**KL Divergence based ranking**

Kullback - Leibler divergence is an asymmetric measure of similarity between 2 probability distributions. Given 2 movies A and B, we first calculate their topic posterior, *\_A* and *\_B* and then combine and normalize the two. This is done by using one of the following strategies:

**Max merge**

*C* = *max* (*\_A; \_B*)

**Average merge**

*C* = (*\_A* + *\_B*)*=*2

*C*, the combined topic distribution, is finally normalized.

For all the movies, *i*, *KL Divergence* (*\_C; \_i*) is calculated. The movies rankings in the results are inversely proportional to this divergence score.

**Our Idea:**

The motivation for this method was that the movies similar to both the movies will have common top topics. For example, if the two movies are about [war, life struggle] and [life struggle, romance], (i.e. their *\_* posterior has high values corresponding to these topics) then returning movies that are about life struggle might be a good idea. For both the movies, the top common topics were chosen. The words were then given to an inverted index which returned the words corresponding to these movies. The returned set of movies were then ranked based on their topic posteriors, the movies for which the top common topics had high stakes were placed above those for which these numbers were lower.