Классификация композиций

по музыкальным инструментам

Николай Стулов

МФТИ

6 марта 2017 г.

Содержание

Цели и задачи

- Цель работы — построить эффективную модель распознавания музыкальных инструментов, работая с различными признаками и алгоритмами.
- Почему машинное обучение? Объем данных велик, при том, что их структура в пространстве признаков не очевидна.
- Прикладное применение — распознавание голоса.

Используемые методы

- Данные IRMAS (Dataset for Instrument Recognition in Musical Audio Signals)
- Язык Python
- Библиотеки Essentia, scikit-learn и XGBoost

python™

Используемые признаки

MonoLoader, ZeroCrossingRate, Energy

```
[ 0.04724265  0.05383465  0.06166265  0.06964324  0.07571642  0.07896665
  0.07866146  0.0746025   0.06825465  0.06016725  0.05041658  0.04054384
  0.03155614  0.02429273  0.01892148  0.01399274  0.00967437  0.00633259
  0.00274667 -0.00027467 -0.00282296 -0.0069277  -0.0117954  -0.01745659
 -0.02659688 -0.03881954 -0.05377362 -0.07271035 -0.09361553 -0.11378826
 -0.13139744 -0.14300974 -0.14664143 -0.14194158 -0.12822351 -0.10718101
 -0.08163701 -0.0535905  -0.02644429 -0.00265511  0.01777703  0.0339671
  0.0450911   0.05165258  0.05334635  0.05020295  0.04330577  0.03323466
  0.02064577  0.00622578 -0.00973541 -0.02624592 -0.04161199 -0.05482651
 -0.06537065 -0.0728019  -0.07673879 -0.07710502 -0.07438887 -0.06953642
 -0.06346324 -0.05742057 -0.05323954 -0.05223243 -0.05421613 -0.05882443
 -0.06640828 -0.07599109 -0.08606219 -0.09677419 -0.10805078 -0.11893064
 -0.12929167 -0.138493   -0.14406262 -0.14291817 -0.13235877 -0.11197241
 -0.08478042 -0.05439924 -0.02433851  0.0013123   0.02064577  0.03411969
  0.04177984  0.0446791   0.04486221  0.04289377  0.03904843  0.03456221
  0.03035066  0.02804651  0.02949614  0.03460799  0.0426954   0.05327006
  0.06483657  0.0757622   0.08479568  0.09091464]
```

Рис. 1: Представление после открытия wav-файла экстрактором MonoLoader

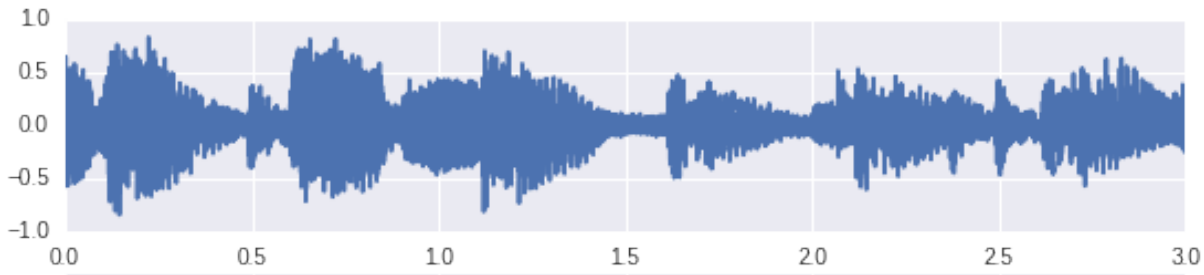


Рис. 2: Графическая интерпретация массива выше

Используемые признаки

Spectrum, Windowing

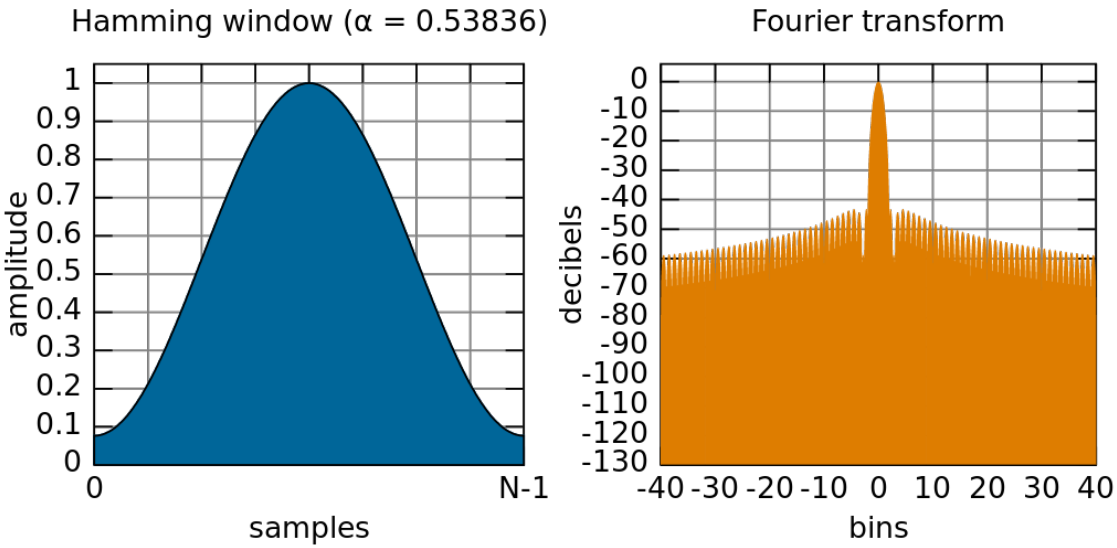


Рис. 3: Окно Хэмминга и его спектральное разложение

$$w(n) = \frac{25}{46} - \frac{21}{46}\cos\left(\frac{2\pi n}{N-1}\right) \approx 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right)$$

Используемые признаки

Centroid, CentralMoments

- Спектральный центроид — «центр масс»спектра (здесь $x(n)$ — средняя амплитуда, а $f(n)$ — главная частота):

$$C = \frac{\sum_{n=1}^{N-1} f(n)x(n)}{\sum_{n=1}^{N-1} x(n)}$$

- n-ый центральный момент — статистическая характеристика:

$$\mu_n = \mathbb{E}[(X - \mathbb{E}(X))^n] = \int_{-\infty}^{\infty} x^n f(x) dx$$

Используемые признаки

Mel-frequency Cepstrum Coefficients

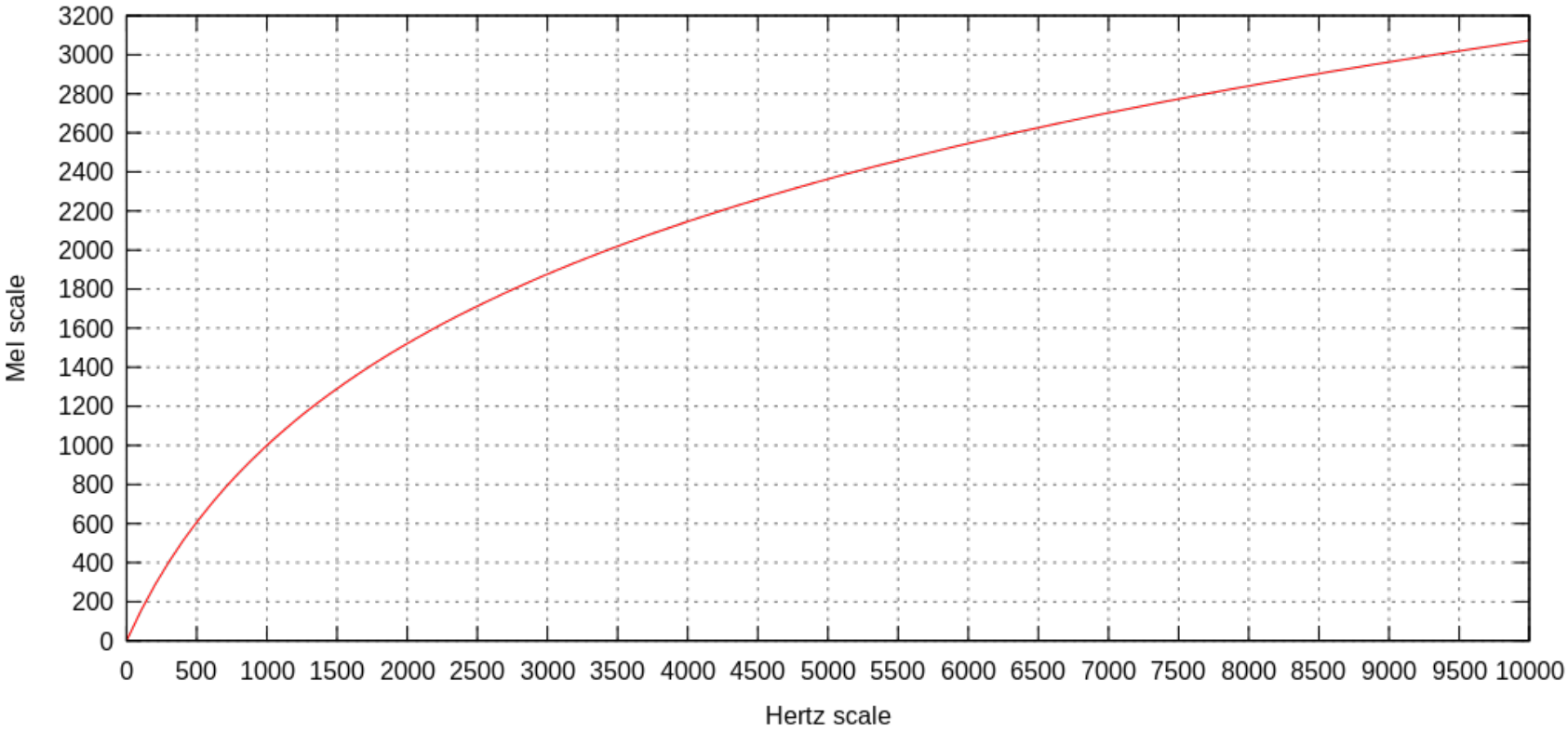


Рис. 4: Шкала частот Мела

Практический эксперимент

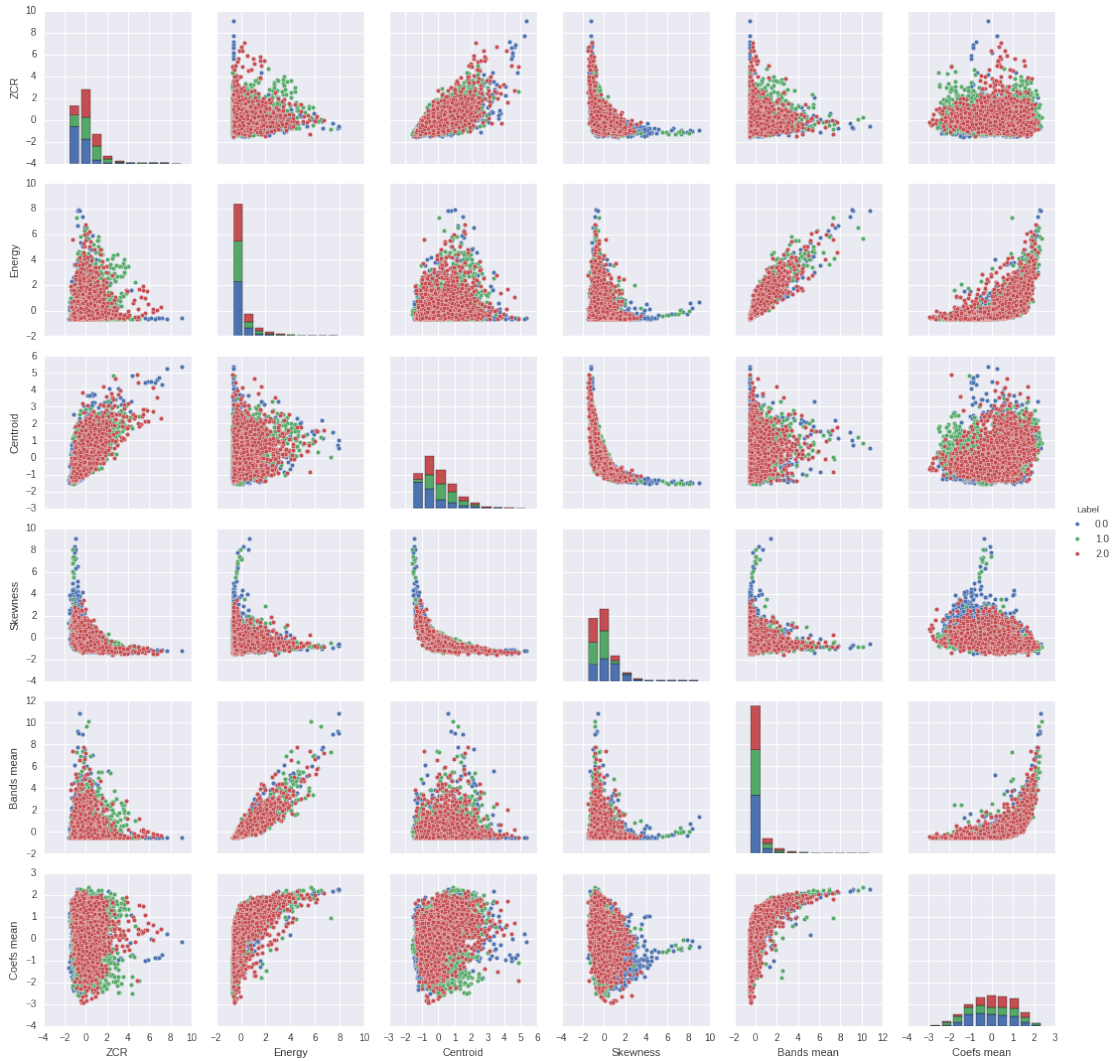


Рис. 5: Двумерные срезы в пространстве признаков

- Точность (accuracy):

$$accuracy = \frac{\sum_{k=1}^{n} [a(x_k) = y_k]}{n}$$

- Кросс-энтропия (logloss):

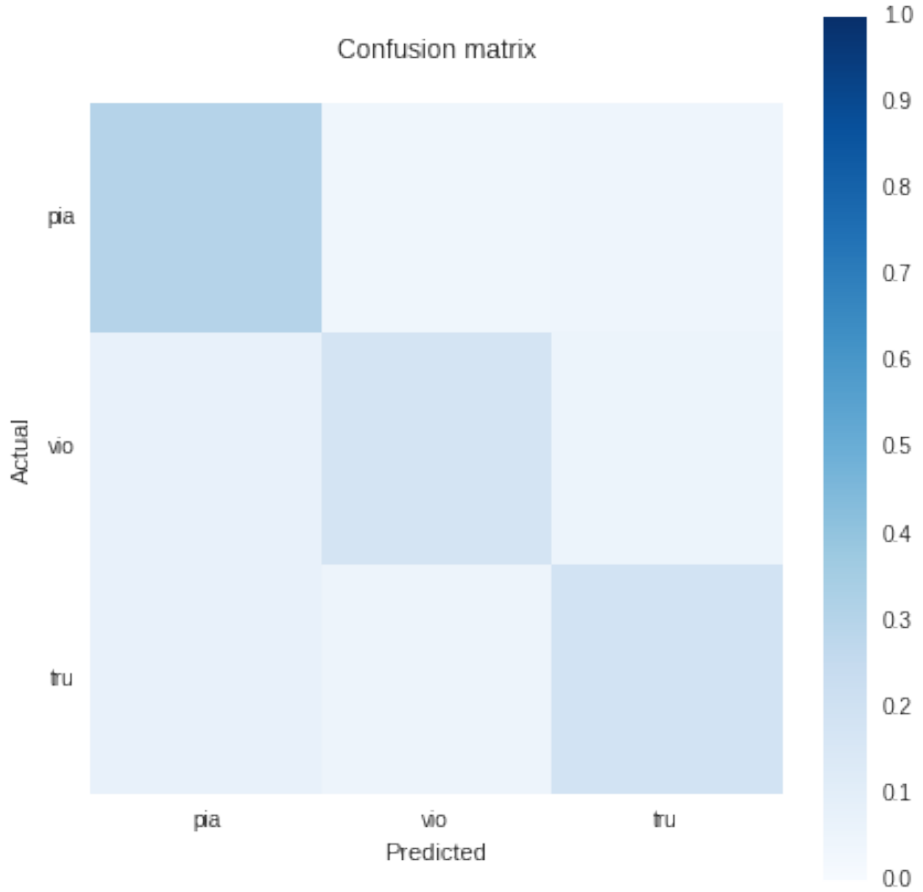
$$logloss = -\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{M} y_{ij} \log p_{ij}$$

Практический эксперимент

LogisticRegression

LogisticRegression

- Параметры: $\alpha = 0.6$
- Точность: 0.665
- Кросс-энтропия: 0.834

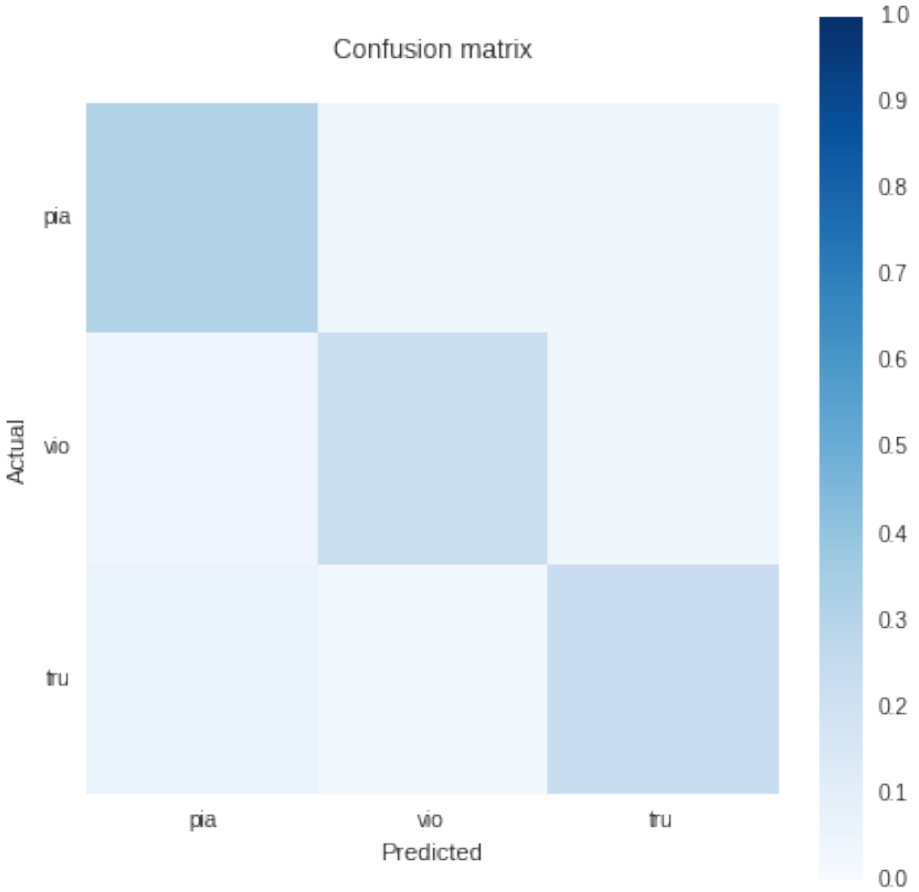


Практический эксперимент

RandomForest

RandomForest

- Параметры: 43 дерева, максимальная глубина 12, не более 3 объектов в листьях
- Точность: 0.765
- Кросс-энтропия: 0.625



Практический эксперимент

Multilabel RandomForest

В процессе разработки был опробован подход с множеством ответов на основании One-Vs-Rest RandomForest со следующими результатами:

- Точность (строгая): 0.668
- Точность (толерантная): 0.846

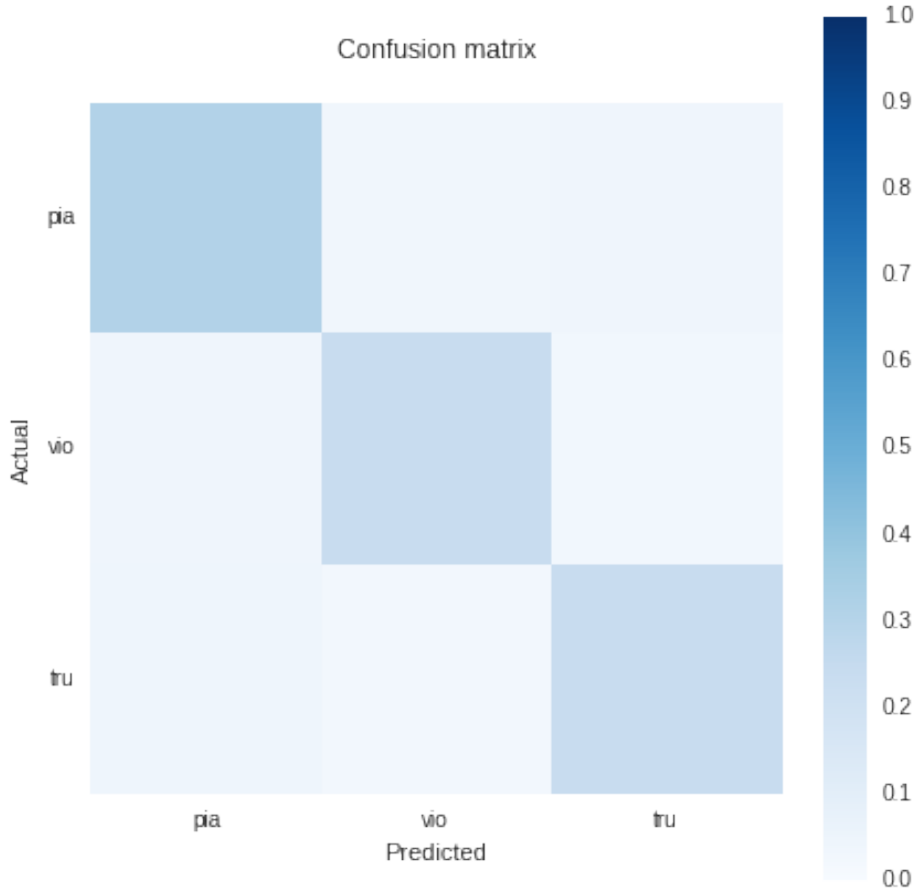
Анализ результатов показал, что в большинстве случаев алгоритм предсказывает ровно одну метку в ответе, поэтому был сделан вывод о том, что гипотеза не верна.

Практический эксперимент

XGBoost

XGBoost

- Параметры: 40 деревьев, максимальная глубина 15
- Точность: 0.787
- Кросс-энтропия: 0.531



Практический эксперимент

Тестирование

Истинные метки

- Фортепиано + труба (0 и 2)
- Фортепиано (0)
- Электрогитара (-)
- Фортепиано + труба (0 и 2)
- Фортепиано + скрипка (0 и 1)

Предсказанные метки

- $[1, 1, 1, 1, 1]$
- $[2, 2, 1, 2, 1]$
- $[1, 1, 1, 1, 1]$
- $[0, 2, 2, 2, 2]$
- $[1, 0, 0, 2, 1]$

Выводы и результаты

- Проведена работа с различными признаками, характеризующими временной ряд
- Обучены три высокоуровневых алгоритма и проведен сравнительный анализ их качества
- Получено качество, более чем вдвое превышающее случайный выбор класса
- Проверена гипотеза о затененности одних инструментов другими

Об алгоритмах

Preprocessing

- Масштабирование данных — влияет на качество некоторых алгоритмов. Выполнено вычитанием среднего по столбцу и делением на стандартное отклонение
- Стратификация отложенной выборки — гарантия того, что в отложенной выборке в равной степени присутствуют все классы

В случае многоклассовой классификации функция распределения и оптимизируемый функционал принимают следующий вид

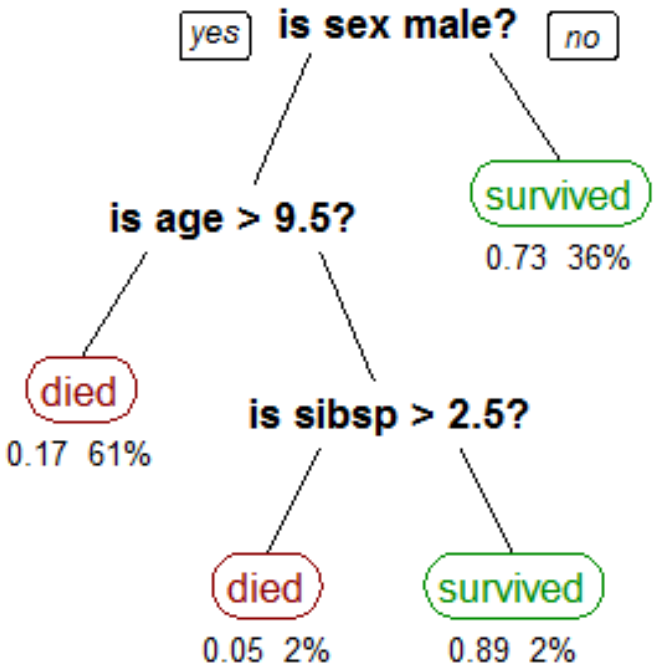
$$h_w = softmax = \frac{1}{\sum_{k=1}^{m} \exp\left(-w_k^T x\right)}$$

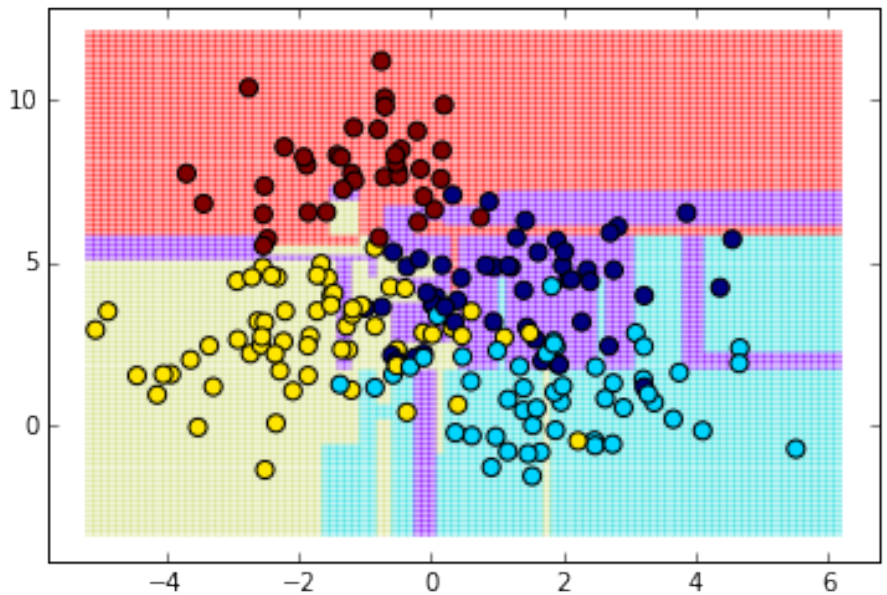
$$L(x, y, w) = -\sum_{i=1}^{n}\sum_{k=1}^{m}\left([y_i = k] \log \frac{\exp\left(w_k^T x_i\right)}{\sum_{j=1}^{m} \exp\left(-w_j^T x_i\right)}\right) + \alpha \sum_{k=1}^{n} |w_k|$$

Об алгоритмах

RandomForest

Композиция алгоритмов, обучение независимо, базовый алгоритм — решающее дерево. Ответ берется как среднее ответов базовых алгоритмов. Для хорошей работы необходима независимость обучения базовых алгоритмов.





Композиция алгоритмов, обучение последовательное. Каждый следующий базовый алгоритм пытается приблизить результат предыдущего к верному ответу.

$$s = \left(-\frac{\partial L}{\partial z}|_{z=a_{N-1}(x_i)} \right)_{i=1}^{\ell}$$

$$b_N(x) = \arg\min_b \sum_{i=1}^{\ell} (b(x_i) - s_i)^2$$

Список источников

[1] «A Large Set of Audio Features for Sound Description», Geoffroy Peeters, 2004

[2] MTG Workshops

[3] Essentia Algorithms Reference

[4] Coursera.org

[5] Machinelearning.ru

[6] Wikipedia.org