



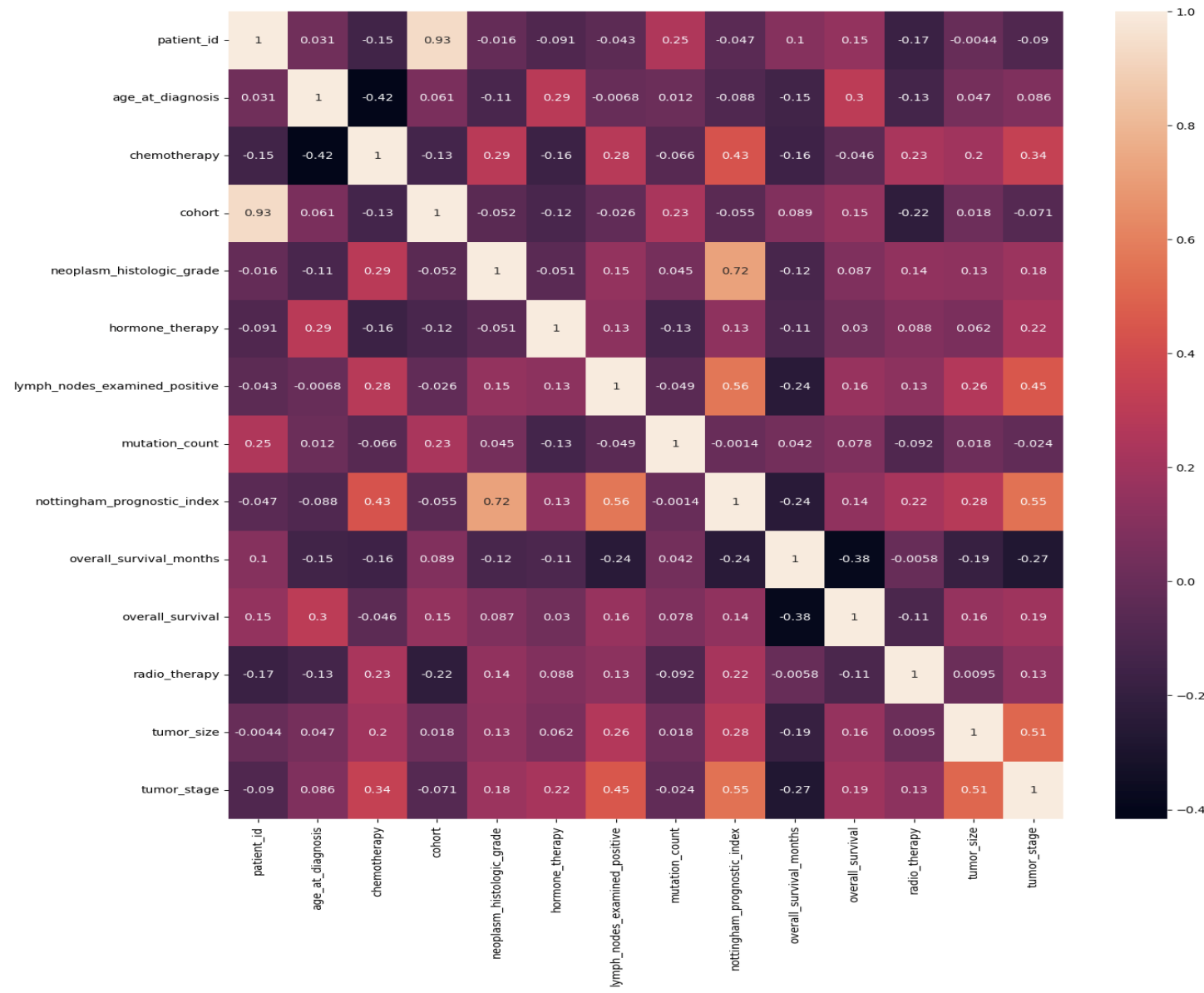
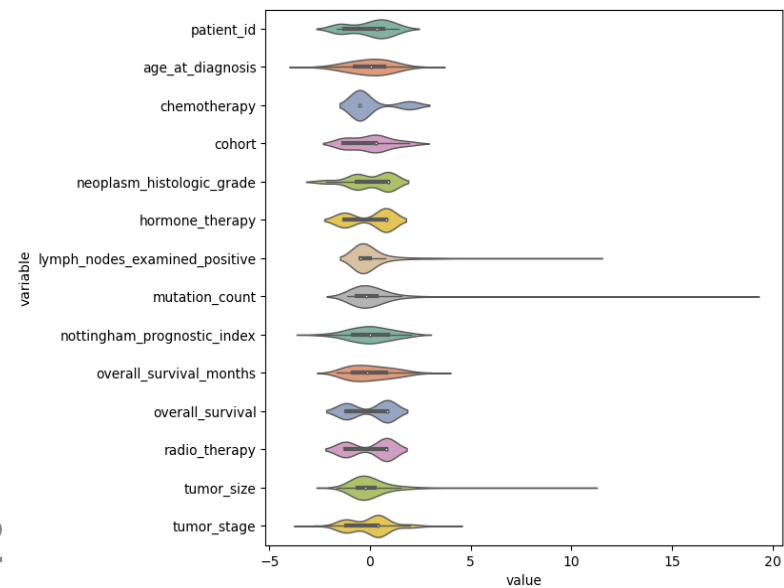
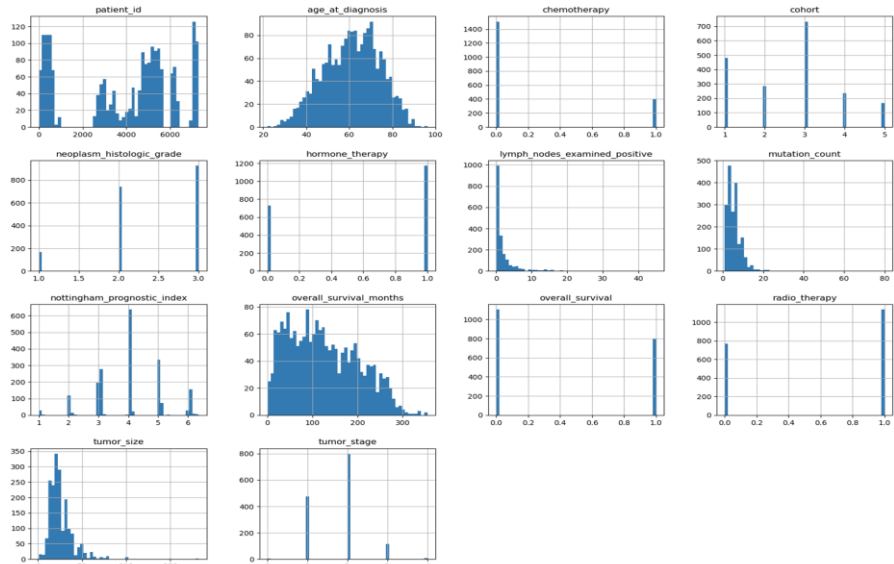
ЦЕНТР  
ДОПОЛНИТЕЛЬНОГО  
ОБРАЗОВАНИЯ  
МГТУ им. Н.Э. Баумана

# Анализ выживаемости пациентов с раком молочной железы методами машинного обучения на примере датасета METABRIC

Стебунова Рената Вячеславовна



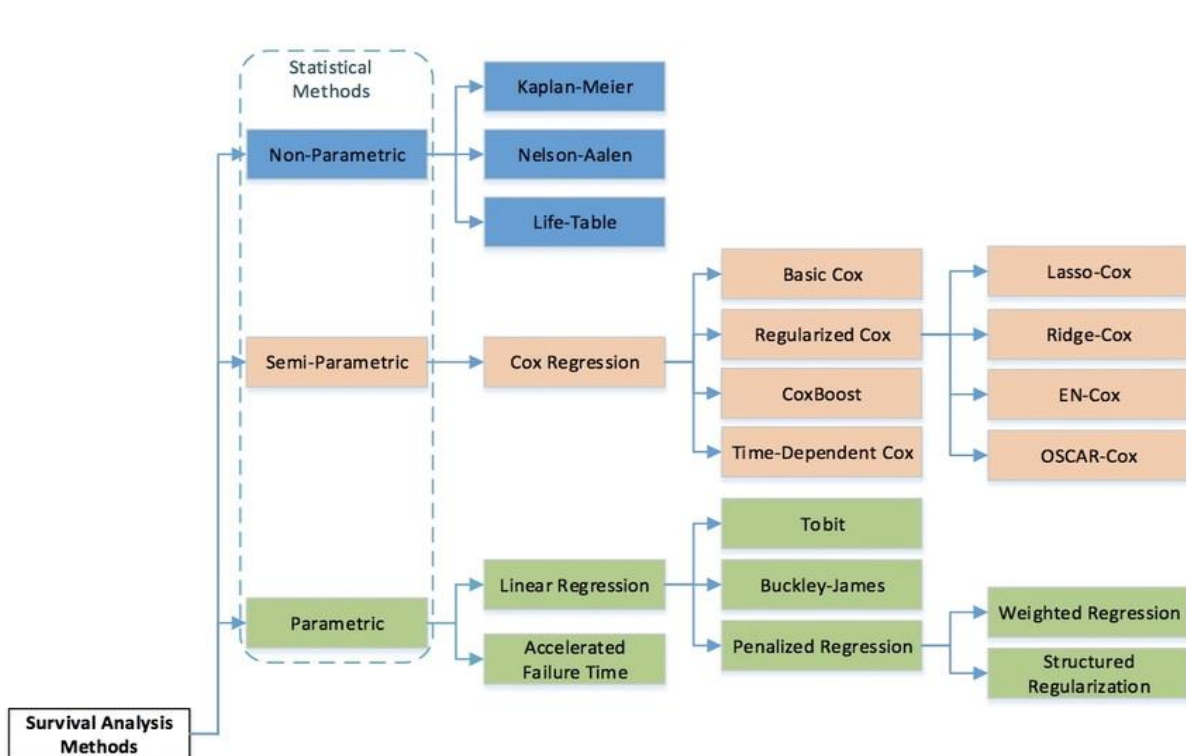
# Разведочный анализ данных



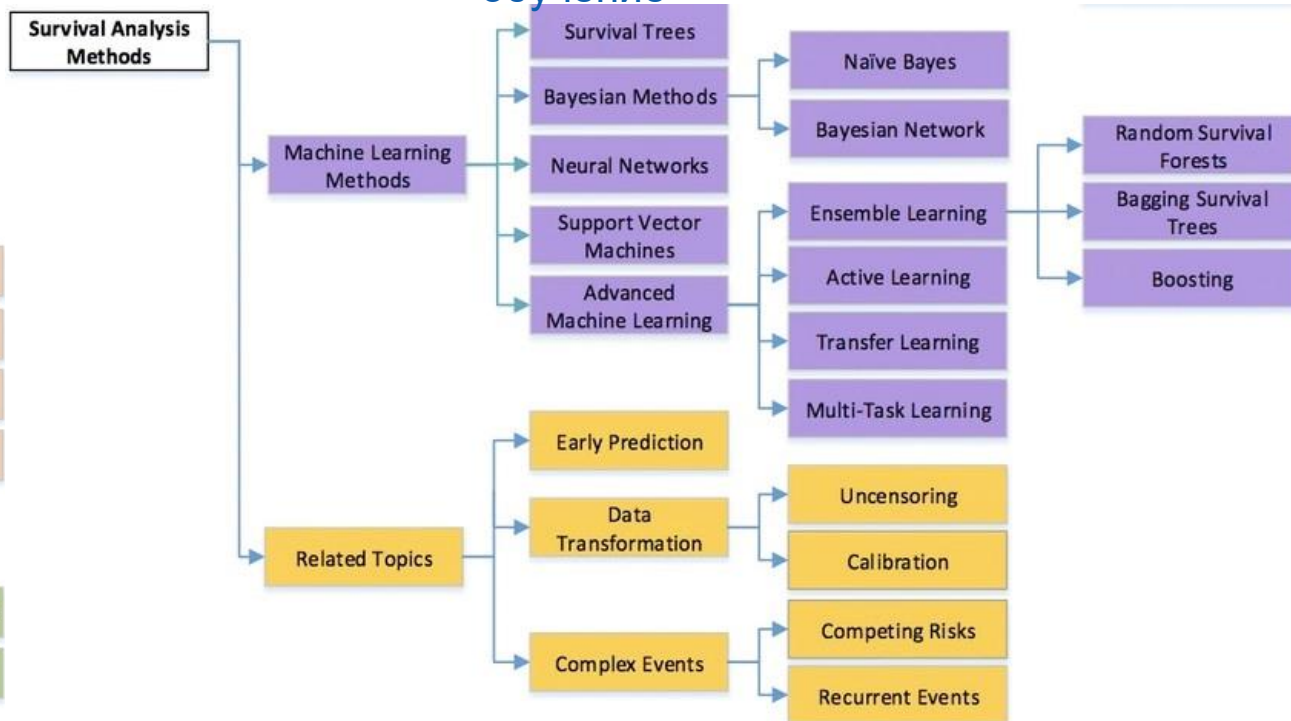


# Методы анализа выживаемости

## Стандартная статистика



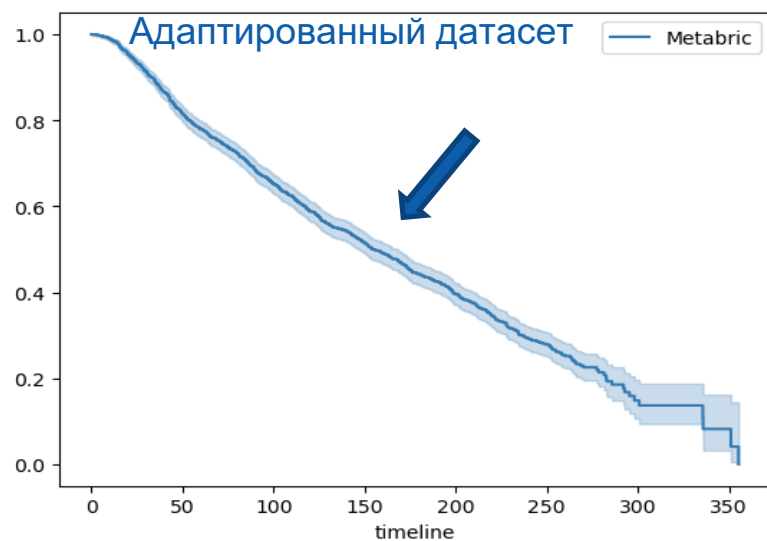
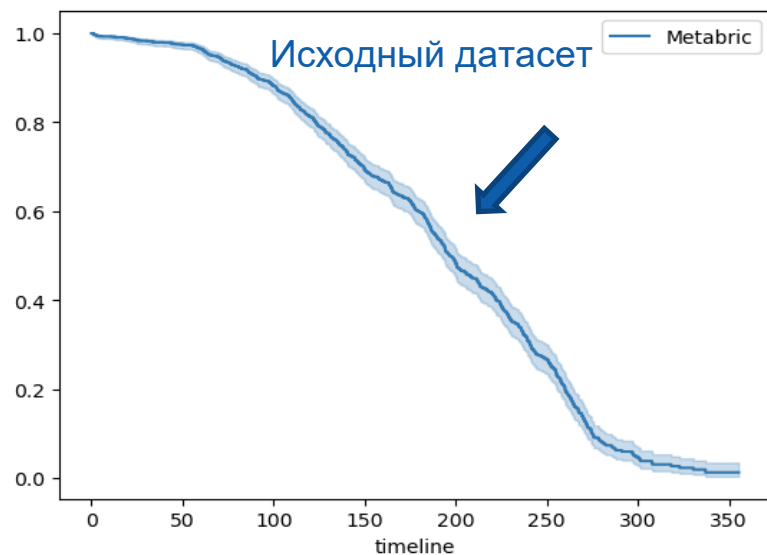
## Машинное обучение



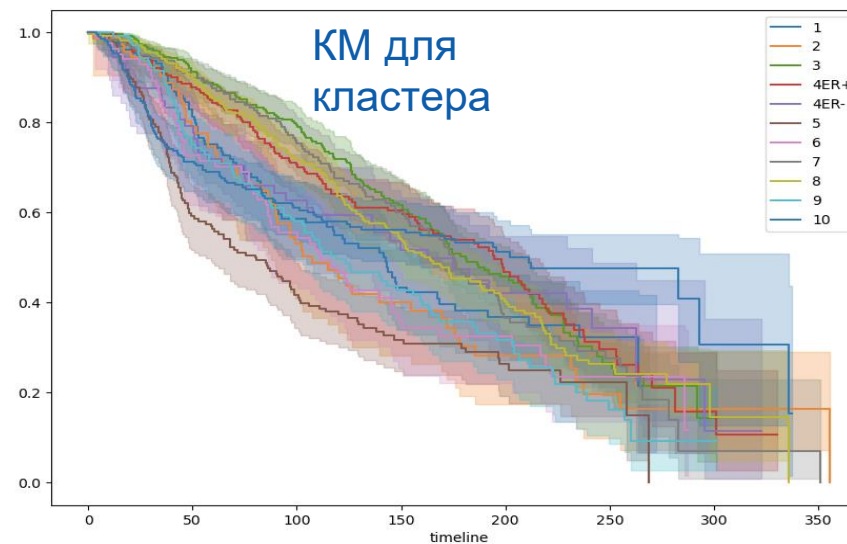
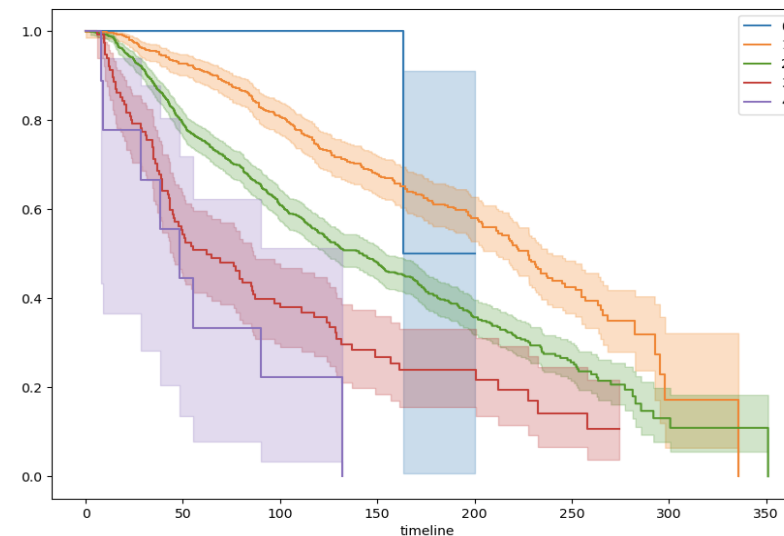
Специальные библиотеки: lifelines, scikit-survival, rpy2 и др.



# Описательная статистика Кривые Каплана-Мейера



КМ для стадии заболевания

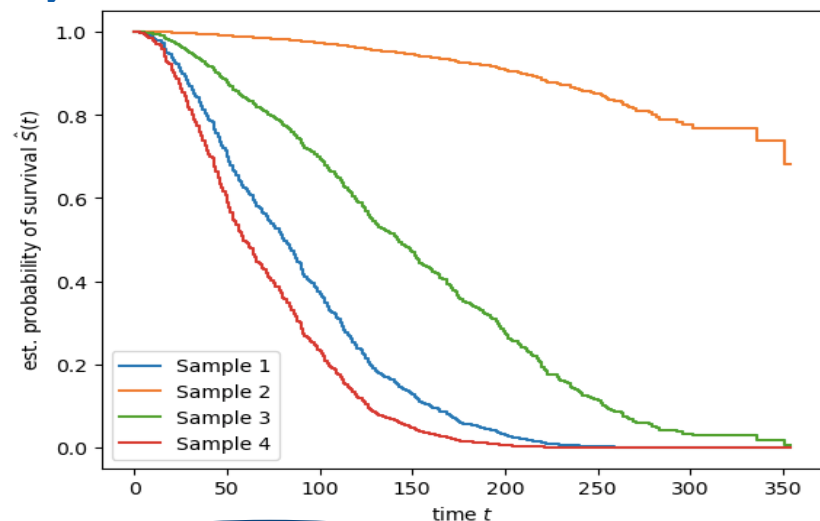




# Регрессия Кокса и регуляризованные регрессии

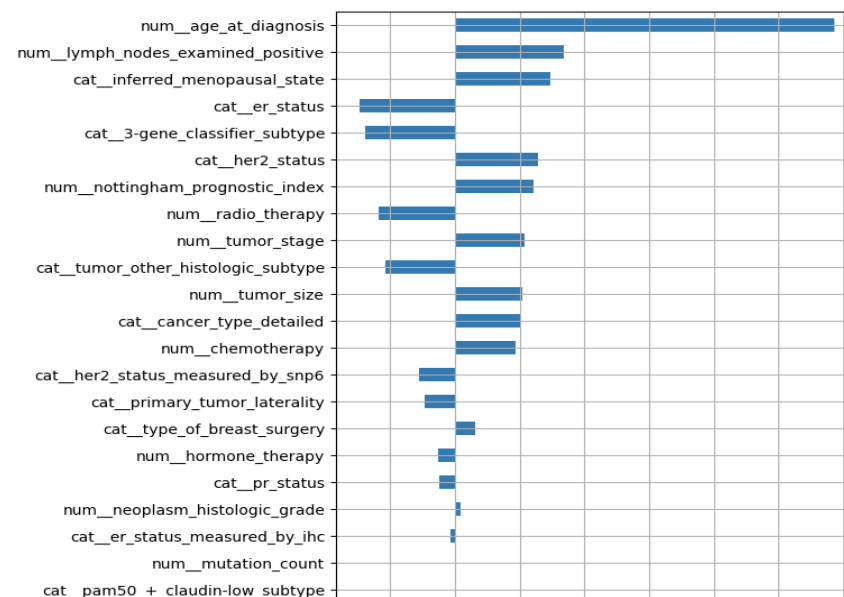
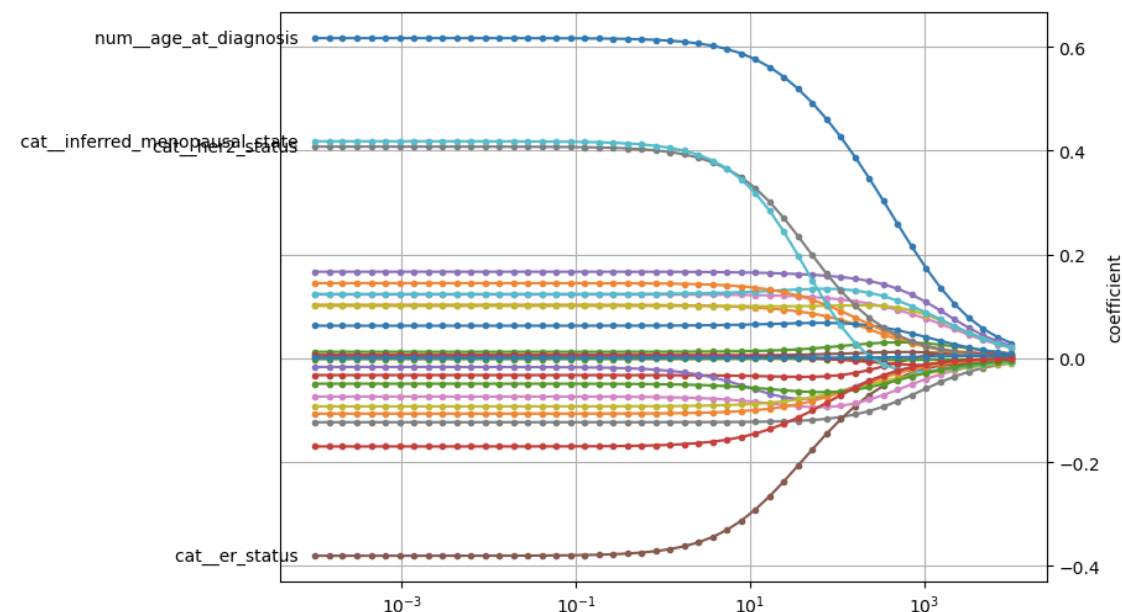
## Гребневая регрессия

## Функции выживания для 4 синтетических



```
Out[34]: cat_death_from_cancer 0.813584
num_nottingham_prognostic_index 0.653552
num_tumor_size 0.626544
num_lymph_nodes_examined_positive 0.619861
num_age_at_diagnosis 0.601891
num_tumor_stage 0.590635
cat_type_of_breast_surgery 0.569659
num_neoplasm_histologic_grade 0.569173
cat_primary_tumor_laterality 0.544821
cat_pr_status 0.544082
num_chemotherapy 0.538274
cat_pam50+_claudin-low_subtype 0.537048
cat_integrative_cluster 0.536231
cat_her2_status_measured_by_snp6 0.535803
cat_er_status 0.529537
cat_er_status_measured_by_ihc 0.529294
num_hormone_therapy 0.522386
cat_inferred menopausal_state 0.520256
cat_tumor_other_histologic_subtype 0.515751
num_mutation_count 0.514720
cat_her2_status 0.512609
cat_cellularity 0.505273
num_radio_therapy 0.504670
cat_3-gene_classifier_subtype 0.503950
cat_cancer_type_detailed 0.486457
dtype: float64
```

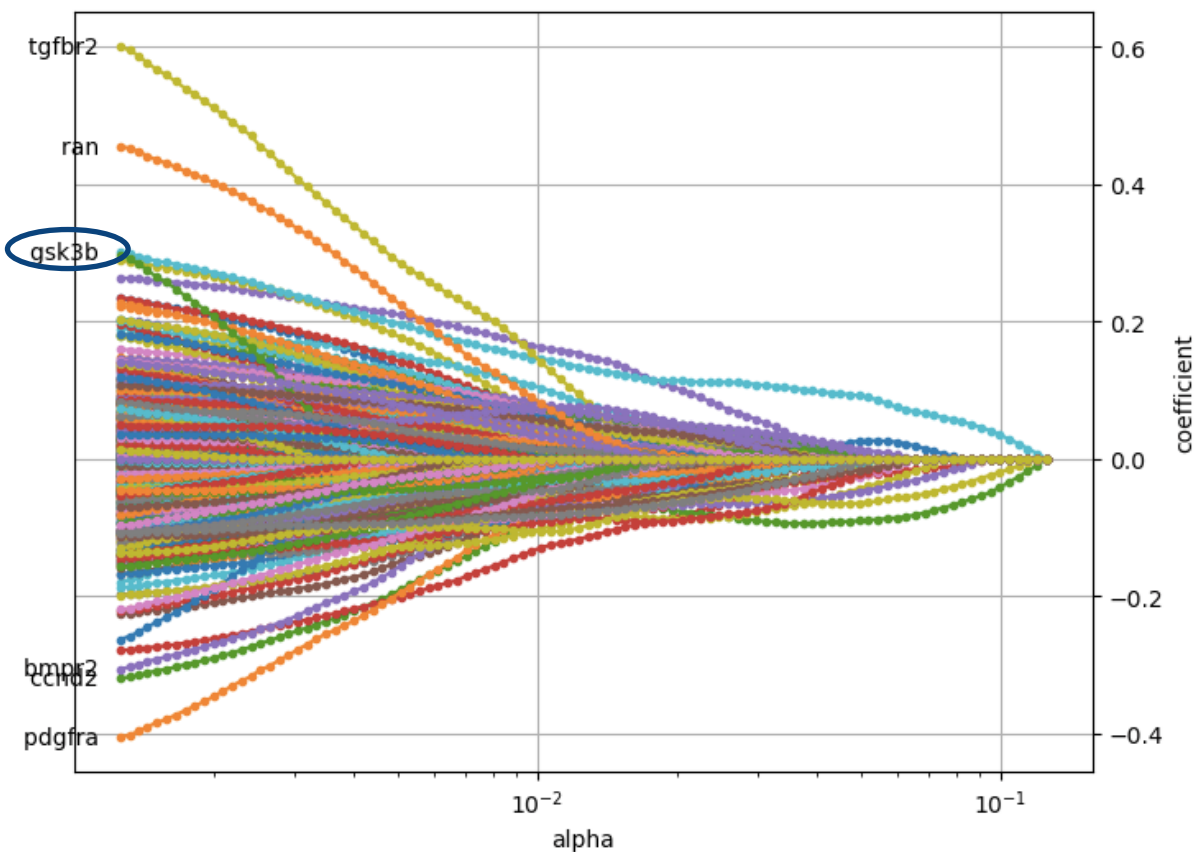
Индекс конкордантности  
0,84 → 0,69



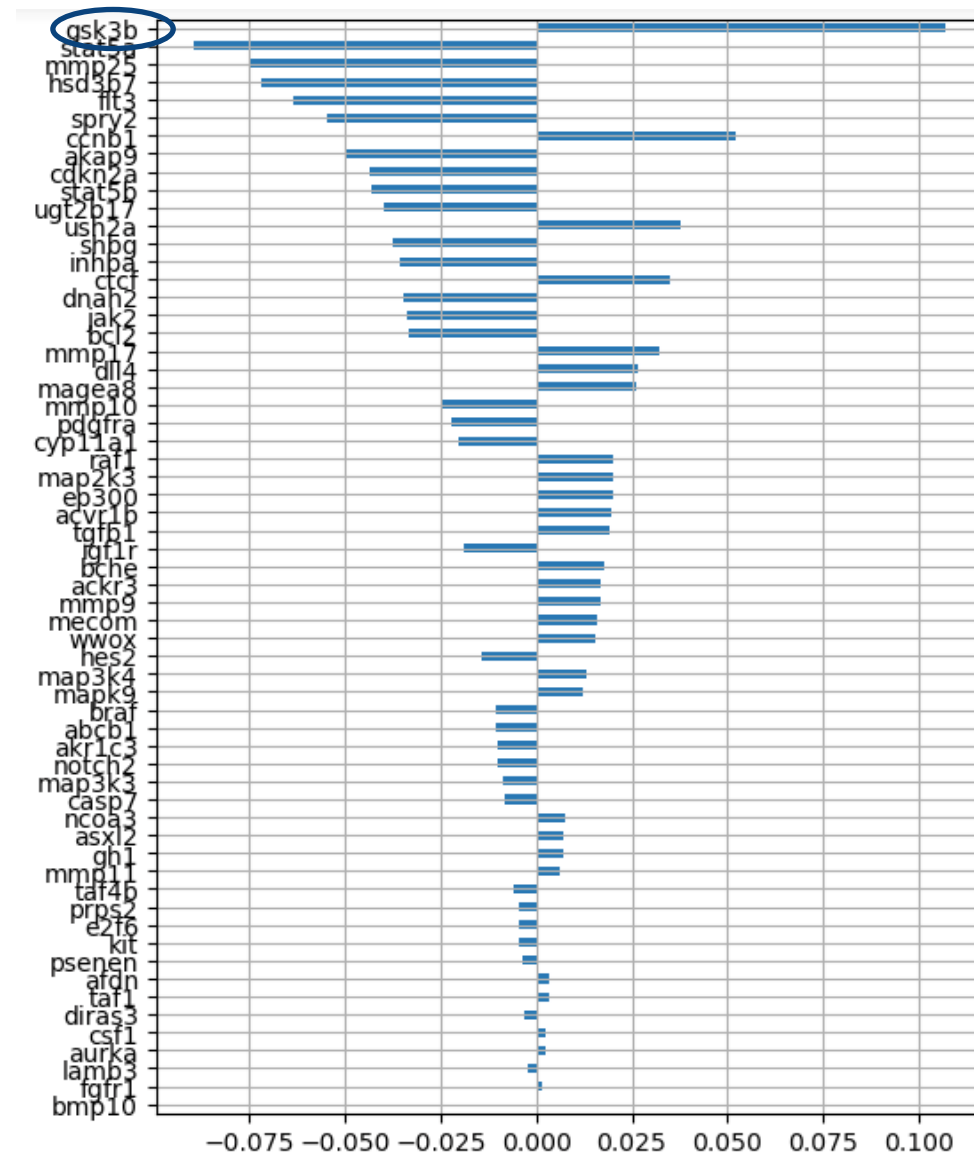


# Регуляризованные регрессии для генетических признаков

Регрессия  
Лассо



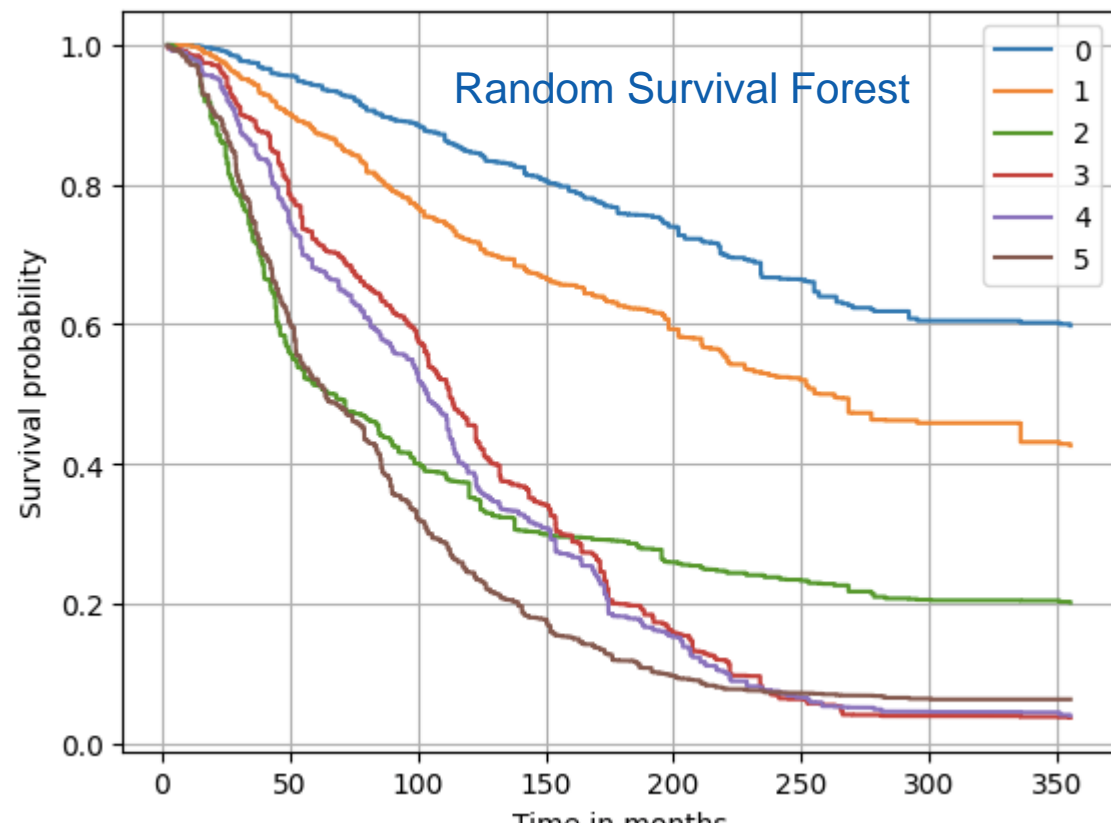
Наилучший признак по лучшему alpha





# Ансамбли

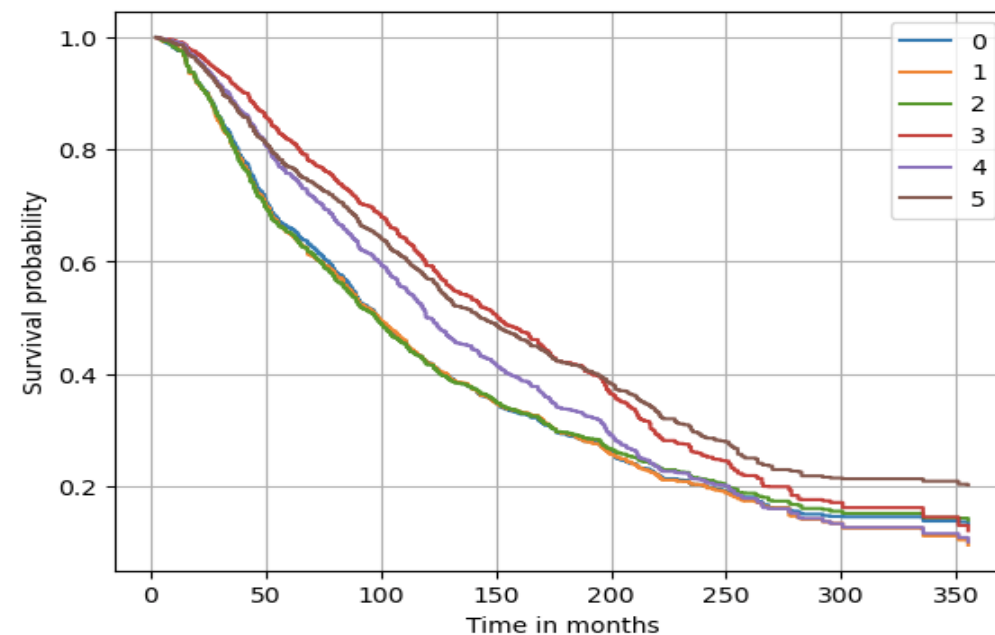
	importances_mean	importances_std
num_age_at_diagnosis	0.065319	0.012908
num_lymph_nodes_examined_positive	0.023484	0.005761
num_nottingham_prognostic_index	0.018623	0.005004
num_tumor_size	0.007129	0.003383
num_tumor_stage	0.003379	0.001702
cat_her2_status_measured_by_snp6	0.002178	0.001157
cat_primary_tumor_laterality	0.001554	0.000750



## C-индекс:

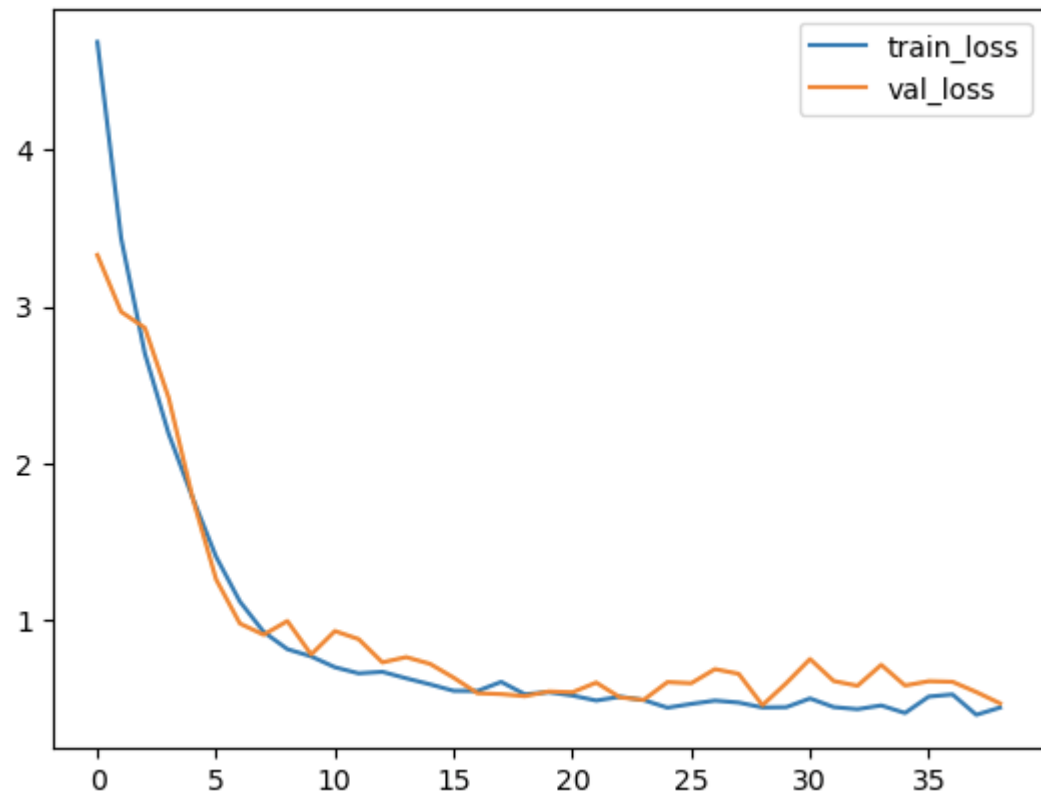
Случайный лес – 0,71  
Градиентный бустинг – 0,71  
(клинические признаки)

Случайный лес – 0,63  
Градиентный бустинг – 0,61  
(генетические признаки)

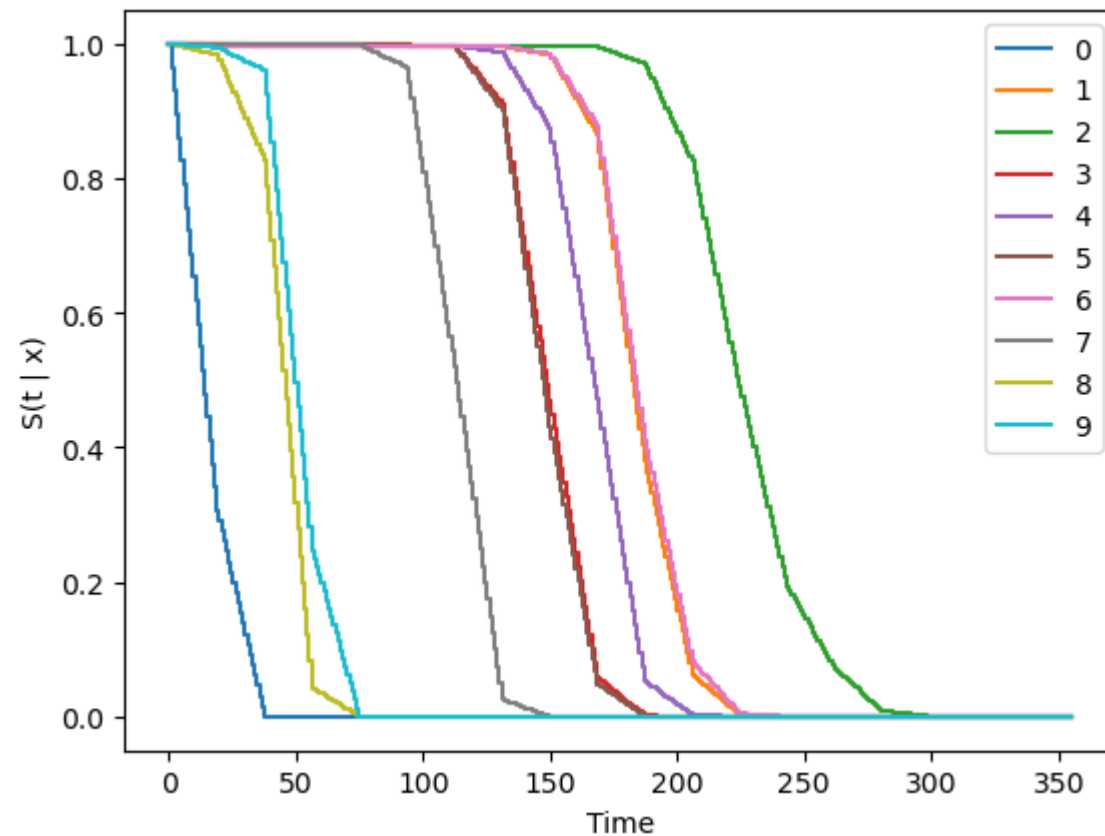




# Нейронная сеть DeepSurv (русск)



Линейная интерполяция кривых выживаемости



С-индекс Антолини – 0,99





ЦЕНТР  
ДОПОЛНИТЕЛЬНОГО  
ОБРАЗОВАНИЯ  
МГТУ им. Н.Э. Баумана



[do.bmstu.ru](https://do.bmstu.ru)