



ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ  
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

**САНКТ-ПЕТЕРБУРГСКИЙ НАЦИОНАЛЬНЫЙ  
ИССЛЕДОВАТЕЛЬСКИЙ АКАДЕМИЧЕСКИЙ УНИВЕРСИТЕТ  
РОССИЙСКОЙ АКАДЕМИИ НАУК**

---

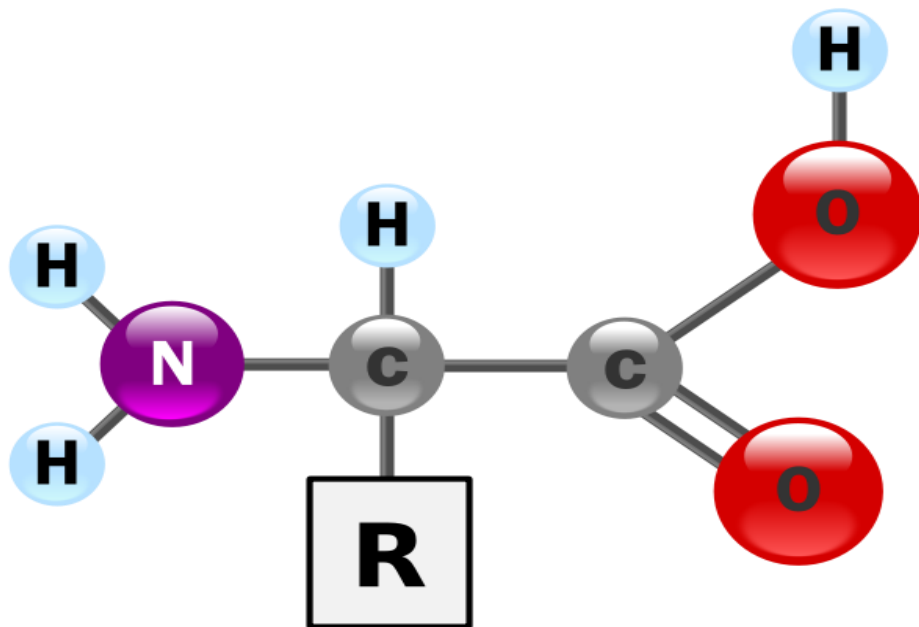
# **Предсказание окисдирования метионина по третичной структуре белка**

Студент: Гайдай И.О.  
Руководитель: Яковлев П.А.

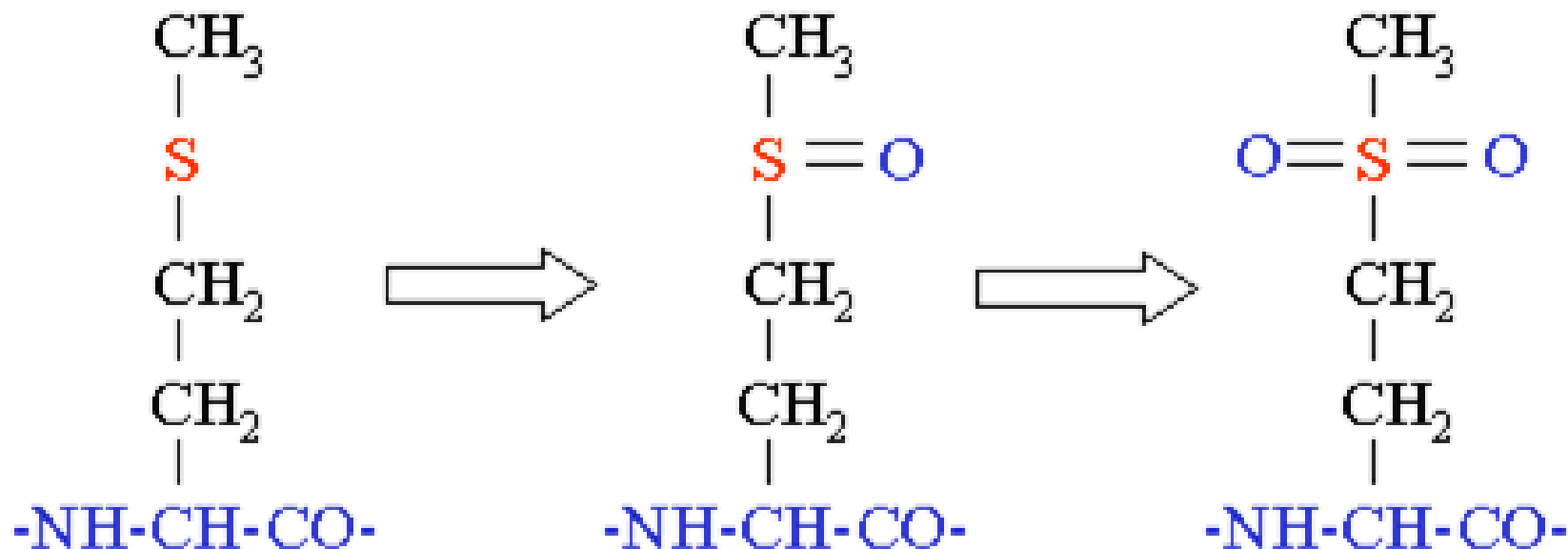
СПбАУ, 2015

# Аминокислоты

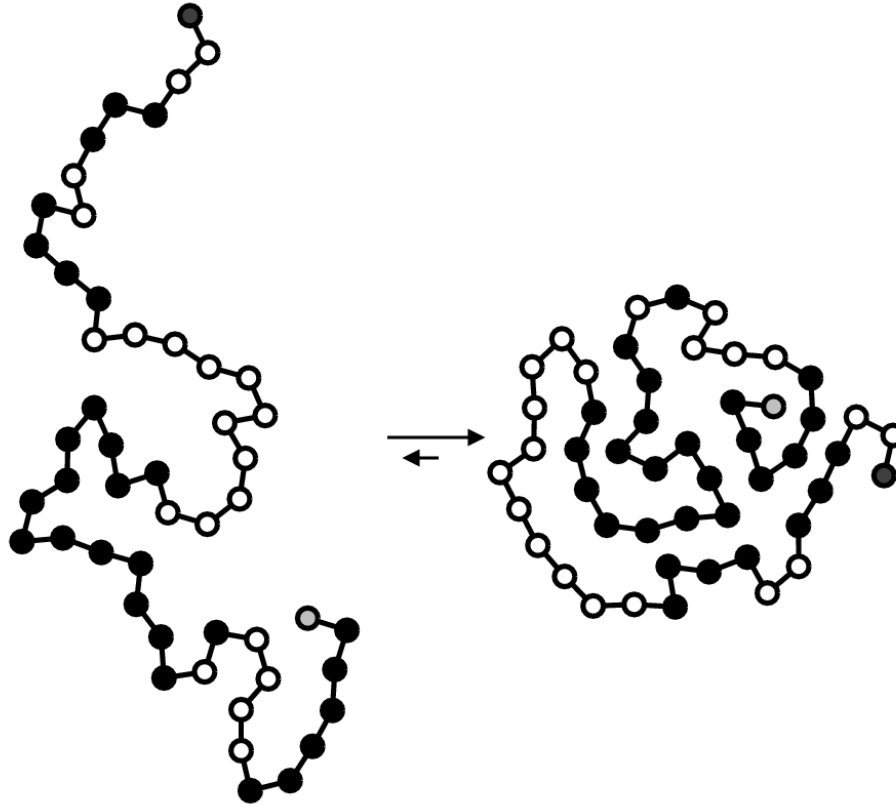
Аминокислоты состоят из атомов. Общая структура аминокислоты:



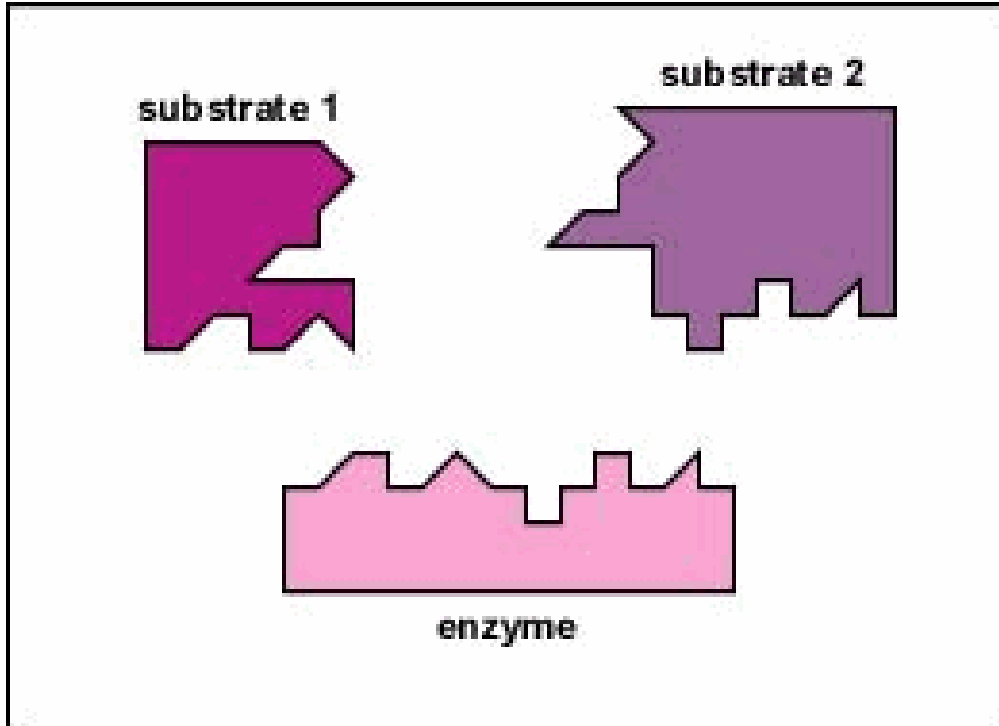
# Оксидирование метионина



# Проблема свертки белка



# Существующие решения



- AMS (2012)
- GPS (2008)
- KinasePhos (2007)
- PPSP (2006)
- PredPhospho (2004)
- Scansite (2003)
  
- COPA(2008)

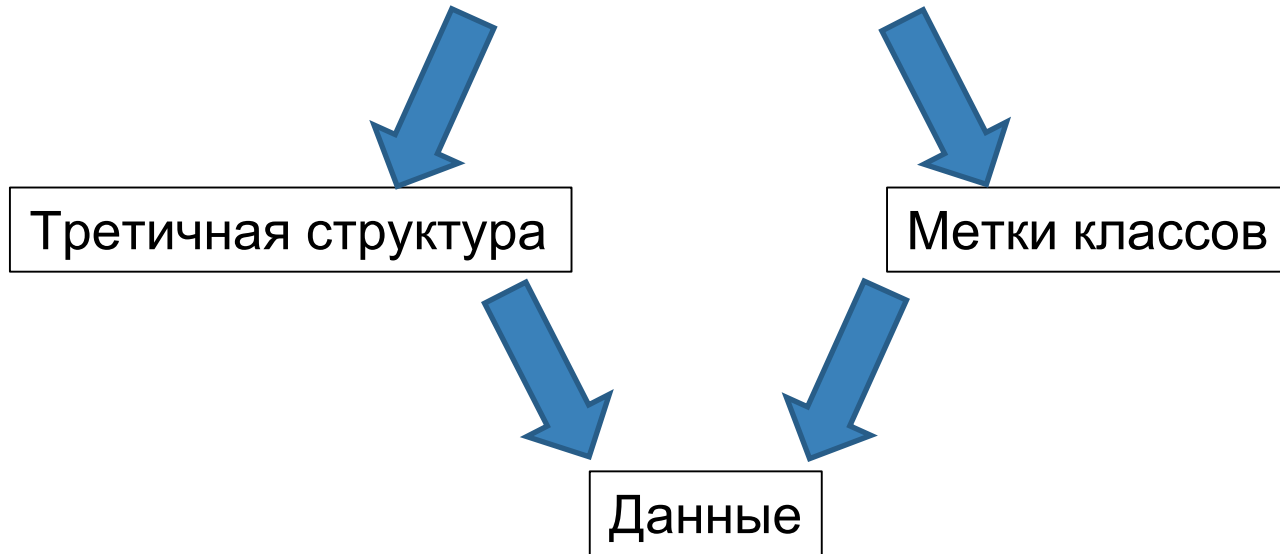
# Цель

Научиться предсказывать вероятность окислирования метионина.

# Задачи

1. Найти достоверные данные
2. Найти значимые предикторы
3. Подобрать оптимальные методы и параметры

# Данные



# Предикторы

Наиболее значимые предикторы:

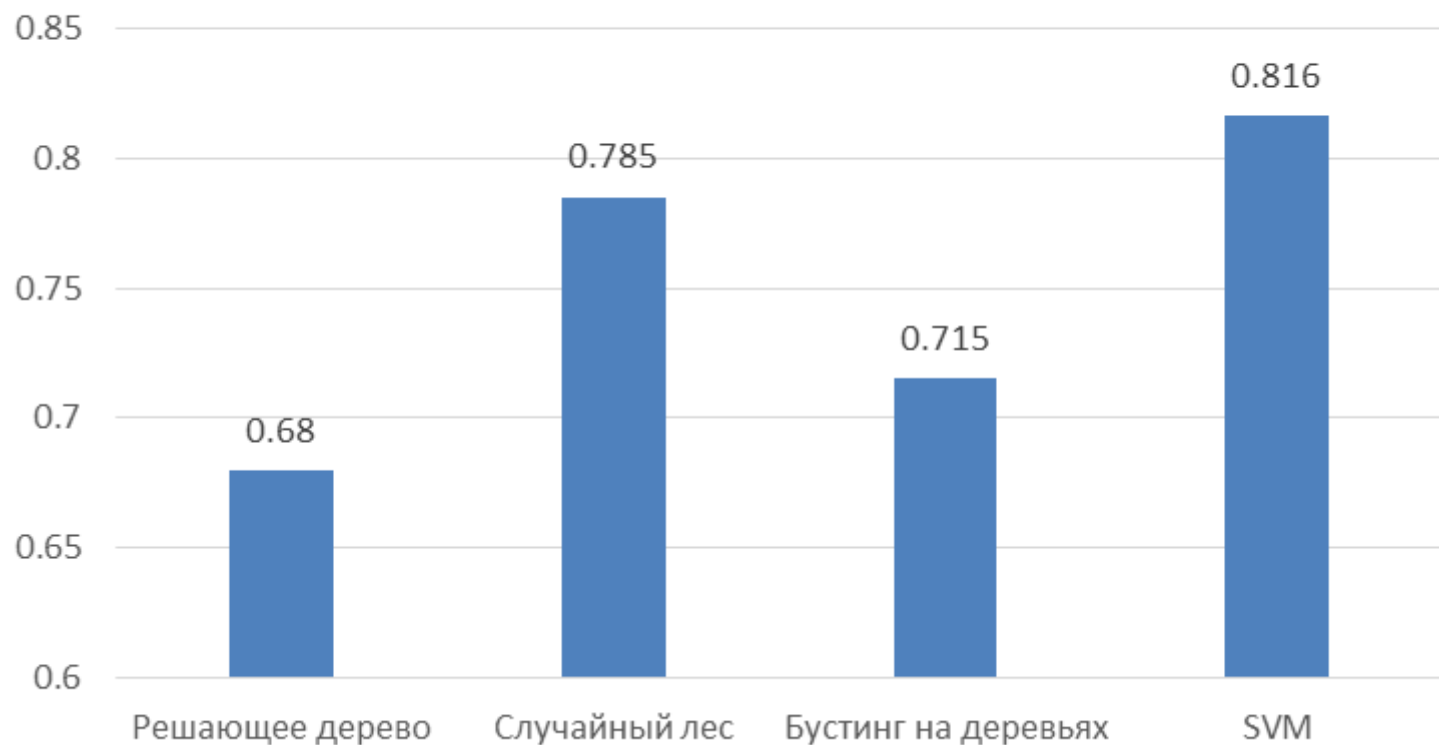
- Типы соседей
- Расстояния до соседей
- Доступность растворителю (SAS, SES)
- Вторичная структура



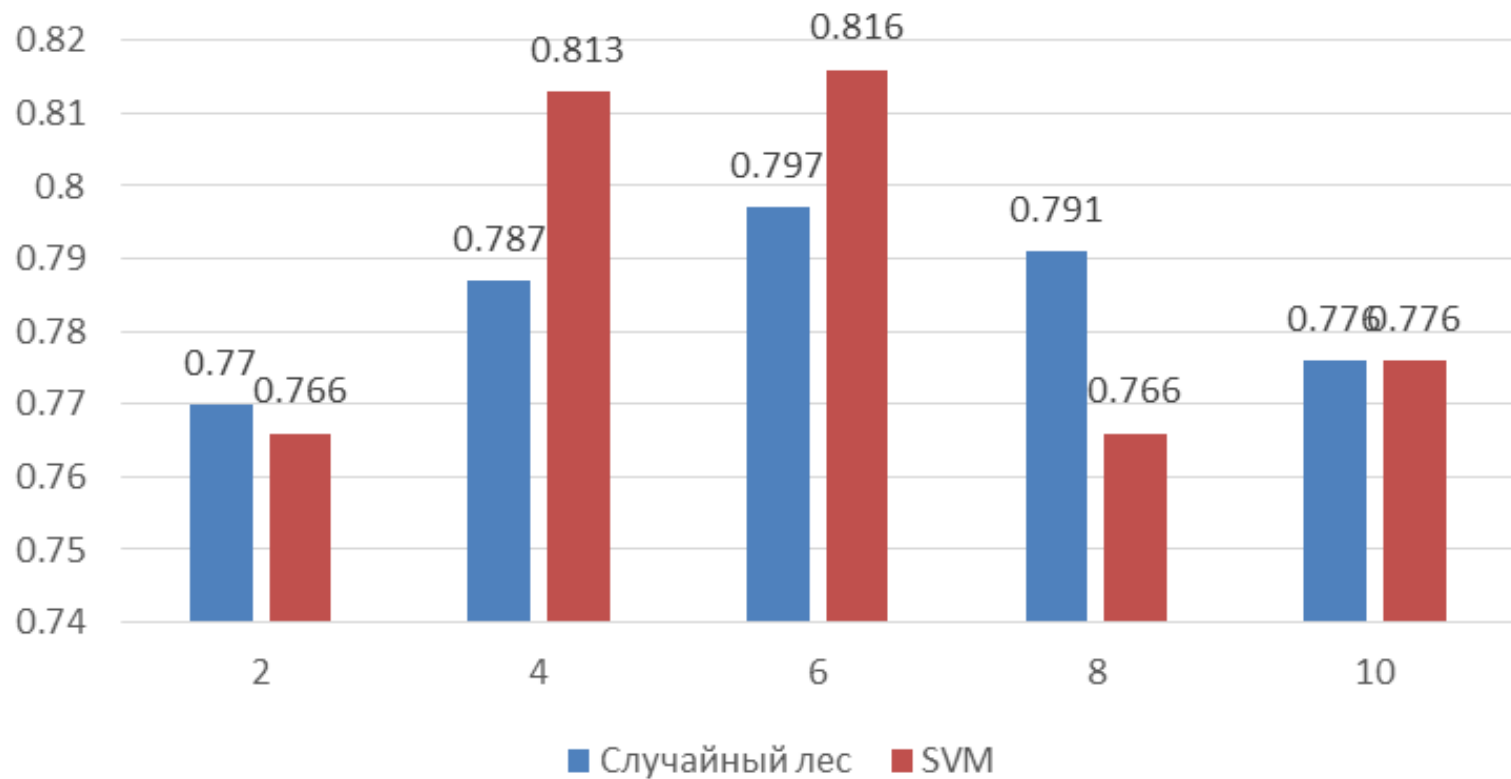
# Методы

- Решающее дерево
- Случайный лес
- Бустинг на деревьях
- SVM

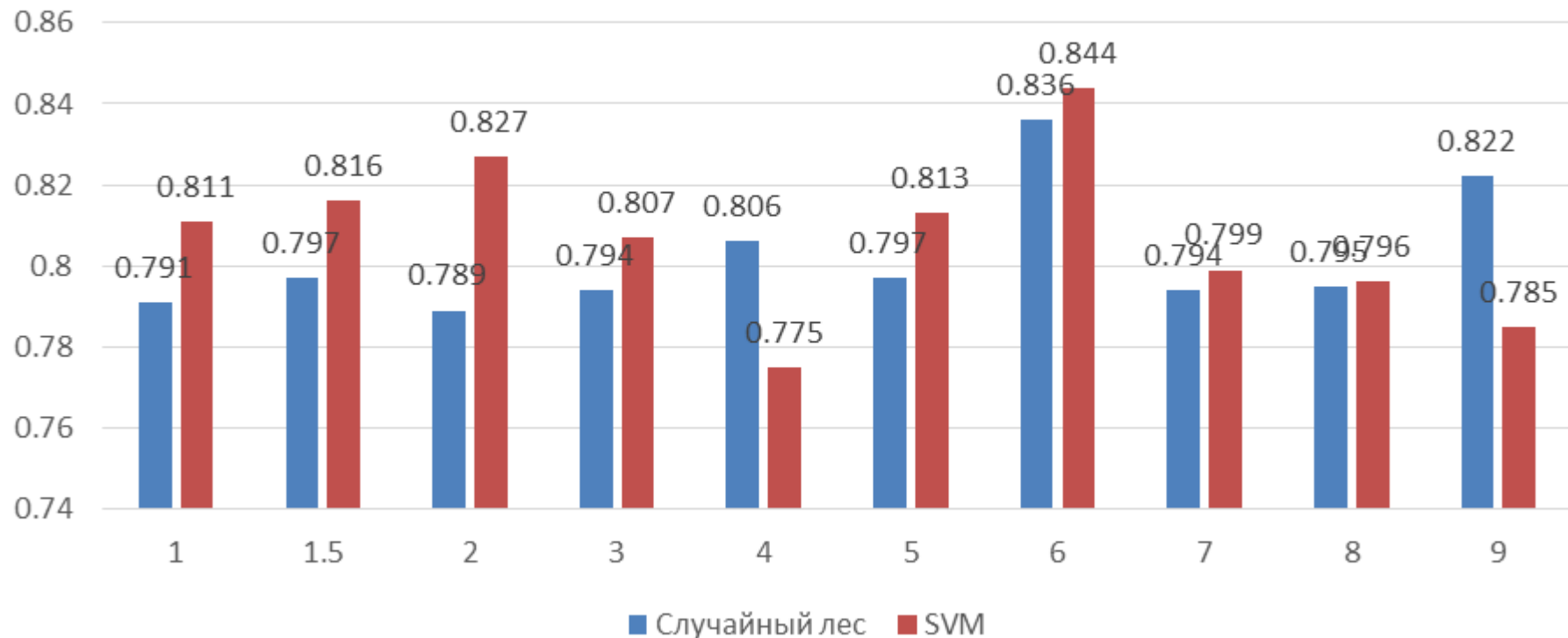
## Сравнение эффективности методов



## Зависимость эффективности от числа соседей

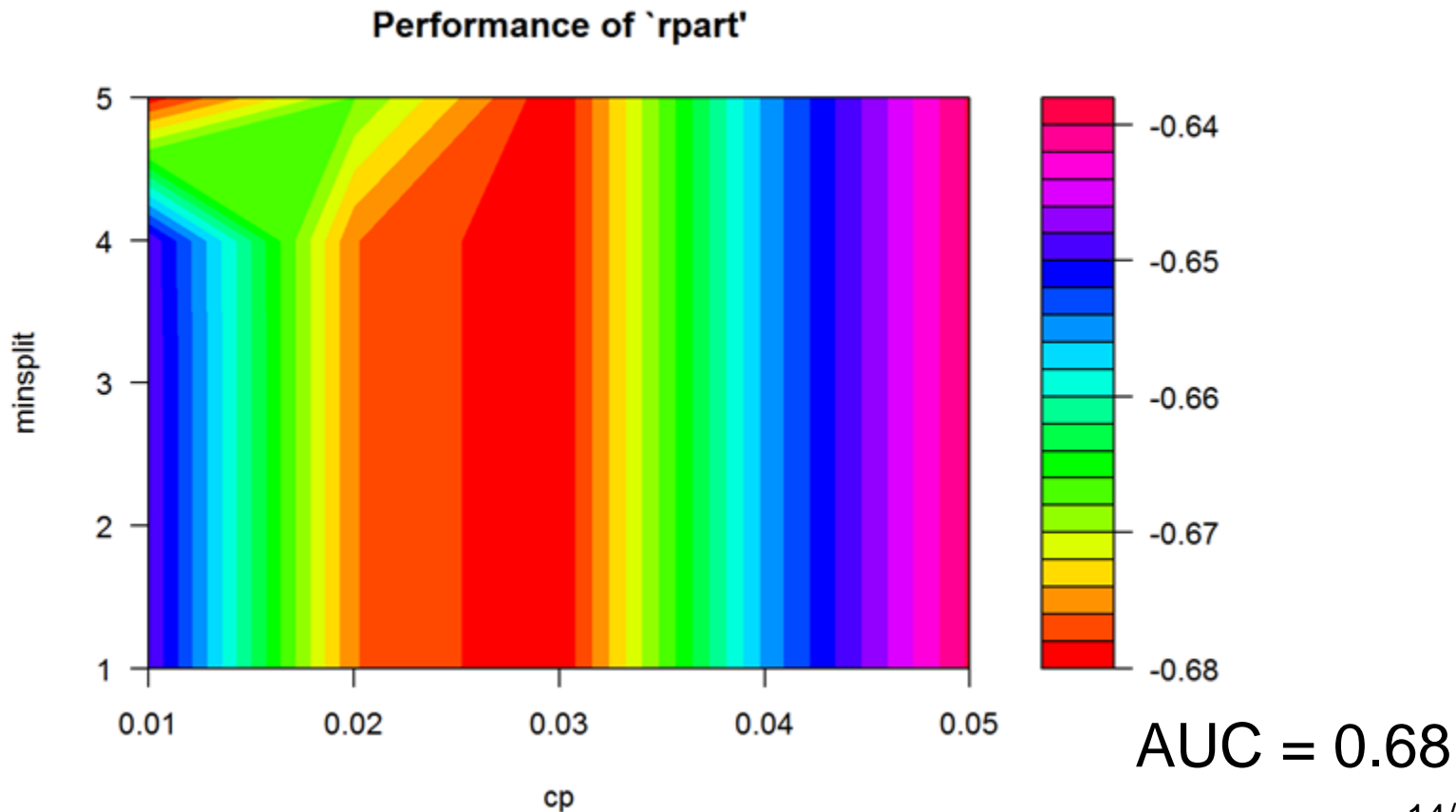


## Зависимость эффективности от грубости поверхности

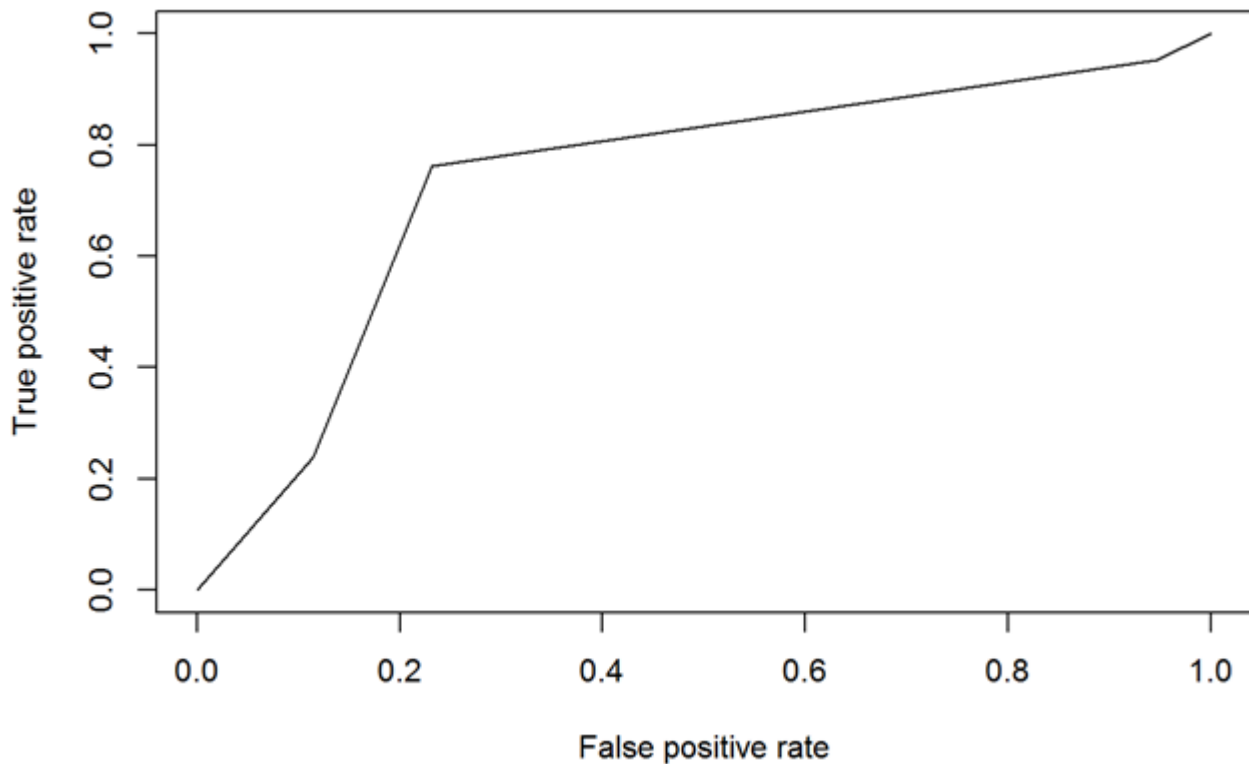


**Спасибо за внимание!**

# Подбор параметров решающего дерева

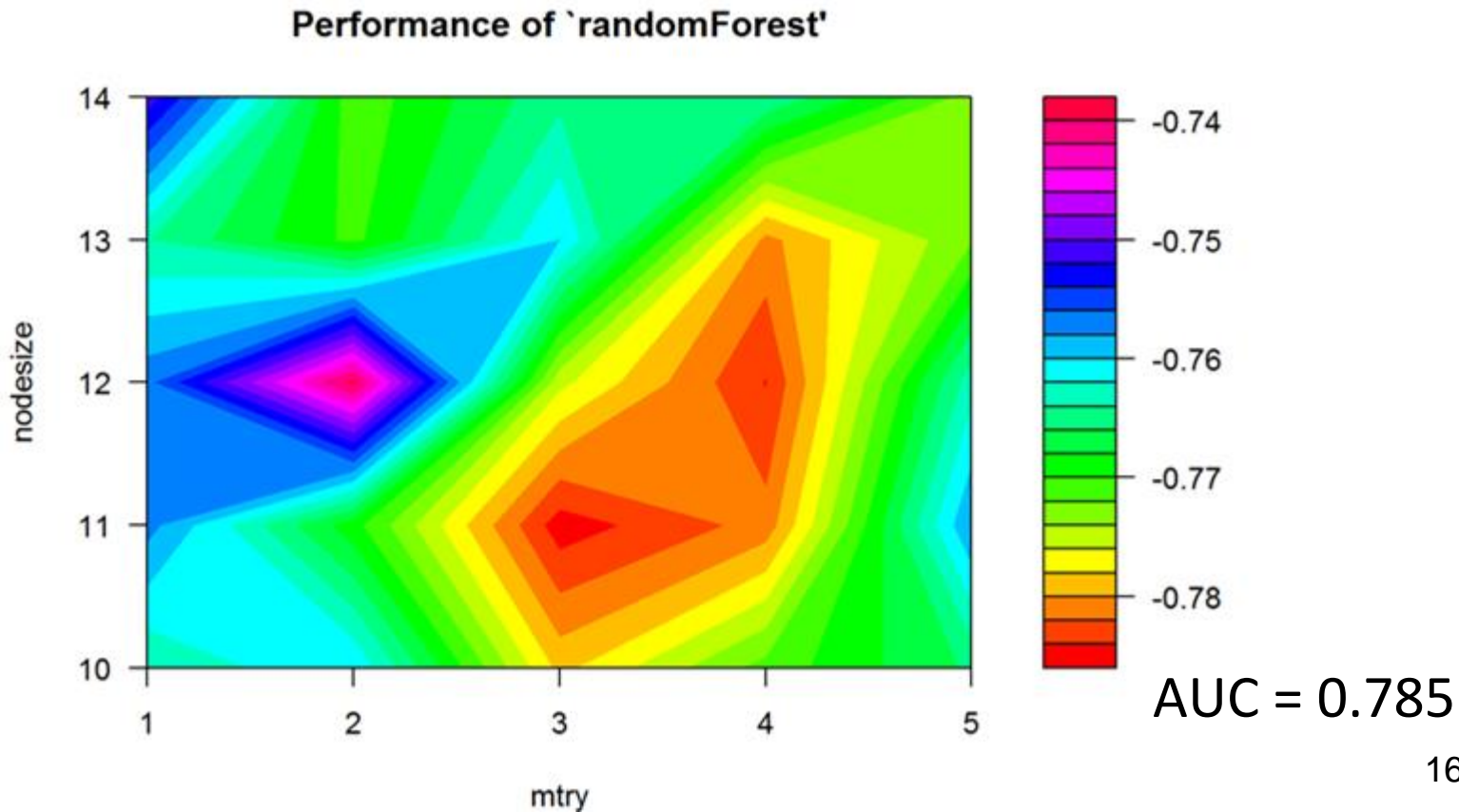


# ROC-кривая для предсказания на тестовой выборке для решающего дерева



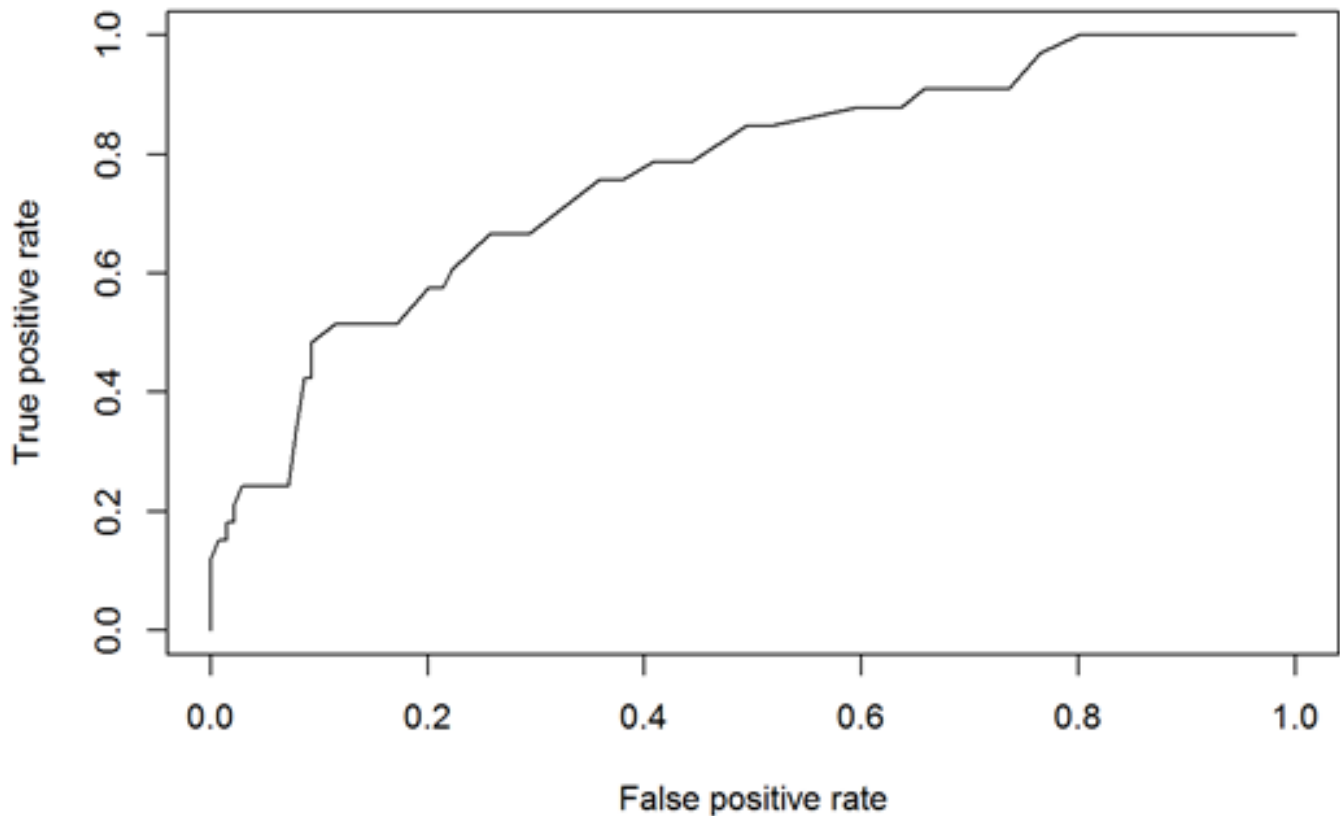
AUC = 0.737

# Подбор параметров случайного леса



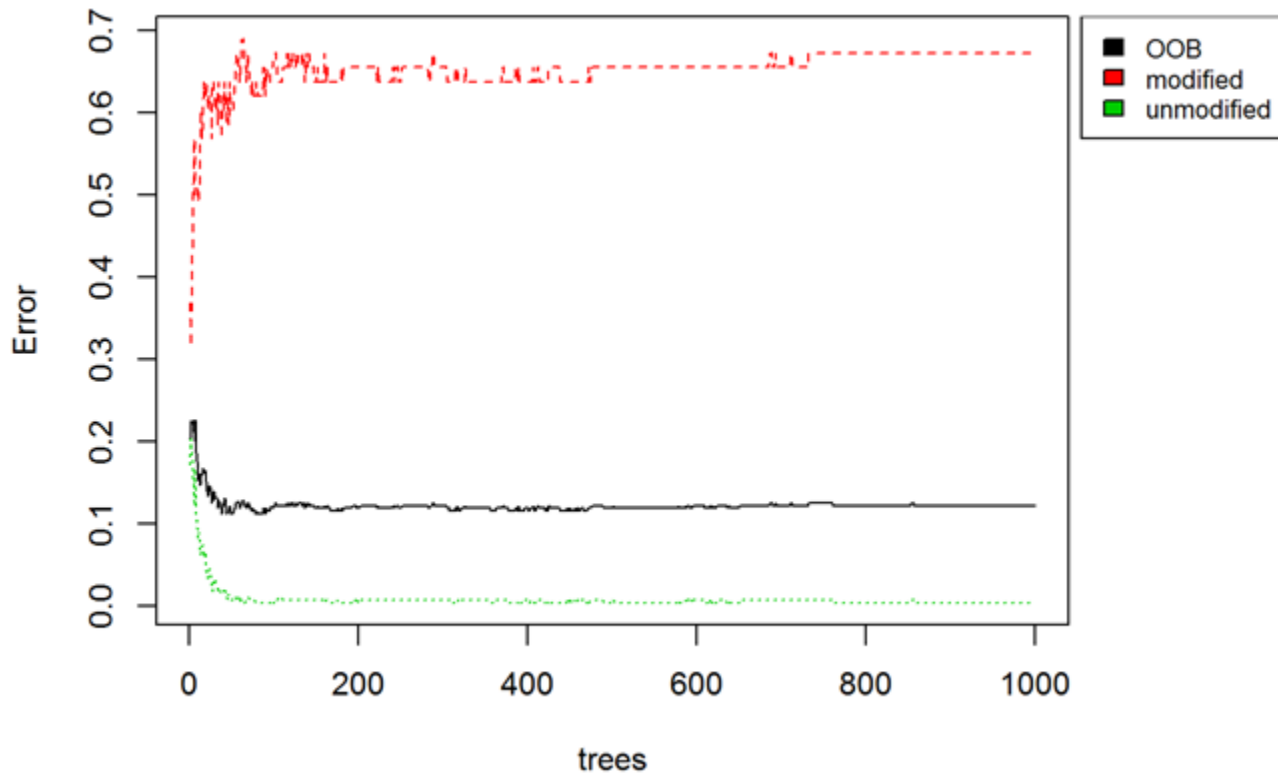


# ROC-кривая для предсказания на тестовой выборке для случайного леса

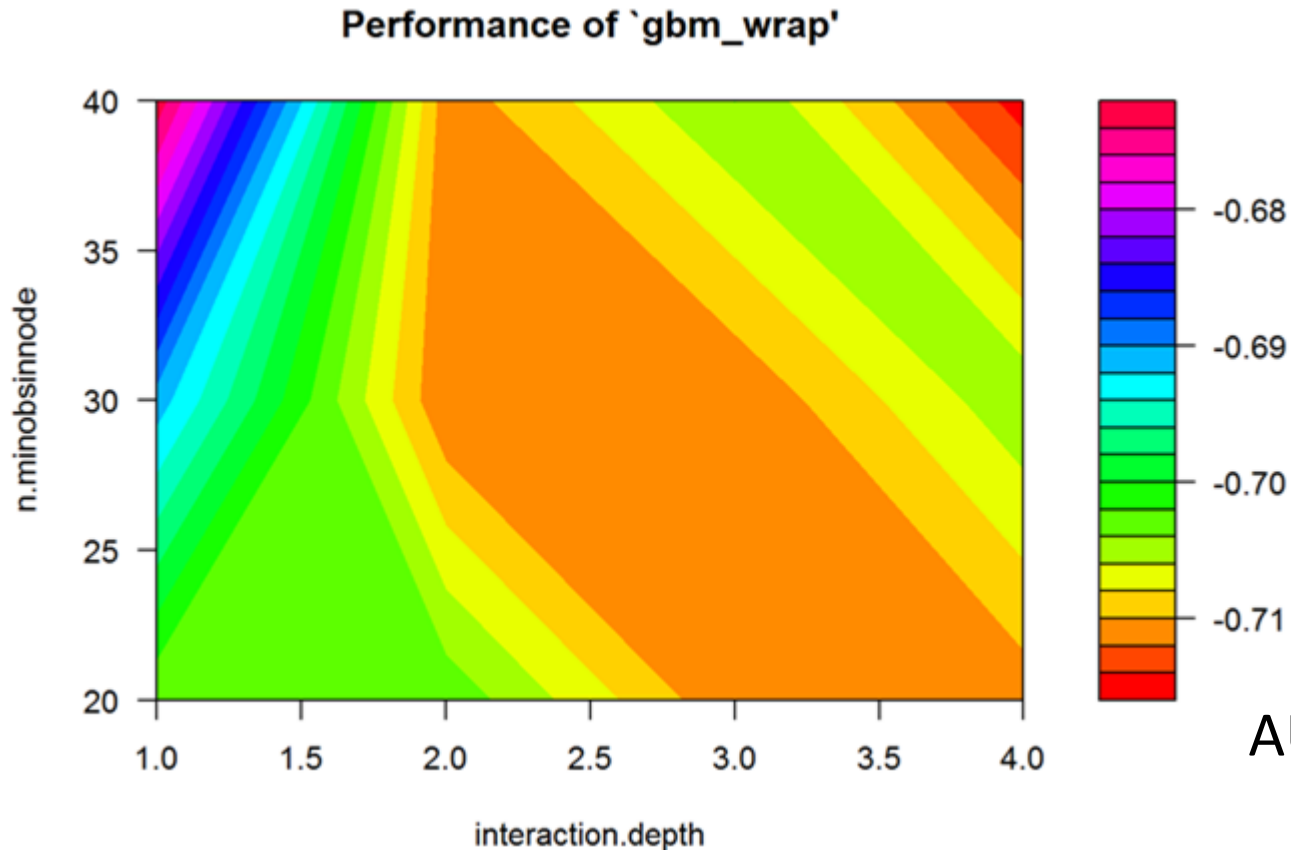


AUC = 0.767

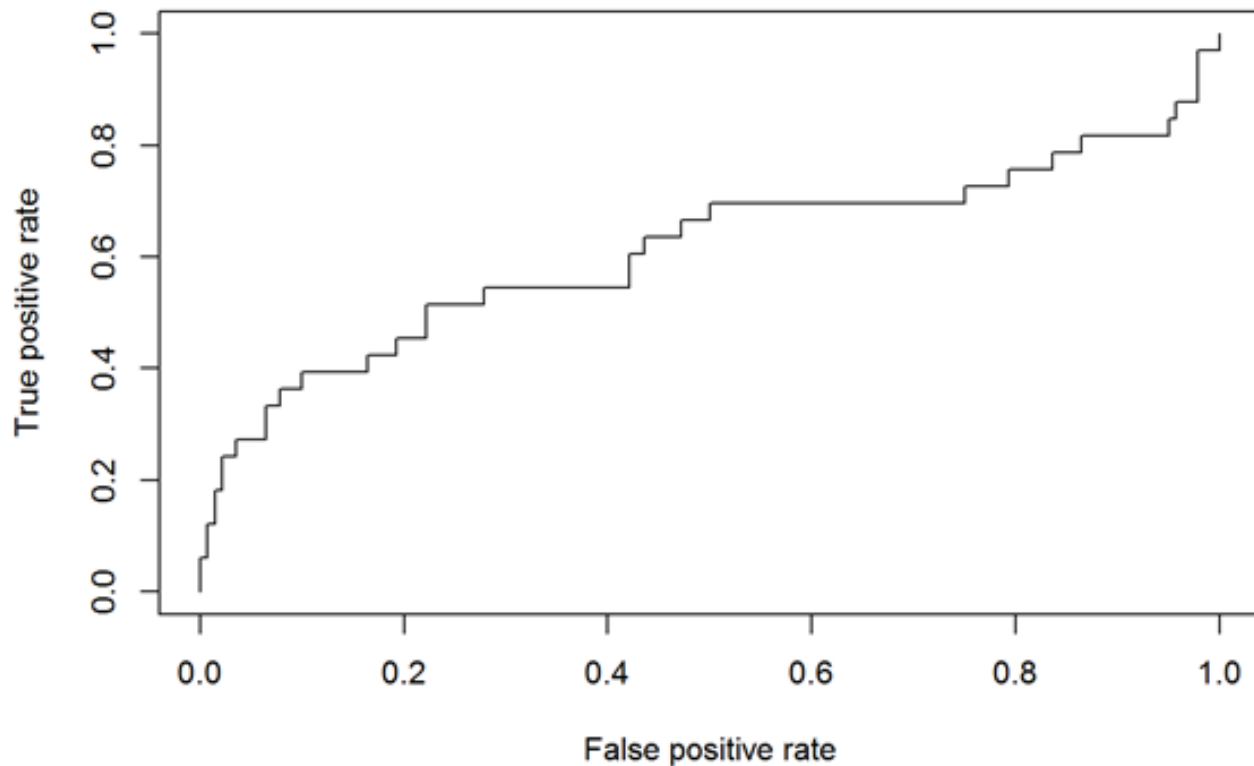
# Зависимость ошибки от числа деревьев в лесу



# Подбор параметров бустинга на деревьях

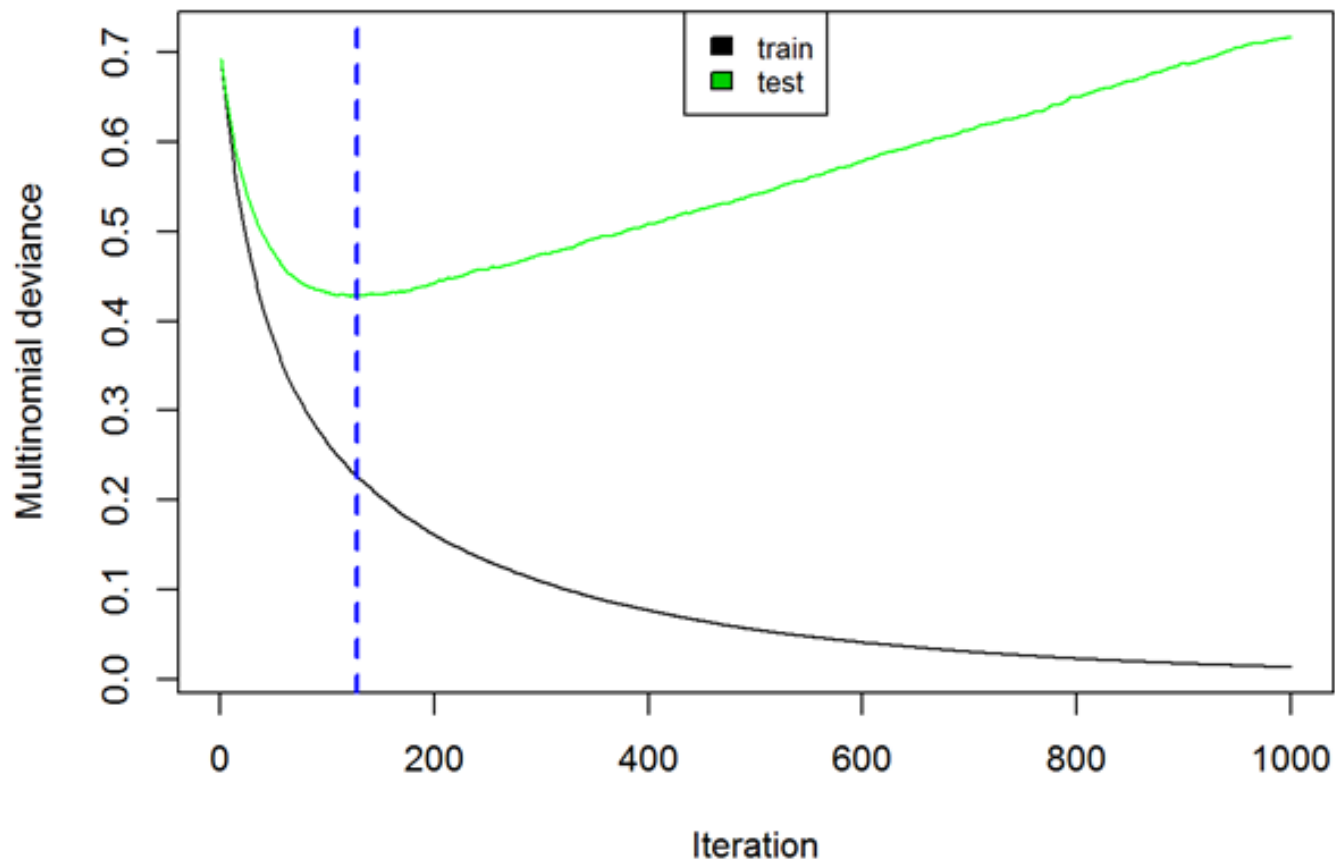


# ROC-кривая для предсказания на тестовой выборке для бустинга на деревьях



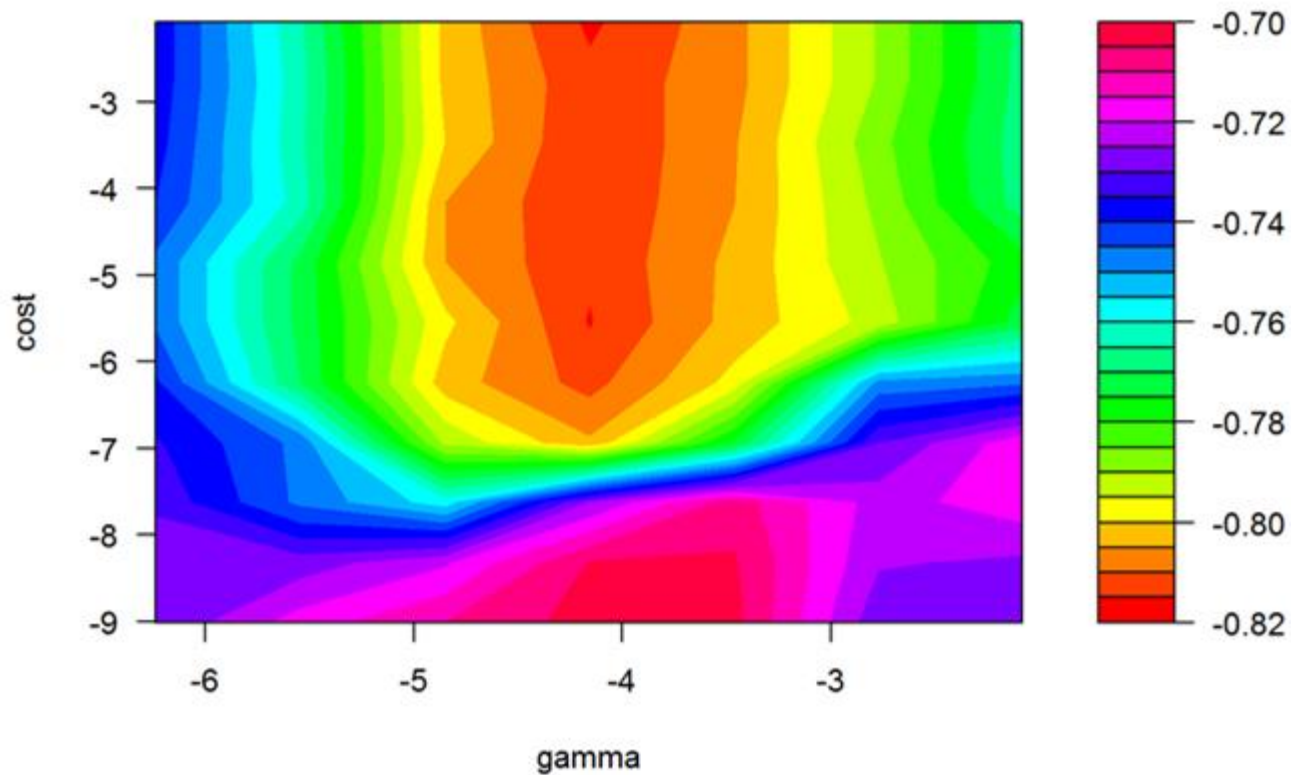
AUC = 0.611

# Зависимость ошибки от числа деревьев



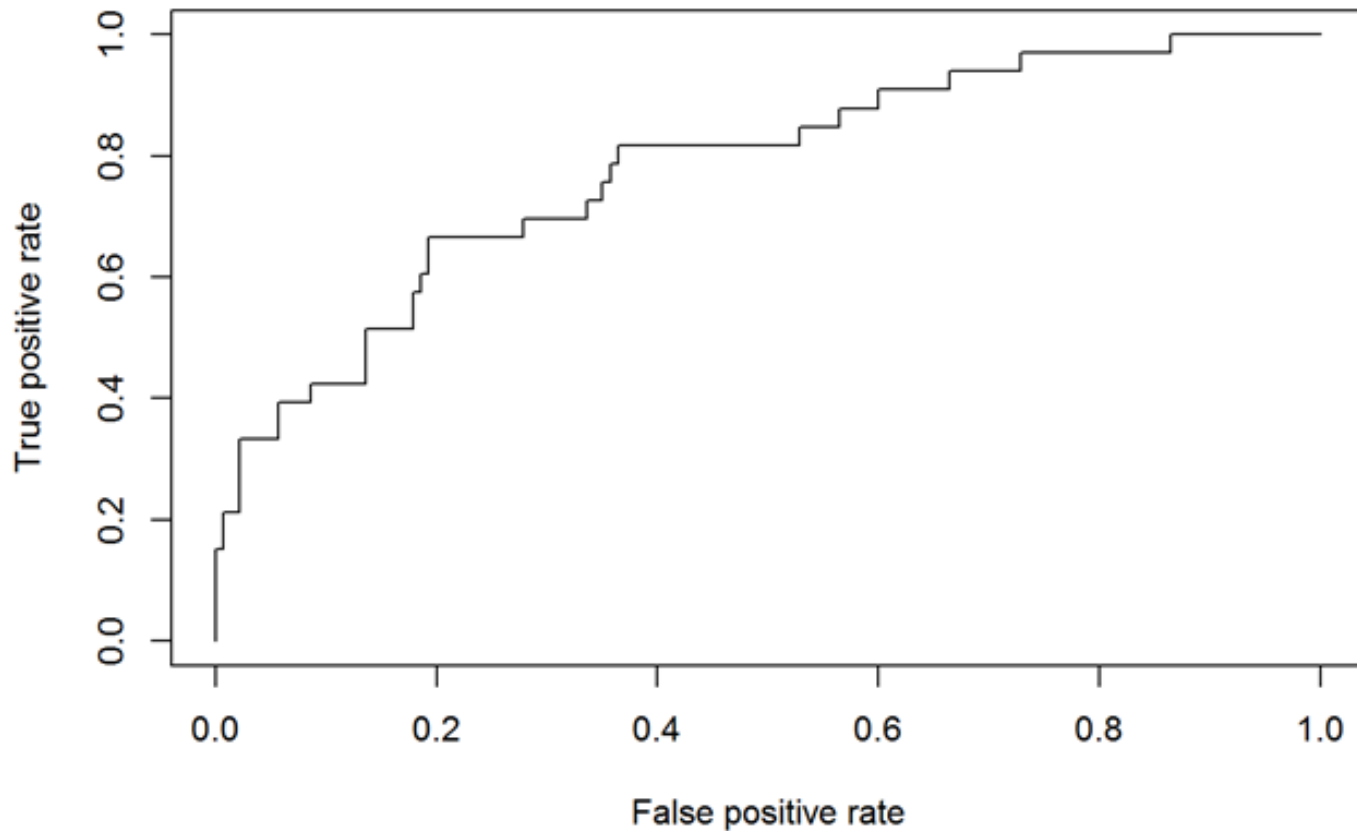
# Подбор параметров SVM

Performance of `svm`



AUC = 0.816

# ROC-кривая для предсказания на тестовой выборке для SVM



AUC = 0.78