

# Решение конкурса

## Определите наличие осложнений у пациента

Илларионова Светлана

21 апреля 2016

# Задача

- По известным медицинским показателям, а так же генетическим данным о пациенте необходимо определить его состояние.
- В предоставленной обучающей выборке вектору признаков соответствует состояние: 0 — у пациента нет осложнений, 1 — есть.
- Для векторов из тестовой выборки требуется предсказать наличие осложнений.

# Описание данных

Для каждого пациента задано описание в виде 1330 признаков:

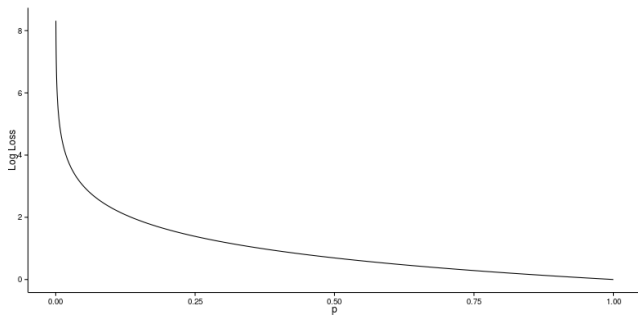
- 330 медицинских показателей
- 1000 генетических показателей (0, 1, 0.5)

Генетические признаки - каждая клетка содержит пару генов, для которых известно мутированы ли они (1), один ген мутирован (0.5), нет мутации(0)

# Функция потерь

Log Loss для случая двух классов:

$$\frac{1}{N} \sum_{i=1}^N [y_i \log p_i + (1 - y_i) \log (1 - p_i)]$$



# Функция потерь

Log Loss для случая двух классов:

$$\frac{1}{N} \sum_{i=1}^N [y_i \log p_i + (1 - y_i) \log (1 - p_i)]$$

- Вклад в сумму осуществляют только верно классифицированные объекты
- Log Loss сильно штрафует классификатор за уверенность в неправильном определении класса объекта
- Идеология Log Loss: нестрашно немного ошибиться, ужасно ошибаться глобально!

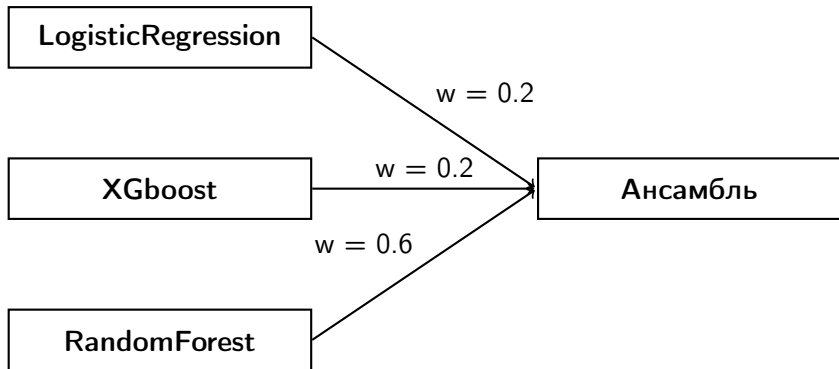
# Обработка пропусков

- Заполнение пропусков (min, max)
- Добавление признаков с информацией о пропусках
- Добавление столбца признаков "число пропущенных значений"

# Обработка признаков

- Максимальное число пропусков 150, если пропусков больше 130 - говорим, что не будет осложнений 0
- Отбор нужного числа признаков: 900 признаков

# Модель





Спасибо за внимание!

