

# Evaluation Criteria

Louis Schlessinger

February 2019

## 1 Introduction

The following are the evaluation criteria used to assess the performance of the automated kernel search process. They include three broad areas: model score, model complexity, and population diversity. Each one is tracked over time and presented as a time-series.

## 2 Model Score

These evaluation criteria measure the performance of a single model. E.g. negative log marginal likelihood, Bayesian information criterion, etc.

### 2.1 Best-so-far

The maximum model score for each generation.

### 2.2 Average Score

The mean model score for each generation.

### 2.3 Score Confidence

The sample standard deviation of model scores for each generation.

## 3 Model Complexity

The model complexity over time shows how the model space is explored and which models in that space perform the best. The mean and standard deviation of each is recorded at each time step.

### 3.1 Number of Hyperparameters

The number of kernel hyperparameters used. E.g., the squared-exponential covariance function in one dimension,  $k_y(x_p, x_q) = \sigma_f^2(-\frac{1}{2\ell^2}(x_p - x_q)^2) + \sigma_n^2\delta_{pq}$ , has 3 hyperparameters  $(\sigma_f, \ell, \sigma_n)$ . This includes parameters such as length scale, period and variance. This shows the evolution of the types of kernels explored over time and which types of kernels performed the best. It can be thought of as a model complexity score of the kernel structure.

### 3.2 Number of Operands

The number of 1-D sub-kernels used in the full kernel. E.g.  $SE_0 + RQ_1 \cdot SE_1$  has 3 operands. This is strongly correlated with the number of model hyperparameters and can also be thought of as a model complexity score.

## 4 Population Diversity

These heterogeneity scores measure distance between pairs of kernels. The mean and standard deviation of each is recorded at each time step.

### 4.1 Kernel Expression Distance

The distance between kernels in algebraic form. The kernel expression distance measures the similarity based on the algebraic expression of the kernel in additive form. To compute this population diversity score, we first take the mean distance between all pairs of kernels (for each generation). To compute the distance between a pair of kernels  $k_i$  and  $k_j$ , we first convert each on to additive form. Then, we vectorize each product within each additive kernel using a simple label encoding. Lastly, the mean Euclidean distance between the two additive kernels' parts are calculated for all possible pairs of products.

Example:

$$k_i = SE_0 + RQ_1(RQ_1 + SE_1)$$

$$k_j = SE_1SE_1 + SE_1RQ_1(RQ_0 + RQ_1) + RQ_0$$

1. Convert to additive form

$$k_i = SE_0 + RQ_1RQ_1 + RQ_1SE_1$$

$$k_j = SE_1SE_1 + SE_1RQ_1RQ_0 + SE_1RQ_1RQ_1 + RQ_0$$

2. Encode additive kernels (assume  $d = 2, \mathcal{B} = \{SE, RQ\}$ )

$$k_i \implies L_i = \left\{ \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} \right\}$$

$$k_j \implies L_j = \left\{ \begin{bmatrix} 0 \\ 0 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \right\}$$

3. Take Cartesian product of all pairs of additive kernels

$$C_{ij} = L_i \times L_j = \{(\ell_i, \ell_j) \mid \ell_i \in L_i, \ell_j \in L_j\}$$

4. Compute mean Euclidean distance of all Cartesian products

$$d_{ij} = \|\ell_i - \ell_j\|_2$$

$$D_{ij} = \frac{1}{|C_{ij}|} \sum_{k=1}^{|C_{ij}|} d_{ij}$$

5. Compute mean  $\bar{D}$  and standard deviation  $\sigma_D$  of  $D$

## 4.2 Covariate Distance

The  $\ell_2$  (Euclidean) distance between a pair of kernel matrices. It can be thought of as a naive similarity score between different explanations of the data offered by each kernel. The distribution of Euclidean distances at each generation can be thought of as population heterogeneity score. For example, suppose we wish to compute the distance matrix  $D \in \mathbb{R}^{n \times n}$ :

$$D_{ij} = \|k_i(X, X) - k_j(X, X)\|_2$$

$$D = \frac{D + D^\top}{2}$$

We can then compute the mean and standard deviation of  $D$ .

$$\bar{D} = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n D_{ij}$$

$$\sigma_D = \sqrt{\frac{\frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n (D_{ij} - \bar{D})^2}{n^2 - 1}}$$

Equivalently, we can vectorize  $D$ : let  $\vec{v} = \text{vec}(D) \in \mathbb{R}^{N \times 1}$ . Then,

$$\bar{v} = \bar{D} = \frac{1}{n^2} \sum_{i=1}^{n^2} v_i = \frac{1}{N} \sum_{i=1}^N v_i$$

$$\sigma_v = \sigma_D = \sqrt{\frac{\sum_{i=1}^{n^2} (v_i - \bar{D})^2}{n^2 - 1}} = \sqrt{\frac{\sum_{i=1}^N (v_i - \bar{v})^2}{N - 1}}$$

where  $N = n^2$