

> Появляется что-то наподобие **knn** или **парзеновского окна** (стр 2)

plained. A promising way for understanding their complex mechanisms is to study the connection between forests and kernel estimates, that is estimates m_n which take the form

$$m_n(\mathbf{x}) = \frac{\sum_{i=1}^n Y_i K_k(\mathbf{X}_i, \mathbf{x})}{\sum_{i=1}^n K_k(\mathbf{X}_i, \mathbf{x})}, \quad (1)$$

where $\{(\mathbf{X}_i, Y_i) : 1 \leq i \leq n\}$ is the training set, $(K_k)_k$ is a sequence of kernel functions, and k ($k \in \mathbb{N}$) is a parameter to be tuned. Unlike the most used Nadaraya-Watson kernels ([Nadaraya, 1964](#); [Watson, 1964](#)) which satisfy a homogeneous property of the form $K_h(\mathbf{X}_i, \mathbf{x}) = K((\mathbf{x} - \mathbf{X}_i)/h)$, kernels K_k are not necessarily of this form. Therefore, the analysis of kernel estimates defined

> Plan

The paper is organized as follows. Section 2 is devoted to notations and to the definition of KeRF estimates. The link between KeRF estimates and random forest estimates is made explicit in Section 3. In Section 4, two KeRF estimates are presented and their consistency is proved along with their rate of consistency. Section 5 contains experiments that highlight the good performance of KeRF compared to their random forests counterparts. Proofs are postponed to Section 6.

> Briemann, Uniform and Centered Trees:

results: Breiman's, centred and uniform forests. In Breiman's original procedure, splits are performed to minimize the variances within the two resulting cells. The algorithm stops when each cell contains less than a small pre-specified number of points (typically between 1 and 5; see [Breiman, 2001](#), for details). Centred forests are a simpler procedure which, at each node, uniformly select a coordinate among $\{1, \dots, d\}$ and performs splits at the center of the cell along the pre-chosen coordinate. The algorithm stops when a full binary tree of level k is built (that is, each cell is cut exactly k times), where $k \in \mathbb{N}$ is a parameter of the algorithm (see [Breiman, 2004](#), for details on the procedure). Uniform forests are quite similar to centred forests except that once a split direction is chosen, the split is drawn uniformly on the side of the cell, along the preselected coordinate (see, e.g., [Arlot and Genuer, 2014](#)).

> Рассмотрим обычный RF

To be more specific, random forest estimates satisfy, for all $\mathbf{x} \in [0, 1]^d$,

$$m_{M,n}(\mathbf{x}, \Theta_1, \dots, \Theta_M) = \frac{1}{M} \sum_{j=1}^M \left(\sum_{i=1}^n \frac{Y_i \mathbb{1}_{\mathbf{X}_i \in A_n(\mathbf{x}, \Theta_j)}}{N_n(\mathbf{x}, \Theta_j)} \right),$$

where $A_n(\mathbf{x}, \Theta_j)$ is the cell containing \mathbf{x} , designed with randomness Θ_j and data set \mathcal{D}_n , and

$$N_n(\mathbf{x}, \Theta_j) = \sum_{i=1}^n \mathbb{1}_{\mathbf{X}_i \in A_n(\mathbf{x}, \Theta_j)}$$

is the number of data points falling in $A_n(\mathbf{x}, \Theta_j)$. Note that, the weights $W_{i,j,n}(\mathbf{x})$ of each observation Y_i defined by

$$W_{i,j,n}(\mathbf{x}) = \frac{\mathbb{1}_{\mathbf{X}_i \in A_n(\mathbf{x}, \Theta_j)}}{N_n(\mathbf{x}, \Theta_j)}$$

Есть проблема: веса при объектах в малочисленных клетках больше чем в клетках с большим числом объектов: это нелогично вот пример

>> Пример:

rough estimates. Indeed, as an extreme example, trees of non adaptive forests can contain empty cells which leads to a substantial misestimation (since the prediction in empty cells is set, by default, to zero).

>> Чтобы это исправить можно например написать вот так:

In order to improve the random forest methods and compensate the misestimation induced by random forest weights, a natural idea is to consider KeRF estimates defined, for all $\mathbf{x} \in [0, 1]^d$, by

$$\tilde{m}_{M,n}(\mathbf{x}, \Theta_1, \dots, \Theta_M) = \frac{1}{\sum_{j=1}^M N_n(\mathbf{x}, \Theta_j)} \sum_{j=1}^M \sum_{i=1}^n Y_i \mathbb{1}_{\mathbf{x}_i \in A_n(\mathbf{x}, \Theta_j)}. \quad (3)$$

>> теперь каждое наблюдение одного веса !! это будет называться **KeRF**

> Оценки получаемые KeRF можно записать в ядерном виде:

$$\tilde{m}_{M,n}(\mathbf{x}, \Theta_1, \dots, \Theta_M) = \frac{\sum_{i=1}^n Y_i K_{M,n}(\mathbf{x}, \mathbf{X}_i)}{\sum_{\ell=1}^n K_{M,n}(\mathbf{x}, \mathbf{X}_\ell)},$$

$$K_{M,n}(\mathbf{x}, \mathbf{z}) = \frac{1}{M} \sum_{j=1}^M \mathbb{1}_{\mathbf{z} \in A_n(\mathbf{x}, \Theta_j)}.$$

This connection function can be seen as a geometrical characteristic of the cells in the random forest. Indeed, fixing \mathbf{X}_i , the quantity $K_{M,n}(\mathbf{x}, \mathbf{X}_i)$ is nothing but the empirical probability that \mathbf{X}_i and \mathbf{x} are connected (i.e. in the same cell) in the M finite random forest.

> Затем они говорят какой смысл будет иметь ядро при бесконечном числе деревьев

$K_n(\mathbf{x}, \mathbf{z}) = \mathbb{P}_\Theta [\mathbf{z} \in A_n(\mathbf{x}, \Theta)]$. Denote by \mathbb{P}_Θ the probability with respect to Θ ,

> По поводу близости оценки RF с KeRF

It is worth noticing that controlling the number of observations in each cell while obtaining a simple partition shape is difficult to achieve. On the one hand, if the tree construction depends on the training set, the algorithm can be stopped when each leaf contains exactly one point and thus KeRF estimate is equal to random forest estimate. However, in that case, the probability $K_n(\mathbf{x}, \mathbf{z})$ is very difficult to express since the geometry of each tree partitioning strongly depends on the training set. On the other hand, if the tree construction is independent of the training set, the probability $K_n(\mathbf{x}, \mathbf{z})$ can be made explicit in some cases, for example for centred forests (see Section 5). However, the number of points in each cell is difficult to control (every leaf cannot contain exactly one point with a non-adaptive cutting strategy) and thus KeRF estimate can be far away from random forest estimate. Consequently, one cannot deduce an explicit expression for random forest estimates from the explicit expression of KeRF estimates.

> В итоге вот краткий план этой статьи:

- 1) Вводим что такое KeRF
- 2) говорим о теоретической близости KeRF и RF (но только если мы не контролируем число объектов в звене?)
- 3) Говорим о explicit виде K для Uniform и Centered виде KeRF
- 4) эксперименты на модельных данных которые не показывают особо результатов

5) Очень много доказательств