

# Новый метод прогнозирования успешности лекарственного лечения онкологических заболеваний

Михаил Артемьев

ШАД, отделение "Анализ данных"

Научный руководитель: проф. Мучник И. Б.

23 мая 2018

- 1 Постановка задачи
- 2 Классификатор на основе детекторов классов
- 3 Оценка качества классификации
  - Лечение рака почки сорафенибом
  - Лечение рака лёгкого сорафенибом
  - Лечение рака груди паклитакселом
- 4 Заключение

- Решаемая задача: прогнозирование результата таргетной терапии рака.
- Имеющееся решение (N. Borisov et al.) довольно громоздкое, медленное.
- Цель: проверить простой и быстрый классификатор.

Оценка качества – Leave-One-Out ROC AUC.

Также смотрим на FNR, FPR – долю ошибок на ответчиках и на не ответчиках.

- Данные о пациентах – активации маркерных сигнальных путей и результат терапии.
  - Лечение рака почки сорафенибом – 28 объектов, 46 признаков
  - Лечение рака лёгкого сорафенибом – 37 объектов, 46 признаков
  - Лечение рака груди паклитакселом – до 508 объектов, 30 признаков
- Данные о клеточных линиях – активации маркерных сигнальных путей и значения  $IC_{50}$ . 227 дополнительных объектов.

# Классификатор на основе детекторов классов

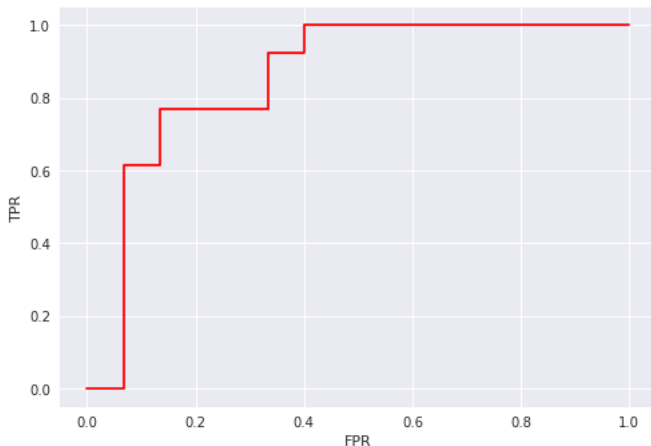
Наивный байесовский классификатор:

$$p(C, F_1, \dots, F_n) = p(C) \prod_{i=1}^n p(F_i|C)$$

$$\tilde{C} = \arg \max_C \sum_{i=1}^n \log p(F_i|C) + \log p(C)$$

Эвристики:

- монотонная классификация: оцениваем не логарифм плотности, а функцию распределения.
- монотонность признаков оцениваем по клеточным линиям.
- работаем в базисе из возрастающих и убывающих главных компонент.
- вес компоненты – модуль её коэффициента монотонности.

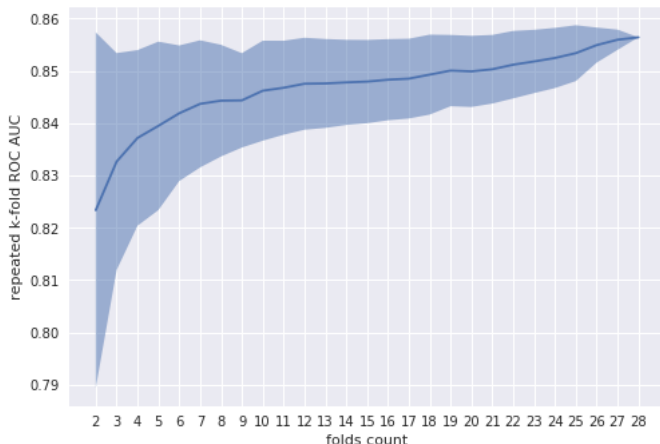


ROC AUC = 0.86, FPR = 0.07, FNR = 0.38.

N. Borisov et al.: ROC AUC = 0.77, 0.81, 0.82.

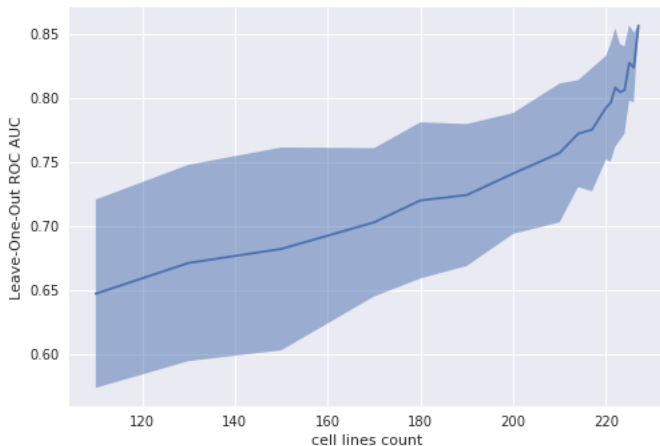
Робастность относительно выборки пациентов.

Зависимость ROC AUC от количества фолдов кросс-валидации:

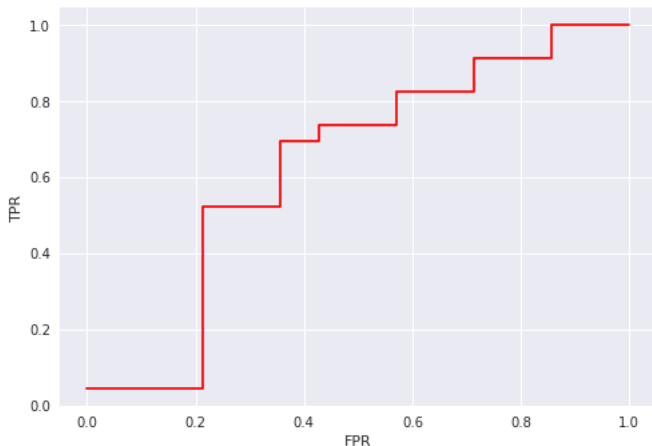


Робастность относительно выборки клеточных линий.

Зависимость ROC AUC от количества используемых клеточных линий:



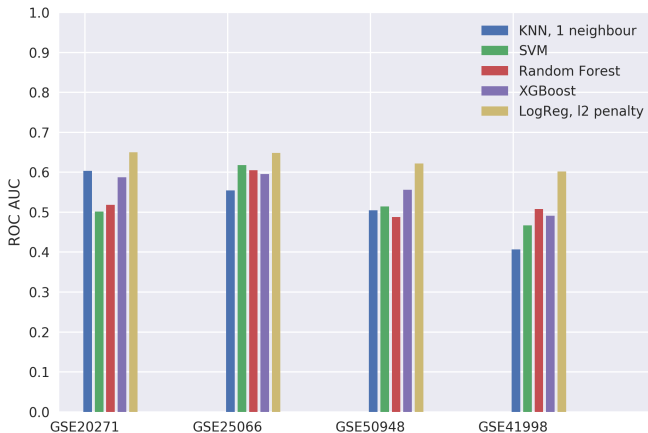




ROC AUC = 0.63, FPR=0.36, FNR = 0.35.

N. Borisov et al.: ROC AUC = 0.72, 0.77, 0.78.

Детекторный классификатор не работает: ROC AUC от 0.44 до 0.55.  
Работающий классификатор можно построить и без использования  
клеточных линий:



- Был построен новый классификатор для прогноза результата таргетной терапии.
- Построенный классификатор хорошо работает для некоторых пар заболевание-лекарство.
- Классификация устойчива к изменению обучающей выборки пациентов.
- При наличии большой выборки пациентов использование клеточных линий не оправданно.

# Приложение: подробное описание классификатора

- 1 Вычислить  $c_i$  – коэффициенты корреляции признаков с  $IC_{50}$
- 2 Найти главные компоненты выборки клеточных линий отдельно в линейной оболочке возрастающих признаков ( $c_i > 0$ ), отдельно – в линейной оболочке убывающих признаков ( $c_i < 0$ ).
- 3 Перевести выборки пациентов и клеточных линий в найденный базис главных компонент. Размерность при этом не понижается.
- 4 Оценить монотонность компонент  $m(f) = \mu(f, IC_{50})$ , где  $\mu$  – мера монотонности.
- 5 По каждой паре (компонента PCA, пациент обучающей выборки) построить детектор класса пациента – пороговое решающее правило.
- 6 Вес детектора  $\sim \frac{|m(f)|}{c}$ , где  $f$  – соответствующая компонента,  $c$  – количество объектов соответствующего класса в обучающей выборке. Объект классифицируется как ответчик, если  $\frac{p}{p+n} > \tau$ , где  $p, n$  – суммарный вес сработавших детекторов ответчиков и не ответчиков соответственно,  $\tau$  – порог классификации.