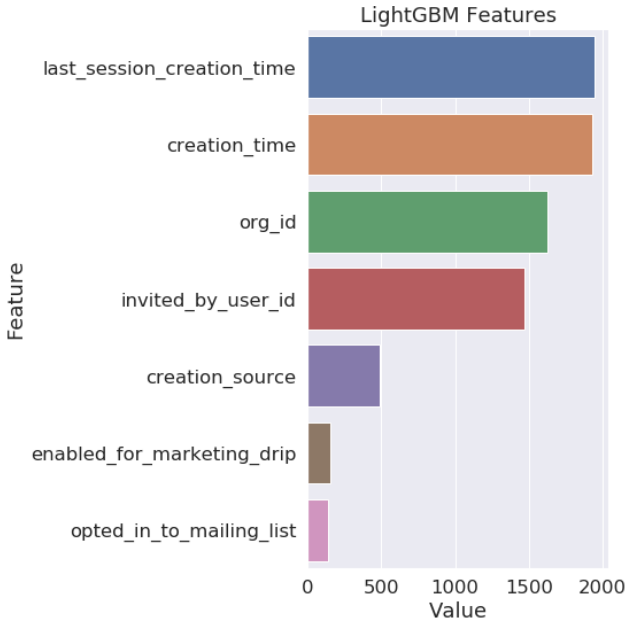
This is a very hard problem, and it would take much, much longer than two hours to do it well. The main issue is that we don't know what are reasonable features to consider. If we consider last_session_creation_time along with all the other user characteristics, then it's possible to predict user adoption with near certainty (accuracy=96% without optimizing hyperparameters). Presumably the causation is running in the opposite direction: adopted users are likely to have recent logins because they have adopted.

If we take that analysis seriously, the feature importances from a LightGBM analysis look like this:



Last session creation time and account creation time are the most important factors. Organization and the inviting user are also important, and creation source is a relevant but relatively less important factor.

We have little information about the meaning of org_id and (inviting) user_id, which are presumably arbitrarily assigned numbers, but it seems that certain ranges or values are predictive of user adoption.

If we eliminate last_session_creation_time from consideration, the relative importances of the other factors are similar, but the predictions (even after optimizing hyperparameters) are not very good. No test users are predicted to adopt with probability 0.5 or greater, and test set ROCAUC is only 0.634.

If we also eliminate creation_time, test set ROCAUC goes down to 0.578.

In an attempt to gain some insight into the relevance of invited_by_user_id, I linked the main data with data for the inviting user and used those data in place of the ID. There was very predictability left, with test set ROCAUC=0.548.

Code:

relax.ipynb: main analysis

relax_no_session_time.ipynb: exclude last_session_creation_time

relax_no_creation_time.ipynb: exclude creation_time

relax_invited_by_user: use characteristics of inviting user in place of ID