**Section 4: Bonus**

* **Part 2: LDA and SVM data:**

We have tried two ideas for LDA and SVM part. We compare another acquisition function namely Probability of improvement (PI) with EI and random search. Also, we relearn the hyperparameters after every observation by maximizing the marginal likelihood and compare the performance with our current model (EI based Bayesian method) and random search.

**Probability of improvement (PI):**

The following figure shows the comparison among PI, EI and random search on LDA and SVM datasets.



Both PI and EI have similar performance. However, as we can see from the left figure above, PI has slightly worse performance between observation 5 and 10 on LDA datasets. As PI only considers improvement but not the size of the improvement, probably in this region (between observation 5 and 10 on LDA dataset) it falls in a slocal optimum which it eventually overcome with the increase number of observations. Overall, both PI and EI performs much better than random policy.

**Relearning the model’s hyperparameters:**

The following figures show the learning curves for this experiment.



We performed this experiment using EI acquisition function. We thought if we relearn the hyperparameters after every observation the performance would increase. From the above figures we see that this hypothesis is almost true. On SVM data we are seeing that relearning after every observation (orange curve) slightly outperforms the default one which we denote as “Not Relearning Bayesian” (blue curve). But for LDA data we cannot conclude it as sometimes orange line is higher (number of observations around 10 to 20) and sometimes blue line is higher (number of observations around 23 to 30). We think this behavior depends on two things. First, how is the initial model? The BIC score of the model for LDA dataset (BIC score -115.41) was much better than SVM (BIC score -47.06). We think if the initial model is good enough, the relearning idea may not provide much improvement as the model is well informative already. Second, how much data were used to train the initial model? For LDA whole dataset consists of only 288 observations. And initially we trained it with 32 observations. So, the initial model is based on 11.1% data. On the other hand, the initial model for SVM dataset is based on 2.3% data (32/1400\*100). As a result, the model for SVM has more scope to improve with relearning than LDA which we are seeing in the above figure.