Work done

The week has been spent in a last effort to obtain better results for the networks and to solve the various convergence issues.

Firstly, the learning rate has been adjusted from 1 to 0.01.

Accuracy	word2vec train	word2vec test	glove train	glove test	fasttext train	fasttext test
Dense	0.7407158017158508	0.707619845867157	0.7455262541770935	0.7277397513389587	0.7578410506248474	0.6917808055877686
CNN	0.8091206550598145	0.7315924763679504	0.794111967086792	0.704623281955719	0.7107946872711182	0.7324486374855042
LSTM	0.8508754968643188	0.7140411138534546	0.864537239074707	0.7260273694992065	0.8509716987609863	0.6930650472640991

Accuracy	word2vec train	word2vec test	glove train	glove test	fasttext train	fasttext test
Dense	0.7107946872711182	0.7324486374855042	0.7107946872711182	0.7324486374855042	0.7107946872711182	0.7324486374855042
CNN	0.7107946872711182	0.7324486374855042	0.7107946872711182	0.7324486374855042	0.7107946872711182	0.7324486374855042
LSTM	0.8144121766090393	0.6720890402793884	0.8173946738243103	0.616866409778595	0.813450038433075	0.6750856041908264

Contrary to what was expected, a lower learning rate produced worse results.Next, we attempted with an SGD optimizer, without any major improvements.

Accuracy	word2vec train	word2vec test	glove train	glove test	fasttext train	fasttext test
Dense	0.7107946872711182	0.7324486374855042	0.7107946872711182	0.7324486374855042	0.7107946872711182	0.7324486374855042
CNN	0.7107946872711182	0.7324486374855042	0.7107946872711182	0.7324486374855042	0.7107946872711182	0.7324486374855042
LSTM	0.7107946872711182	0.7324486374855042	0.7126226425170898	0.7311643958091736	0.7107946872711182	0.7324486374855042

An additional embedding was added, a word2vec trained on wikipedia and biomedical text in the hopes that it would produce better results. There was no discernible difference.

Accuracy	word2vec train	word2vec test	word2vecMed train	word2vecMed test	glove train	glove test	fasttext train	fasttext test
Dense	0.775351167251587	0.6288527250289917	0.7500481009483337	0.639126718044281	0.7565903663635254	0.7033390402793884	0.7710217237472534	0.6699486374855042
CNN	0.7107946872711182	0.7324486374855042	0.7107946872711182	0.7324486374855042	0.7107946872711182	0.7324486374855042	0.7107946872711182	0.7324486374855042
LSTM	0.9914373755455017	0.6605308055877686	0.987877607345581	0.6635273694992065	0.9931691288948059	0.6797945499420166	0.9882624745368958	0.5984588861465454

Given that the convnet we have been using has no variance in accuracy results regardless of what input, learning rate, and optimizer we implemented, we decided to drop it, and focused on the dense and LSTM networks, producing variants with more neurons and dropouts, respectively.

Accuracy	word2vecMed train	word2vecMed test	glove train	glove test	fasttext train	fasttext test
Dense	0.8155666589736938	0.6613869667053223	0.8225899338722229	0.6626712083816528	0.8018087148666382	0.6207191944122314
Really Dense	0.7107946872711182	0.7324486374855042	0.7107946872711182	0.7324486374855042	0.7107946872711182	0.7324486374855042
LSTM	1.0	0.6258561611175537	1.0	0.6879280805587769	1.0	0.65625
Dropout LSTM	1.0	0.6571061611175537	1.0	0.6459760069847107	0.9996151328086853	0.6515411138534546

It is clear the LSTM heavily overfits, even when we introduce dropouts. We therefore worked with two more variants of dense, adding one or two layers of heavier neurons. We also focused exclusively on the biomedical word2vec and glove embeddings, as the fasttext performed worse.

Accuracy	word2vecMed train	word2vecMed test	glove train	glove test
Dense	0.7932460904121399	0.6618150472640991	0.7107946872711182	0.7324486374855042
Really Dense	0.7107946872711182	0.7324486374855042	0.7107946872711182	0.7324486374855042

The heavier dense networks have the most accuracy, and some variant will probably be used in the final project.

Problems

Even with significant efforts, the best accuracy obtained is 73%. It is entirely likely that we will not produce better results on this particular dataset. As such, we will use dense network variants as a concept of proof going forward with the project, as the deadline approaches.

To do for next week:

The next sprint will be focused on Django and building the actual tool interfaces for the website. In particular, the next week will be spent on Django familiarization.