EP2420 *Project 1 - Week 2*

Federico Giarrè

November 9, 2023

2.2 Study the Relationship between Estimation Error and the Size of the Training Set

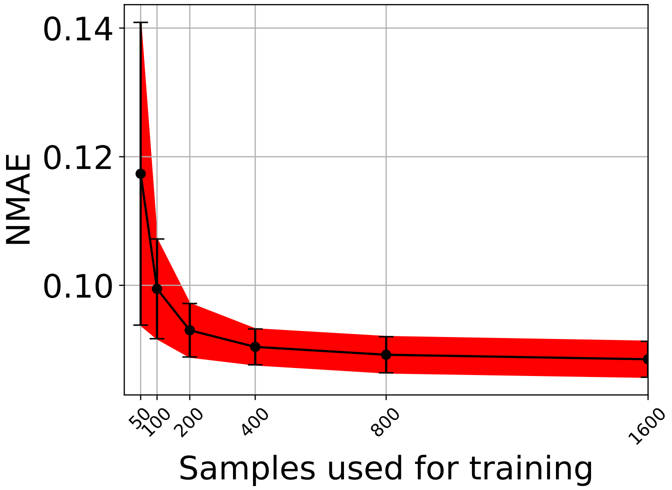

Figure 1: Normalized Mean Absolute Error with respect to an increasing Training Set size

In order to grasp the relationship between the estimation error and the size of the training set, different linear regression models were trained using Training Sets of different sizes. In particular, over the full 3600 samples dataset, 1000 samples were extracted to be used as Test Set, and models were trained with 50, 100, 200, 400, 800 and 1600 samples respectively. This process has been repeated 50 times in order to catch how the estimation error (expressed as NMAE), varies with different Training Set sizes, In the plot presented in Fig. 1, the NMAE is shown for the different Training Set sizes, where the black line indicates the average of the NMAEs for each sample size and the red area and error bars indicate the standard deviation of the values. As it is possible to notice, increasing the size of the training set decreases the mean NMAE. Additionally, a trend that is noticeable from the plot is that the standard deviation is very high at small sample sizes, but quickly stabilize and remain the same after the 400 sample mark.

2.3 Quiz on time complexity

The time complexity of training a Linear Regression model is $O(p^2n+p^3)$[1], where p is the number of features and n is the number of samples. With this complexity in mind, it is possible to calculate the computational time complexity of both the optimal and heuristic methods. The optimal method requires finding every possible subset of feature and training a model on top of that subset. Since the number of subsets of a set of p features is 2^p, we can conclude that the time complexity of the optimal method is $O(2^p(p^2n+p^3))$. On the

[1] https://www.thekerneltrip.com/machine/learning/computational-complexity-learning-algorithms/

other hand, the heuristic method requires to compute the Pearson Correlation Coefficient (PCC) between each feature and the label and training 12 models as described in the Project description. In order to do that we'll need to compute the PCC (that has a complexity of $O(p)$) one time for each feature, and then train the p models, resulting in a complexity of $O(p + p(p^2n + p^3))$. In this specific case, the usage of the heuristics over the optimal is roughly 341 times better[2],finding the difference in number of models trained.

3 Reduce the Number of Device Statistics to Estimate the Service Metric

3.1 Optimal Method

As mentioned in the previous Section , finding the best combination of features through the optimal method requires training one model for each possible subset of features. Finally, by computing the NMAE for each model, it is possible to see which model performed better and thus which feature combination results in better predictions. After **43.11** seconds of training and evaluation, the optimal method's result is that the best model has a NMAE of **9e-01**, with the used features being **(runq-sz,%%memused, proc/s, cswch/s, all_%%usr, pgfree/s, plist-sz, file-nr, idel/s, tps)** and thus discarding the ldavg-1 and totsck.

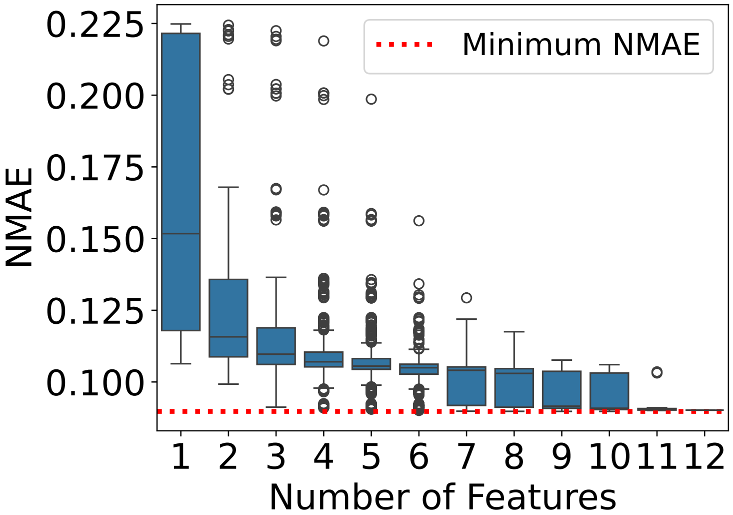

Figure 2: Distribution of NMAE results with respect to the number of features used

In Fig. 2 the distribution of NMAE values are plotted as a box plot, where each box represents the NMAE results for a certain number of features used in training. A dotted line is then plotted to show the minimum NMAE reached by the upper-cited model. The plot shows some interesting results that can be summed up as follows: *i)* Models trained on one or two features can not reach near the optimal result; *ii)* Models trained with three, four, five, and six features can reach as low as close to the optimal NMAE, but only models trained on three features reach near that mark with the lower whiskers, while the others can reach only with outliers; *iii)* Models trained on seven and eight features can reach near the minimum NMAE with their Q1 (and thus with 25% of values); *iv)* Models trained on nine or ten features reach near the minimum NMAE with almost 50%, since their medians barely touch it. One of the models trained with 10 features is the one producing the smallest NMAE;*v)* Finally, models trained on eleven and twelve features obtain a very narrow box (or a line in the case of the twelve features one), with values standing just slightly above the optimal model's NMAE (apart from an outlier).

[2] 4356300 operations instead of 1486946304, heuristic performs 341x times faster. This can be found also by performing $\frac{4096}{(12+\frac{12}{364608})}$