**Section 3: Bayesian optimization**

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

Learning curves for LDA and SVM data are given below. Bayesian optimization using EI acquisition function is performing better than random search on both data set.



Next we compute the mean gap for EI and for random search presented in the following table:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Bayesian optimization (EI acquisition function) | Random search (30 observations) | Random search (60 observations) | Random search (90 observations) | Random search (120 observations) | Random search (150 observations) |
| LDA data | 0.93 | 0.56 | 0.84 | 0.86 | 0.96 | 0.96 |
| SVM data | 0.88 | 0.66 | 0.76 | 0.8 | 0.82 | 0.83 |

The corresponding p values from paired t-test are:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Random search (30 observations) | Random search (60 observations) | Random search (90 observations) | Random search (120 observations) | Random search (150 observations) |
| LDA data | 0.0009 | 0.05 | 0.1 | \*0.65 | \*0.60 |
| SVM data | 0.01 | 0.06 | 0.88 | 0.29 | 0.37 |

\*random search performs better for >=120 observations on LDA data. So the (\*) p-value represents the p-value where random search is better than LDA. Note that this is not the case for SVM data.

For both LDA and SVM data p value becomes 0.05 around 60 observations of random samples.

Interestingly for LDA data random search outperforms Bayesian method when the random observations become 120. But this is not the case for SVM. We have found out that SVM dataset has 1400 observations in total. Whereas LDA has only 288. When random search is provided with 120 observation, random search is including 41.67% data. We think that is the reason better performance of random search on LDA data, but not on SVM data.