

Разработка эффективных методов построения  
ансамблевых регрессионных линейных моделей,  
основанных на максимизации корреляции с откликом

Борисов Иван Максимович  
д.ф.-м.н. Сенько Олег Валентинович

Строим d МНК линейных регрессий:  $R_i = b_i + w_i x_i$ ,  $\bar{R} = (R_1, \dots, R_d)$ ,  $P(\bar{c}, \bar{R}) = \sum_{i=1}^d c_i R_i$

Хотим  $c^*$ :  $\forall \bar{c} = (c_1, \dots, c_d) : \sum_{i=1}^d c_i = 1, c_i \geq 0 \Rightarrow \rho(P(\bar{c}^*, \bar{R}), y) \geq \rho(P(\bar{c}, \bar{R}), y)$ ,

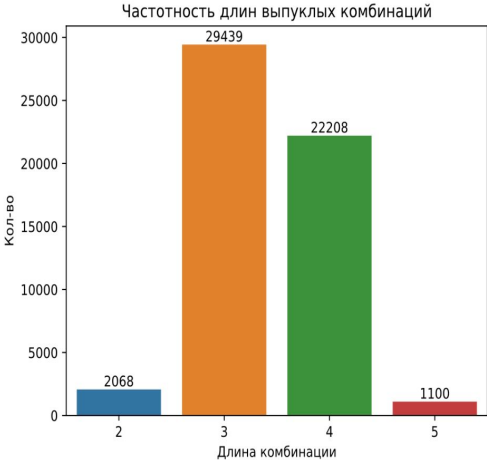
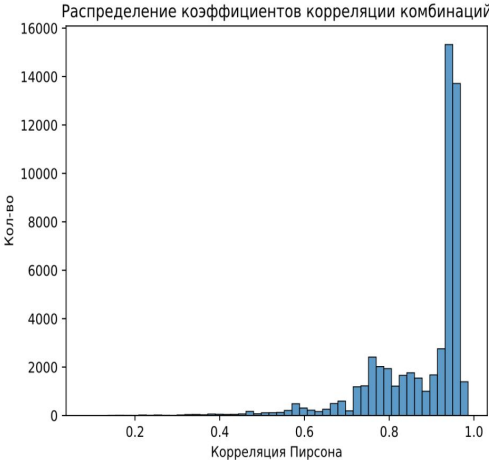
Фиксируем множество индексов  $I = \{1, \dots, l\}$ , решаем:

$$\rho(Y, P(\bar{c}, \bar{R})) = \frac{\sum_{i=1}^l c_i \mathbb{D}R_i}{\sqrt{\mathbb{D}Y} \sqrt{\sum_{i=1}^l c_i \mathbb{D}R_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l c_i c_j \varrho(R_i, R_j)}} = \frac{\theta}{\sqrt{\mathbb{D}Y} \sqrt{\theta - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l c_i c_j \varrho(R_i, R_j)}} \rightarrow \max_{\theta}$$

$\exists l \in \{2, \dots, d\}$ : полученная комбинация несократима и нерасширяема

Ищем несократимые и нерасширяемые комбинации  $\forall \{i_1, i_2\}$

Итоговое предсказание строим как “среднее” по всем полученным комбинациям



Модель	$r^2$	Корреляция Пирсона
ВПК <sub>ср</sub>	0.8993	0.9489
ВПК <sub>кор</sub>	0.8815	0.9389
ВПК <sub>лин</sub>	0.9613	0.9814
Ridge	0.9611	0.9810
Lasso	0.9492	0.9750
ElasticNet	0.9527	0.9766
ARD	<b>0.9627</b>	<b>0.9822</b>
Байесовская	0.9624	0.9821