Take-home Challenge 2: Relax Inc.

by Chantel Clark

A logistic model was chosen to determine which factors predict future user adoption, since coefficients from a logistic model can help us to determine the effect of each factor. A L1 regularization (LASSO) penalty eliminated the less important factors in order to select the most influential predictors. The factors that were considered in this model included: creation source, opted in to mailing list, enabled for marketing drip, organization belonging to, invited by an adopted user, active time (time between last session and account creation), and year account was created.

The first test run gave an accuracy score of 97.3%, with the best predictor being active time ('tdelta'). This makes sense that the longer a user is active, the more likely the user is to become an adopted user (coefficient 0.06346). The next more important factors in this model (and their associated coefficients) are shown in the table below.

feature	coefficients
tdelta	0.063460
invited_by_adopted_user	0.000319
opted_in_to_mailing_list	0.000140
creation_source_GUEST_INVITE	0.000087
org_id_30	0.000048

The active time ('tdelta') coefficient is about 100 times larger than the others, so the test was re-run without active time to see the effects of the other predictors. The next test run had an accuracy of 86.5%, but the model returned no true positives or false positives (it did badly on predicting an adopted user). This model is clearly flawed but could potentially give information about the relationship of factors outside of active time ('tdelta').

feature	coefficients
invited_by_adopted_user	0.171478
creation_source_GUEST_INVITE	0.155092
creation_source_SIGNUP_GOOGLE_AUTH	0.093291
creation_source_SIGNUP	0.039643
enabled_for_marketing_drip	0.019794

In summary, the best predictor was active time (tdelta), followed by invitation by an adopted user. Another relatively strong predictor was creation source - the accounts created by guest invite were most likely to become adopted users, while accounts created through google authentication and signup via website have more potential of becoming adopted users than accounts created by organization invite or for personal projects. Taking both models into consideration, opting into the mailing list and enabling marking drip have relatively low predictive power compared with the other factors.