3. COMPARE LOGISTIC REGRESSION TO CRFs

3.1 ANALYTICS

From my understanding, CRF will perform better than logistic regression for the sequence tagging problem.

The primary reasons why CRF would perform better are as follows:

- CRF being a discriminative model, does not assume that the features are independent of one another.
- The highlight of the CRF model is that it makes predictions by taking context into account. That is, while making a prediction for a particular word, it also takes into account the labels of the surrounding words, the previous and the future words. For instance, the Viterbi algorithm while predicting the best tag sequence for a sentence of 'n' words, to predict the best tag for the nth word, it recursively computes the best tag sequence till word (n – 1) using the transition probabilities and emission probabilities.
- The CRF used here is a linear chain CRF, which essentially predicts a sequence of labels for a sequence of words. Since CRF is a discriminative model, it makes use of conditional probability.
- Also, it makes use of global normalization.
- Also, in CRF, the features for the model play a very critical role in its performance. Hence, a lot of the performance can be tuned based on the features.

On the contrary, Logistic Regression converts a sequence labelling task into word independent classification, wherein, the tag assigned to each word is independent of the context.

Hence, as CRF keeps a track of context and does not assume features to be independent, it performs better then Logistic Regression.

Example:
Consider the CRF and Logistic Regression output for sentence:

```
RT  O
@YahooDrSaturday   O
:  O
This   O
is O
how    O
Arkansas   B-sportsteam
crazily    O
converted  O
4th    O
and    O
25 O
in O
OT O
.  O
What   O
a  O
lateral    O
!  O
https://t.co/ylALEACWe8    O
```

CRF Output:

```
RT      O       O
```

@YahooDrSaturday O O
: O O
This O O
is O O
how O O
Arkansas B-sportsteam B-geo-loc
crazily O O
converted O O
4th O O
and O O
25 O O
in O O
OT O O
. O O
What O O
a O O
lateral O O
! O O
https://t.co/ylALEACWe8 O O

Logistic Regression Output:

```
RT 0  0
@YahooDrSaturday   0  0
:  0  0
This   0  0
is 0  0
how    0  0
Arkansas   B-sportsteam   0
crazily    0  0
converted  0  0
4th    0  0
and    0  0
25 0  0
in 0  0
OT 0  0
.  0  0
What   0  0
a  0  0
lateral    0  0
!  0  0
https://t.co/ylALEACWe8    0  0
```