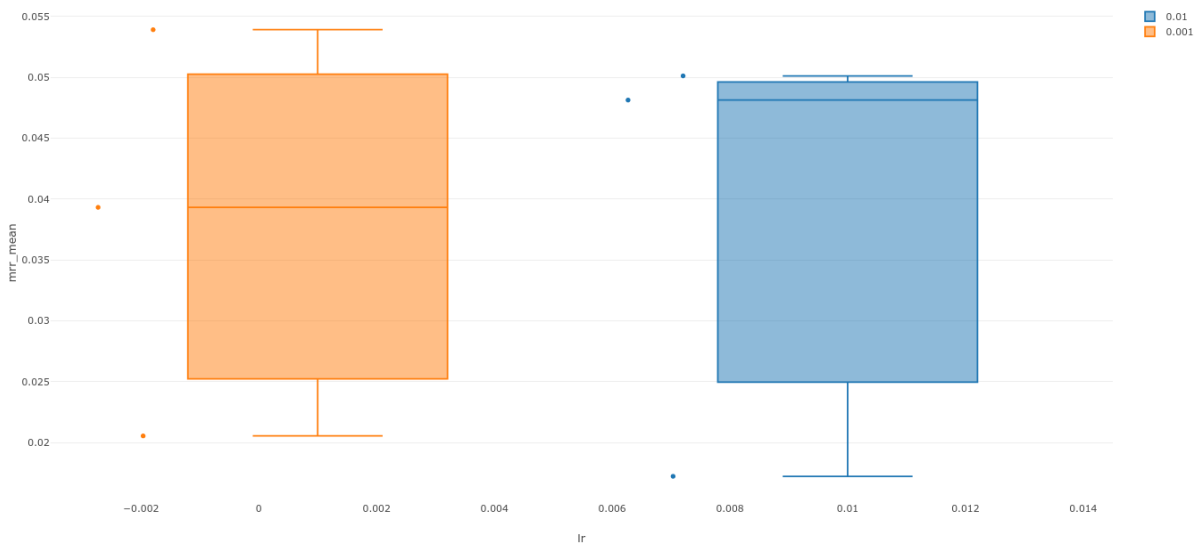
Fig1. Parameter combination

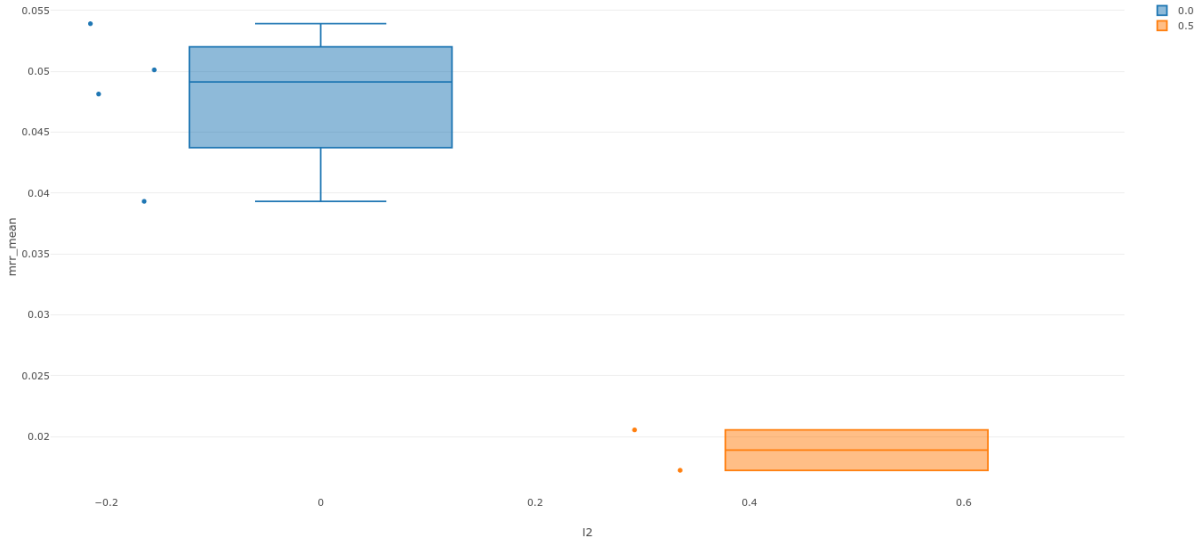
Metrics:

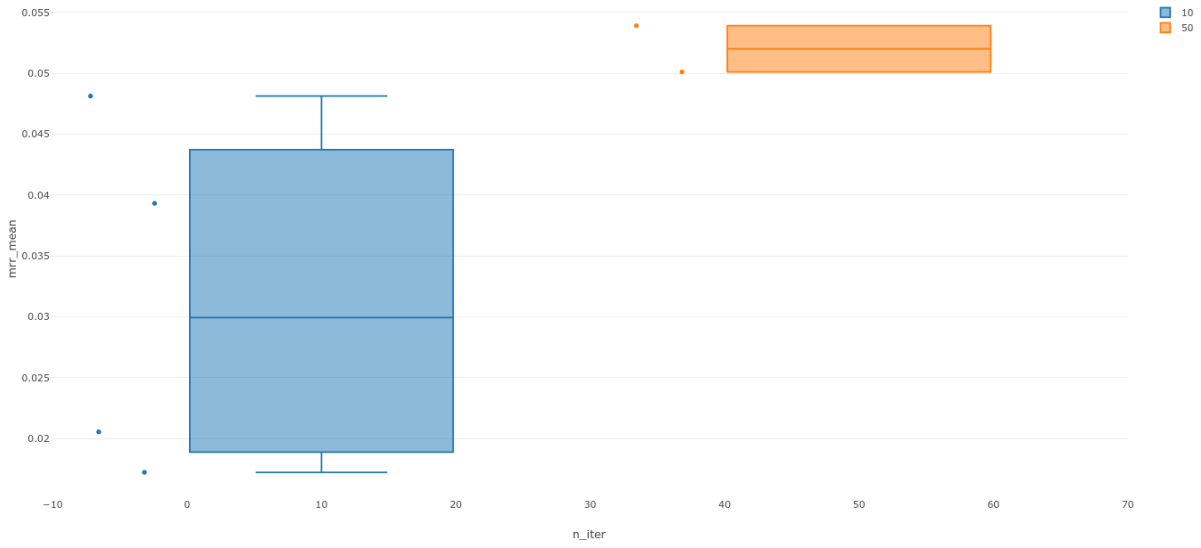
mean_precision	0.052	0.046	0.02	0.02	0.056	0.055
mean_recall	0.052	0.046	0.02	0.02	0.056	0.055
mrr_mean	0.048	0.039	0.021	0.017	0.05	0.054
mrr_std	0.127	0.109	0.085	0.077	0.13	0.147
precision_std	0.105	0.099	0.064	0.063	0.109	0.107
recall_std	0.105	0.099	0.064	0.063	0.109	0.107

Fig2. Performance metrics

Comparison of mrr_mean with the different parameters:







After the evaluation, we have come to the conclusion that the most optimal hyperparameters are lr: 0.01, l2:0 and n_iterations 50.

Evaluation of the Factorizer model:

In this case, we have decided to compare performance metrics modifying a combination of 3 hyperparameters (2 values for each). The hyperparameters that we have modified are:

4) Learning Rate (lr). 0.01 and 0.001
5) L2 ???
6) Number of iterations (n_iter). 10 and 50

l2	0	0	0.5	0.5	0	0
lr	0.01	0.001	0.001	0.01	0.01	0.001
n_iter	10	10	10	10	50	50

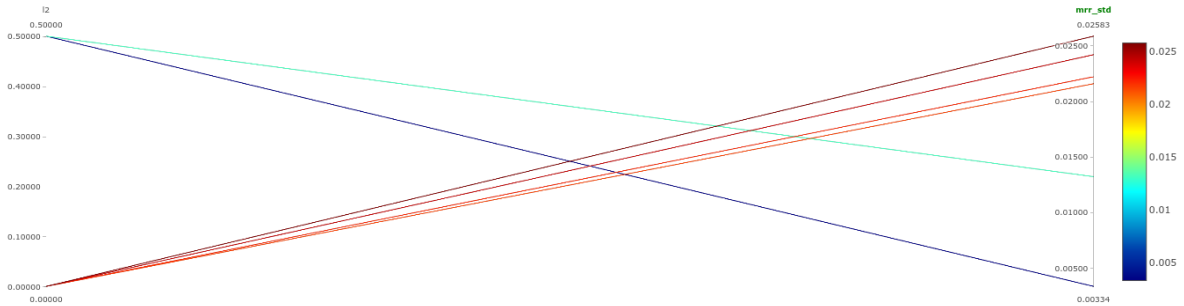
Fig1. Parameter combination

Metrics:

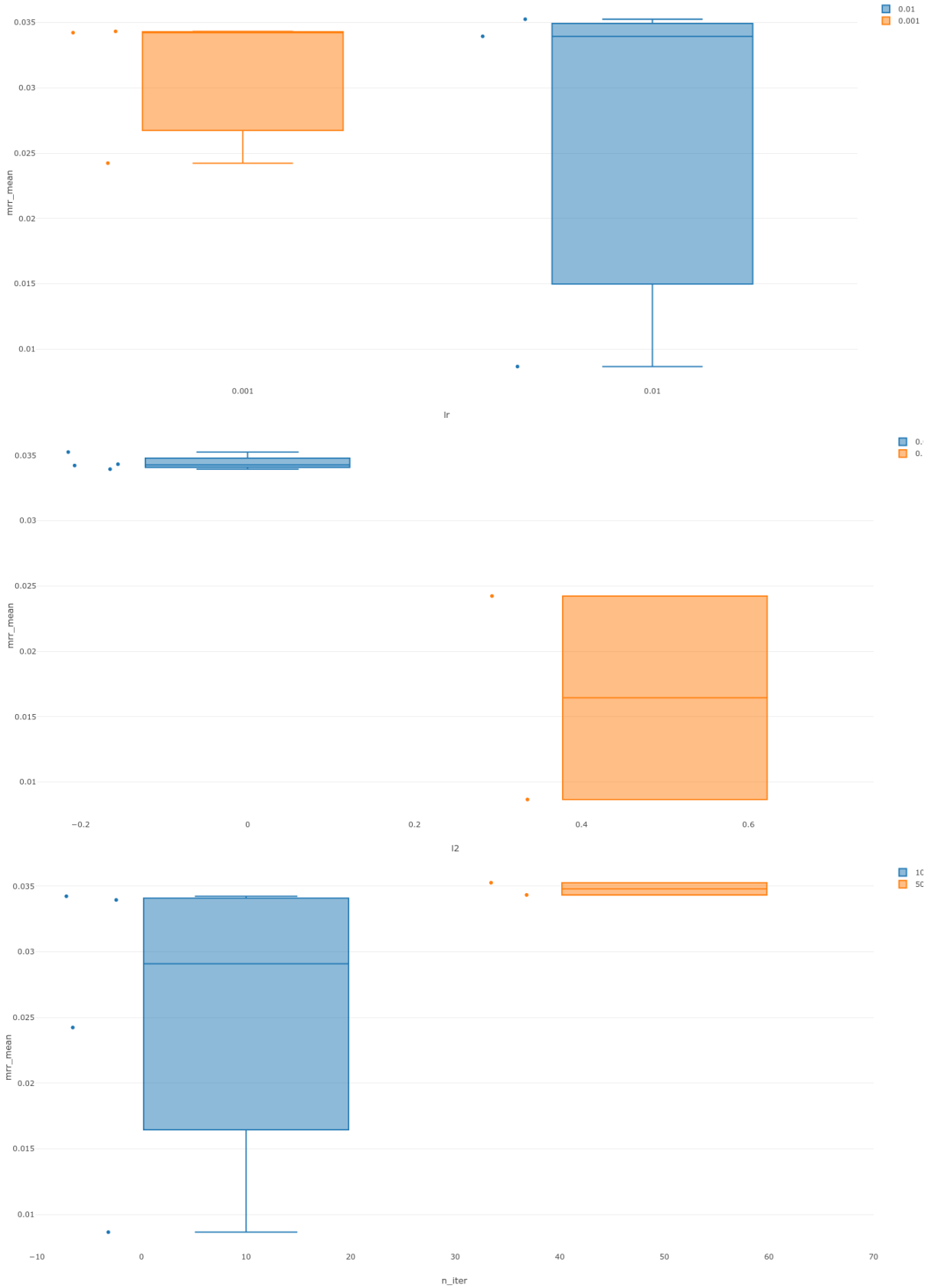
mean_precision	0.444	0.467	0.428	0.137	0.463	0.502
mean_recall	0.071	0.075	0.055	0.01	0.073	0.078
mrr_mean	0.034	0.034	0.014	0.009	0.035	0.034
mrr_std	0.022	0.026	0.013	0.003	0.022	0.024
precision_std	0.208	0.24	0.238	0.178	0.231	0.228
recall_std	0.05	0.069	0.037	0.012	0.052	0.065

Fig2. Performance metrics

In the following plot we can see how higher values of l2 have a worse performance:



Boxplots:


0.01
0.001
0.035
0.03
0.025
0.02
0.015
0.01
mrr_mean
0.001
0.01
lr
0.035
0.03
0.025
0.02
0.015
0.01
mrr_mean
−0.2
0
0.2
0.4
0.6
l2
0.035
0.03
0.025
0.02
0.015
0.01
mrr_mean
−10
0
10
20
30
40
50
60
70
n_iter