Краткосрочное предсказание музыкальных произведений с использованием последовательностей аккордов

Михаил Матросов[1,2,3],
научный руководитель В. В. Стрижов[1,2,3],
консультант Антон Матросов[1]

[1]Московский Физико-Технический Институт
[2]Сколковский Институт Науки и Технологий
[3]Вычислительный Центр Российской Академии Наук

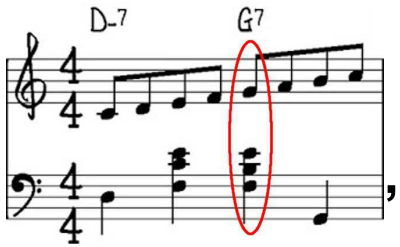
Вычислительный Центр Российской Академии Наук
Июнь 2015, Москва, РФ

Цель исследования

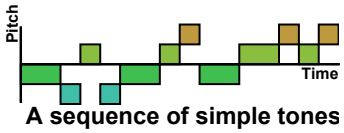
Предсказать следующий элемент в последовательности аккордов в теоретико-групповом представлении, не учитывая темпоральную составляющую (арпеджио, длительности, паузы).

Новизна: более точное (больше информации о каждом аккорде) представление музыкальной последовательности.

Метод: композиция байесовских классификаторов.



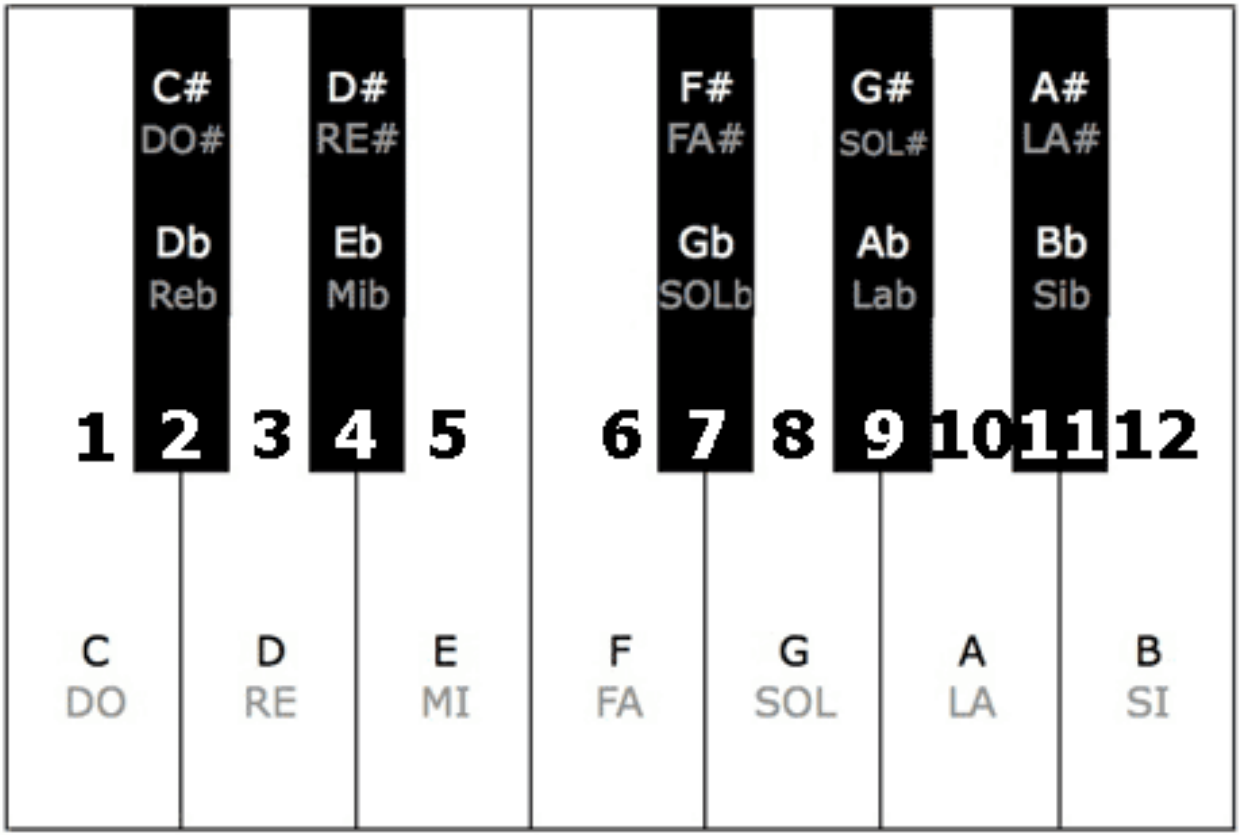
, not



Тизер: 92.5% Хэмминг-сходство (58% по-аккордно) между предсказанием и оригинальной мелодией из тестовой выборки.

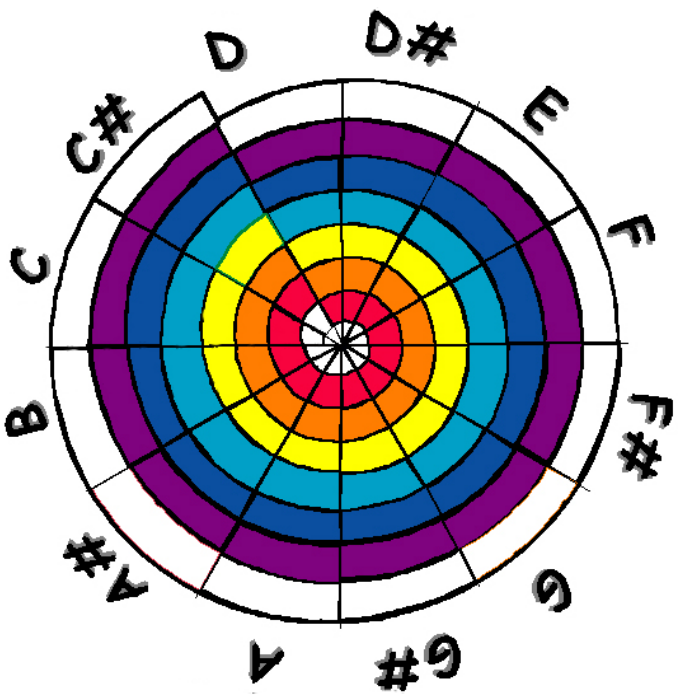
Октава

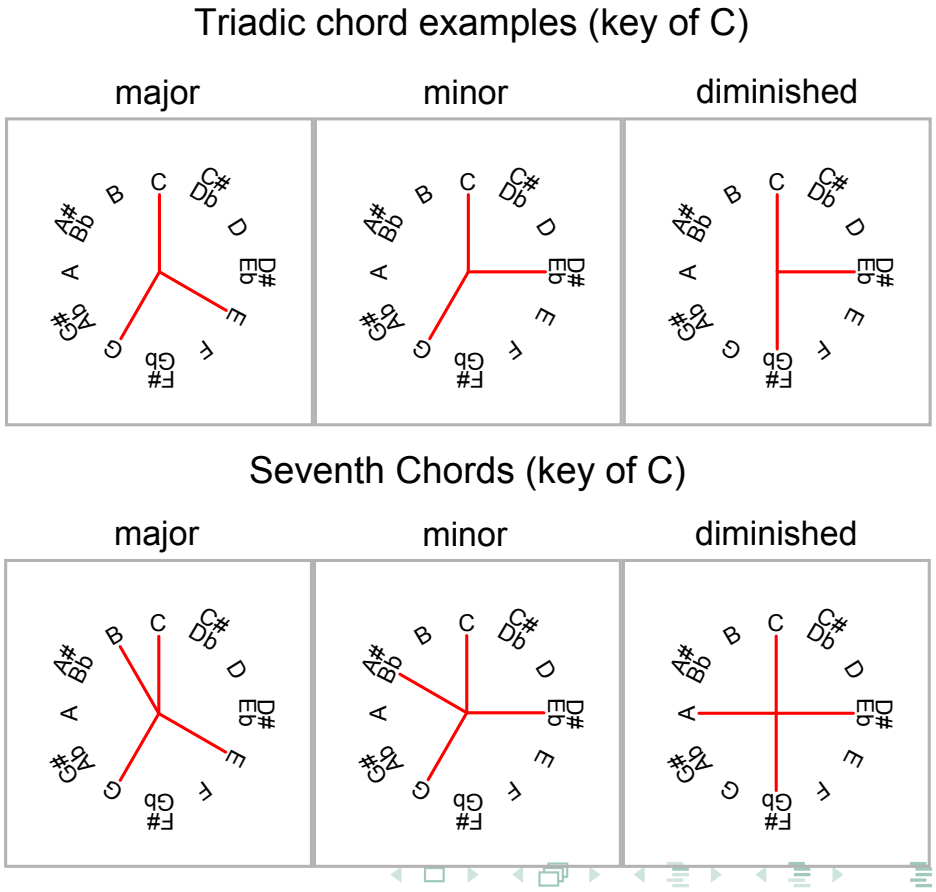
Октава состоит из 12 полутонов. Музыкант может сыграть все ноты в пределах одной октавы, при этом звучание мелодие изменится не значительно. По-этому мы представляем аккорды в пределах только одной октавы.



Аккорды

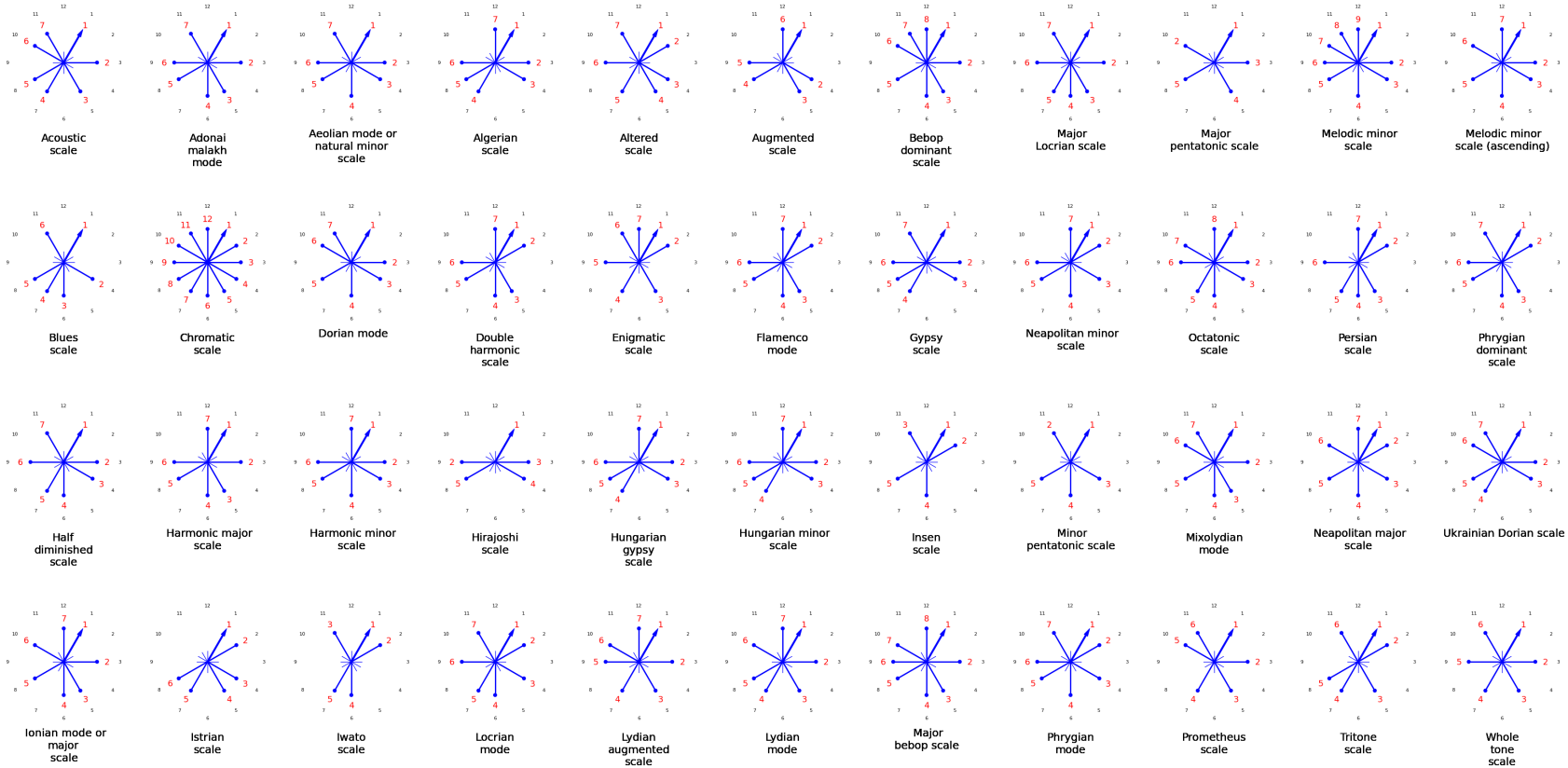
Каждый **аккорд** содержит от 1 до 12 одновременно звучащих полутонов. Аккорд может быть транспонирован на несколько полутонов вверх или вниз. Его так же можно изобразить на круге, при этом вращение эквивалентно изменению высоты тона.





Созвучие и базовый тон

Ниже избражены еще несколько аккордов. Существует 351 возможное созвучие.



Каждое из них может быть сыграно в 12 различных ключах, кроме нескольких симметричных случаев (довольно редких). Итого $2^{12} - 1 = 4095$ возможных аккордов.

Представление аккордов

Мелодия представляется в виде последовательности аккордов (целых от 1 до $2^{12} - 1$), i здесь и далее обозначает время.

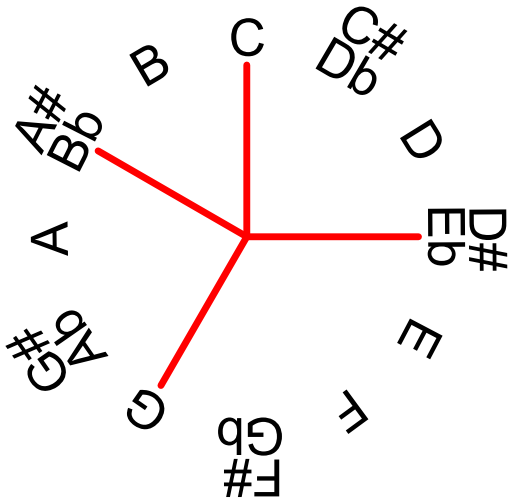
$$\underset{\text{melody}}{\mathbf{c}} = \underset{\text{sequence}}{\{c_i\}}, \; c_i \in \mathbf{C},$$

$\mathbf{C} = \{1, 2, 3, \ldots, 4095\}$ — пространство **аккордов**.
Каждый аккорд имеет свою форму (строй $s \in \mathbf{S}$) и базовый тон ($z \in \mathbf{Z}_{12}$). Таким образом, вся мелодия представляется как последовательность пар (s, z):

$$\underset{\text{chord}}{c} = \underset{\text{strum}}{s} \times \underset{\text{key}}{z},$$

$$\underset{\text{melody}}{\mathbf{c}} = \underset{\text{pairs}}{\{(s, z)_i\}},$$

Произведение означает транспозицию.
$\mathbf{S}$ — множество уникальных **строев**, 351 элементов.
$\mathbf{Z}_{12}$ — группа вычетов по модулю 12.



Последовательность элементов

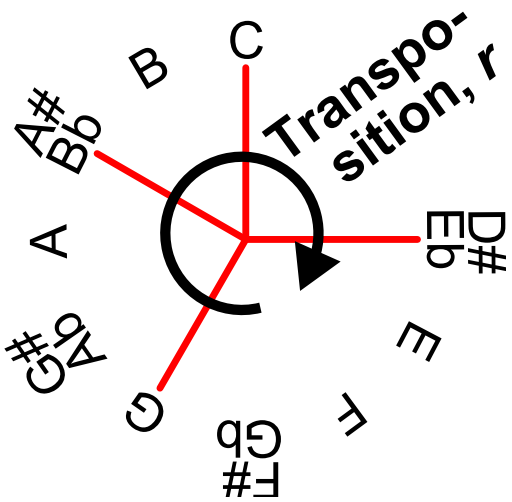
Мелодия может быть транспонирована, так что ноты нужно соизмерять с предшествующими:

$$r_i = z_i - z_{i-1} \mod 12,\ r_1 = z_1.$$

Пара (s, r) называется **элементом** и обозначается $x \in \mathbf{E}$.

$$\mathbf{C} = \mathbf{S} \times \mathbf{Z}_{12} = \mathbf{S} \times \mathbf{R}_{12} = \mathbf{E},$$

C — пространство **аккордов**, $N = 4095$,
S — множество уникальных **строев**, $N = 351$,
$\mathbf{Z}_{12}$ — вычеты по модулю 12, $N = 12$,
$\mathbf{R}_{12}$ — смежные разности Z_{12}, $N = 12$,
E — пространство **элементов**, в котором мы предсказываем.
Пространство — это множество с операцией транспонирования (изменения высоты тона).



N-граммы

N-грамм — подпоследовательность из N элементов $x \in \mathbf{E}$ данной последовательности $\mathbf{x} = \{x_i\}$. N-грамм размера 1 — это просто один элемент $x \in \mathbf{E}$. Например:

$$\mathbf{x}_i^N = \{x_i, x_{i+1}, \ldots, x_{i+N-1}\}.$$

В данной работе N-граммы используются как характеризующие вектора, описывающие текущую точку в музыкальной последовательности.

Набор данных Midi50k

50 000 случайных midi-файлов были собраны в интернете.
Каждый midi-файл преобразован в последовательность аккордов $\mathbf{c} = \{c_i\},\ c_i \in \mathbf{C}$ по следующему алгоритму:

- открыть midi-файл как piano roll,
- отбросить перкуссию,
- квантовать с частотой $2 \cdot tempo$,
- отбросить номер октавы ($pitch = pitch \mod 12$).

Средний midi-файл содержим **600 аккордов** аккордов, что дает суммарно **30 миллионов** аккордов.
Мелодия — последовательность элементов (индекс это время): $\mathbf{x} = \{x_i\},\ x_i \in \mathbf{E}$.
$\mathbb{X}$ набор мелодий: $\mathbb{X} = \{\mathbf{x}_j\}$.

Разбиение выборки

Для оценивания алгоритма полный набор Midi50k $\mathbb{X}_0$ делился на две части разного размера. Каждый раз деление производилось случайным образом — из набора выбиралось подмножество заданного размера M.

$$\mathbb{X}_{\text{training}} \subset \mathbb{X}_0,$$

$$|\mathbb{X}_{\text{training}}| = M.$$

Для проверки алгоритма использовалась оставшаяся часть набора:

$$\mathbb{X}_{\text{testing}} = \mathbb{X}_0 \setminus \mathbb{X}_{\text{training}}.$$

Постановка задачи в терминах минимизации ошибки

Для тренировочного набора данных $\mathbb{X}$ найти вектор параметров алгоритма $\mathbf{w}$, минимизирующий функцию ошибки:

$$\hat{\mathbf{w}} = \arg \min_{\mathbf{w} \in \mathbb{R}^{2K}} S(\mathbf{w}, \mathbb{X}).$$

Функция предсказания

Предсказание делается по взвешенной сумме классификаторов:

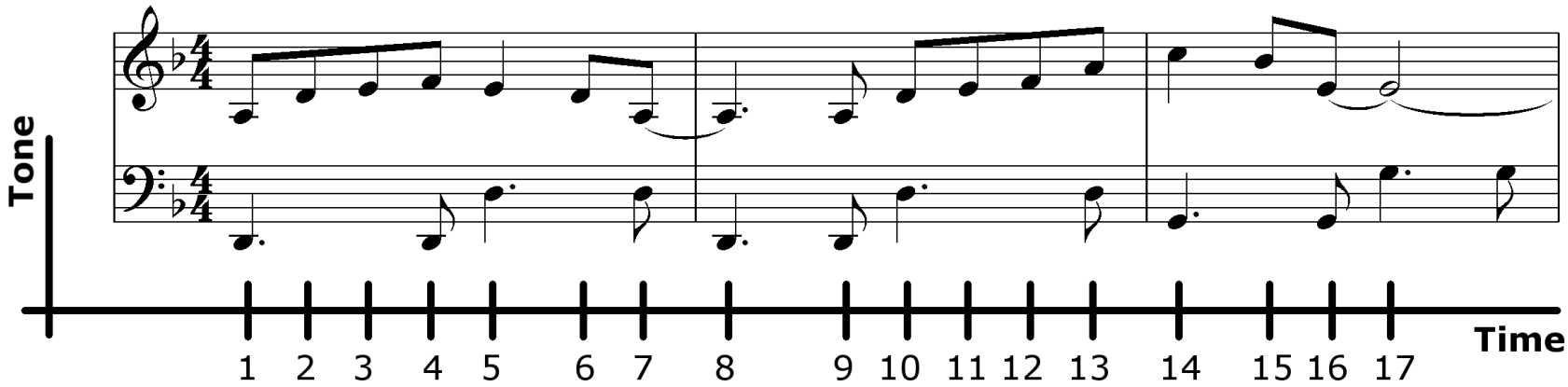
$$x_{i+1} = f(\mathbb{X}, \{x_1, \ldots, x_i\}, \underset{\substack{=\{u_k, v_k\},\\ k=1,\ldots K}}{\mathbf{w}}) = \arg\max_{e \in \mathbf{E}} \sum_{k=1}^{K} (A_{ek} u_k + B_{ek} v_k), \tag{1}$$

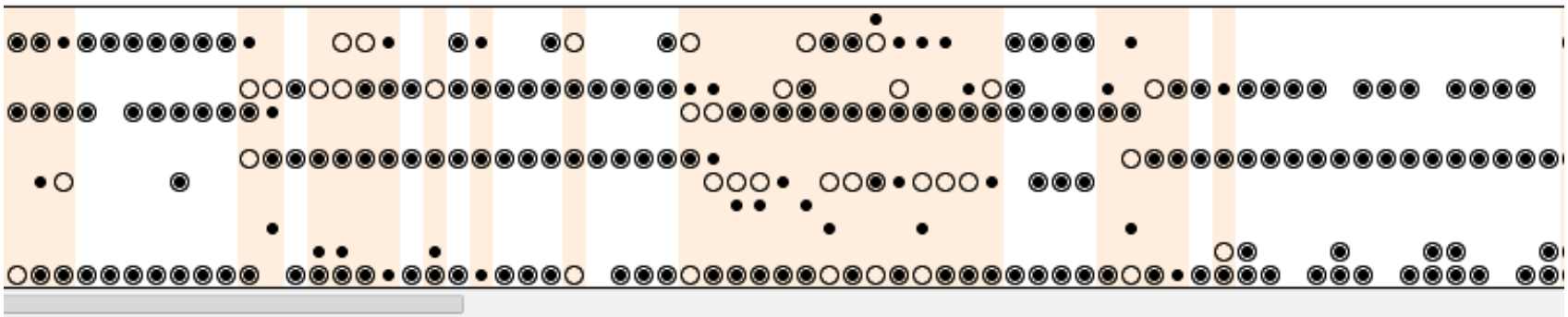
$$A_{ek} \propto N\left(\underbrace{\{x_{i-k+2}, \ldots, x_i, e\}}_{k\text{-gram}} \text{ in } \underbrace{\mathbb{X}}_{\text{Training set}} \right),$$

$$B_{ek} \propto N\left(\underbrace{\{x_{i-k+2}, \ldots, x_i, e\}}_{k\text{-gram}} \text{ in } \underbrace{\{x_1, \ldots, x_i\}}_{\text{Part of melody before i+1}} \right),$$

где "$\propto$"означает, что A_{*k} и B_{*k} нормированы на L1, $x_i \in \mathbf{E}$, $N(g \text{ in } dataset)$ — количество k-граммов $g \in \mathbf{E}^k$ в наборе, $\mathbf{w} = \{u_k, v_k | k = 1, \ldots K\} \in \mathbb{R}^{2K}$ — вектор параметров, K — сложность модели (максимальная длина N-граммов).

Функция ошибки





Пустые круги — истинные полутона, точки — предсказанные, ошибки подсвечены, горизонтальная ось — время.

Функция ошибки

Функция f это классификатор, предсказывающий следующий элемент. Тогда функция ошибки (i означает временной интервал):

$$S(\mathbf{w}, \mathbb{X}) = \sum_{\mathbf{x} \in \mathbb{X}} \sum_{i=1}^{N_{\mathbf{x}}-1} \left[x_{i+1} \neq f(\mathbb{X}, \underbrace{\{x_1, \ldots, x_i\}}_{\text{Prev. part of the melody}}, \mathbf{w}) \right].$$

Brackets stand for 1 if the statement inside is true and 0 if false.
Квадратные скобки дают 1, если выражение внутри истинно, и 0, если ложно.
$\mathbb{X}$ — набор мелодий, $\mathbf{w} \in \mathbb{R}^{2K}$ — вектор параметров.

Сглаженная функция ошибки

Ввведем также сглаженную функцию ошибки, чтобы избавиться от возможных проблем со стабильностью процесса минимизации:

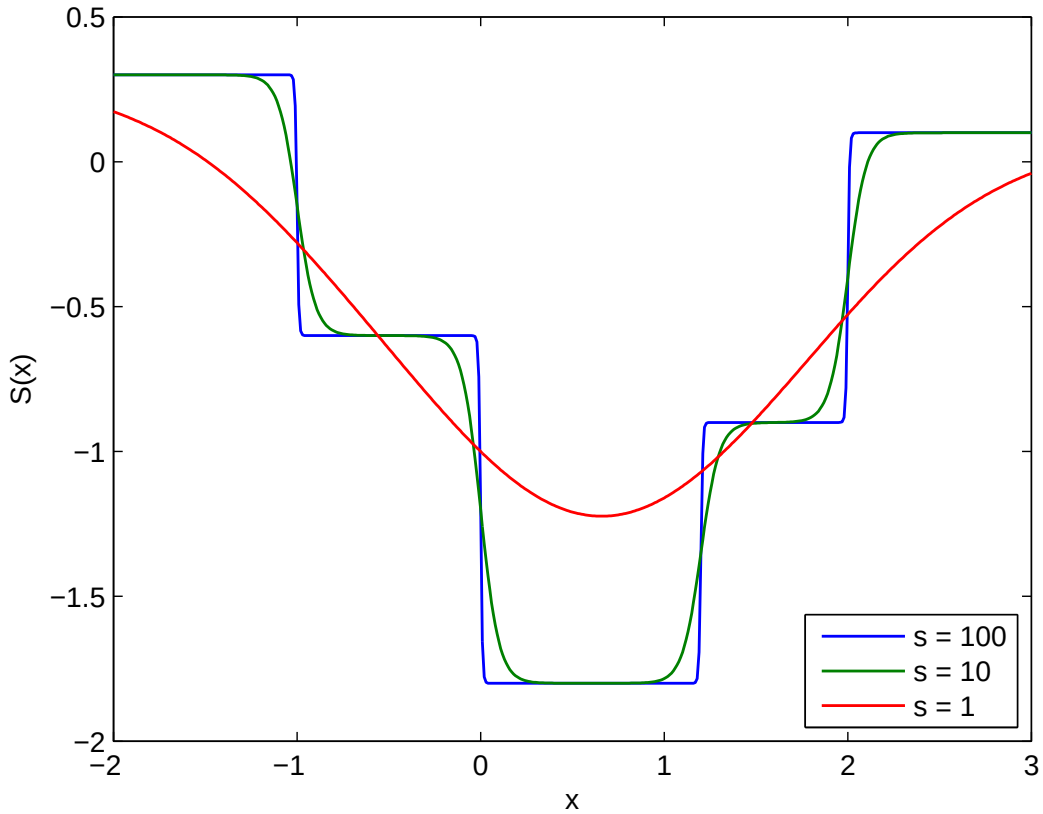
$$S_{smooth}(\mathbf{w}, \mathbb{X}) = \sum_{\mathbf{x} \in \mathbb{X}} \sum_{i=1}^{N_{\mathbf{x}}-1} B\left(x_{i+1}, \hat{f}(\mathbb{X}, \underbrace{\{x_1, \ldots, x_i\}}_{\text{Prev. part of the melody}}, \mathbf{w})\right),$$

$$B(x, \hat{f}) = \tanh(-s \frac{M_1 - M_2}{|M_1 + M_2|} \cdot e),$$

где M_1 и M_2 — первый и второй наибольшие компоненты распределения вероятностей $\hat{f}$ из f в (1), s — масштабный параметр, $e = 1$ если предсказание верно и -1 иначе.

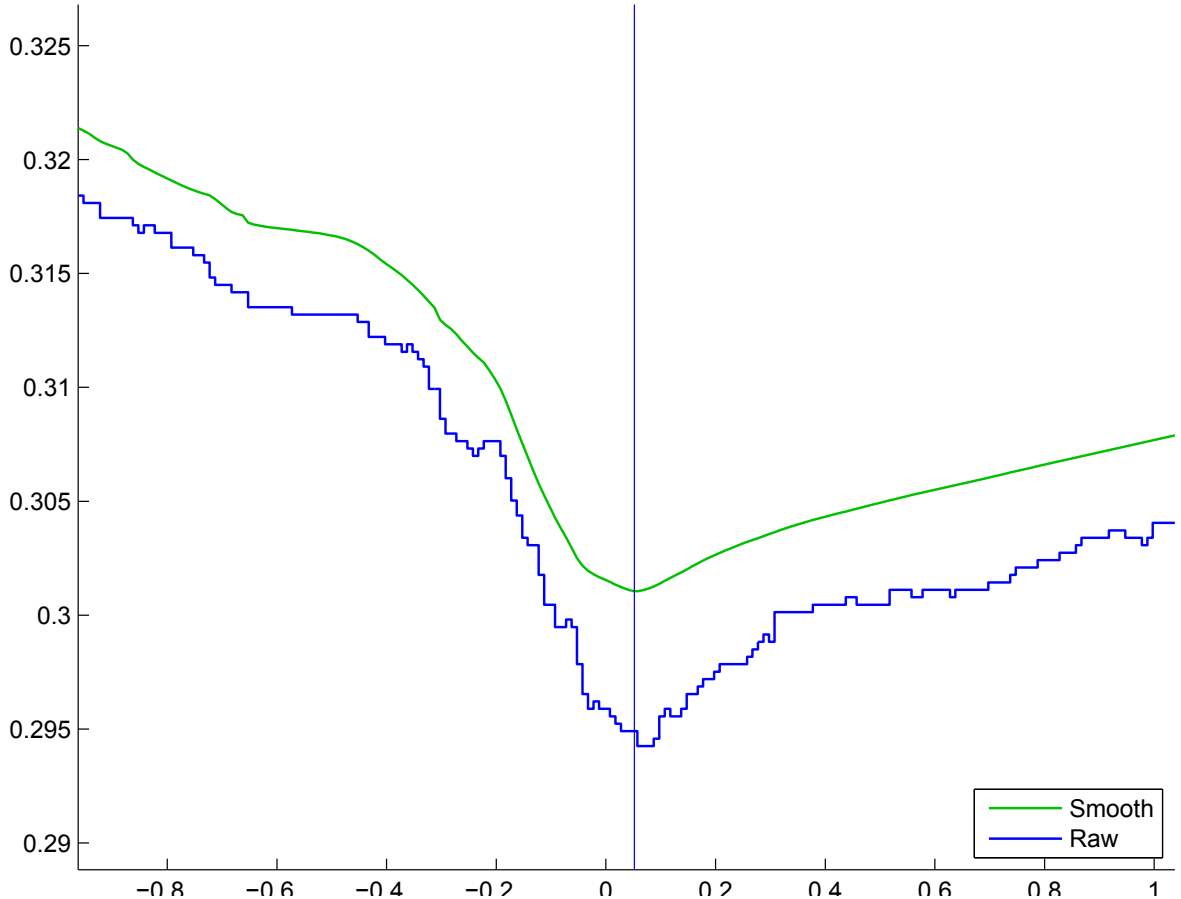
Сглаженная функция ошибки

Меньший параметр масштаба s функции ошибки S_{smooth} делает ступеньки менее заметными, тем самым процесс оптимизации проще, но менее точным.



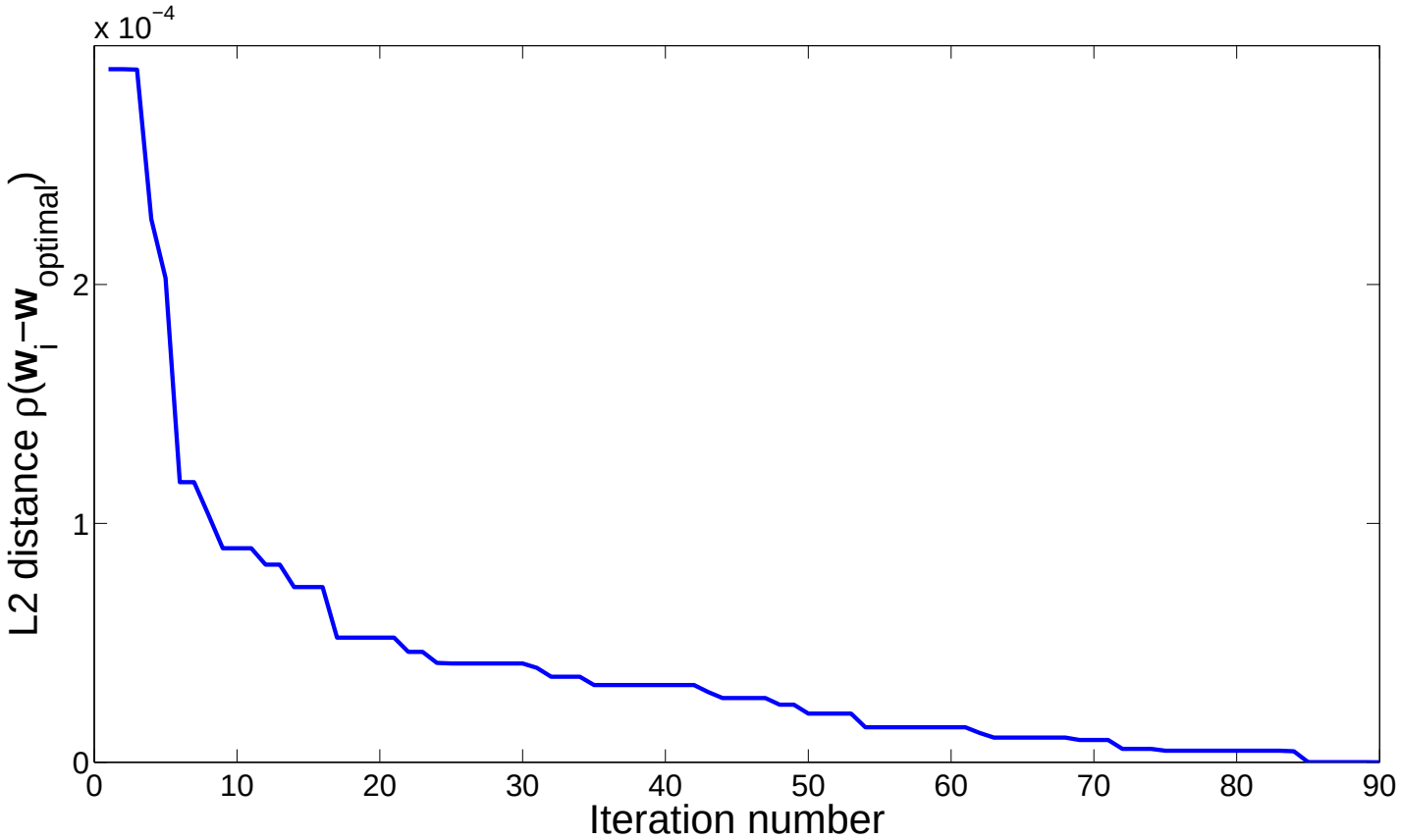
Сглаженная функция ошибки

Сглаженная $S_{smooth}(\mathbf{w}, \mathbb{X})$ и исходная $S(\mathbf{w}, \mathbb{X})$ функции ошибки в большинстве случаев имеют близкие минимумы.



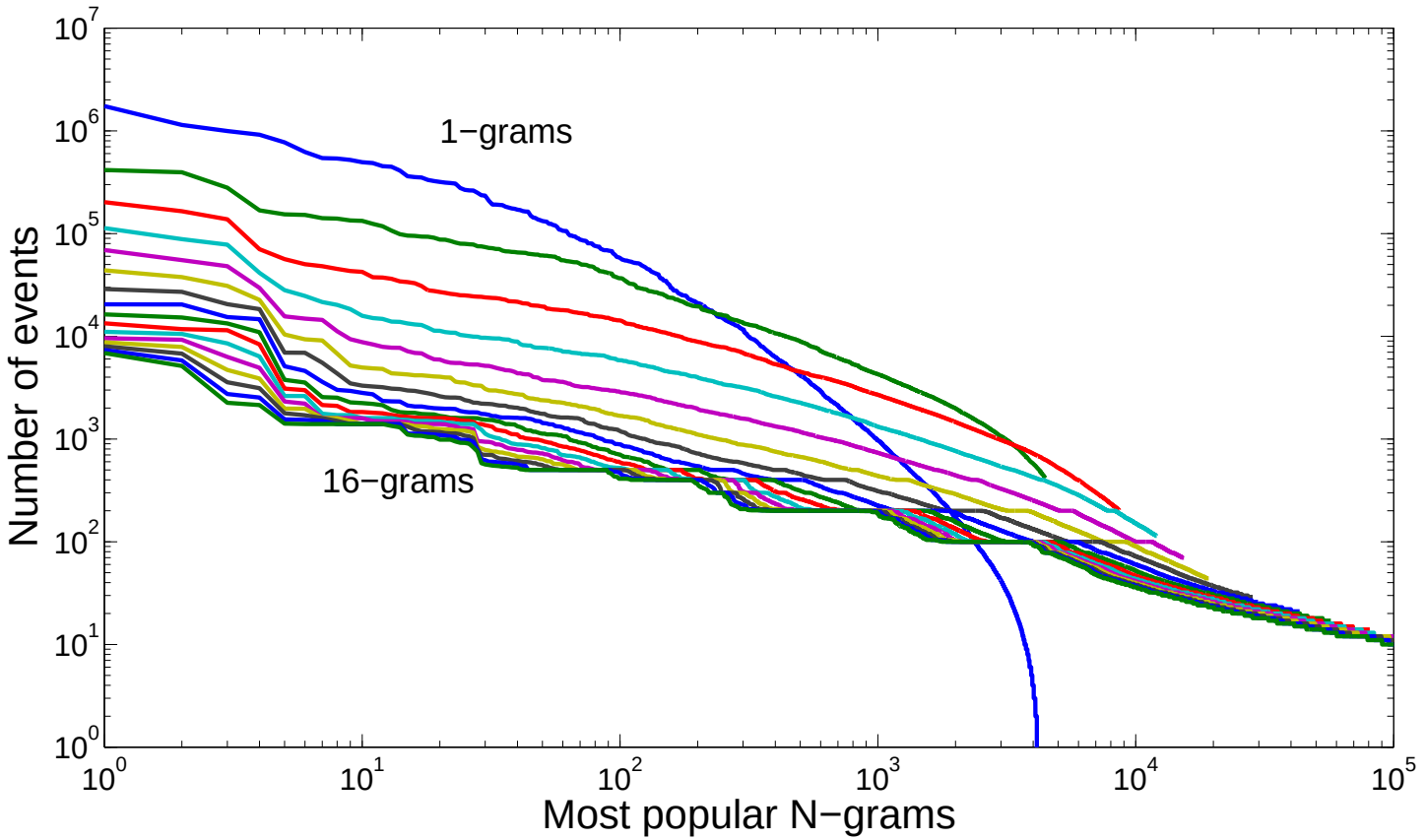
Стохастический градиентный спуск

Предсказание и функция ошибки вычисляется до 100 часов. Лучше делать **маленькие шажки** для небольших частей тренировочного множества — случайное подмножество из **100 мелодий**. Для сравнения, в обычном наборе может быть до 10 000. Так оптимизацию можно провести быстрее.



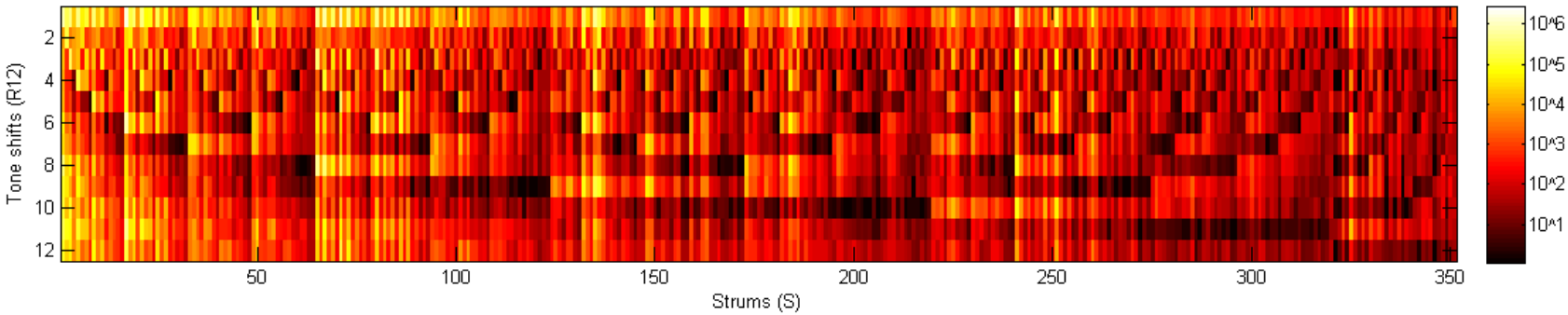
Музыка обладает свойствами естественного языка

Распределение N-граммов по частоте для различных N. Число событий — это число появлений N-грамма в Midi50k. Наклон равен -0.6, распределение похоже на распределение слов в естественном языке (закон Ципфа).



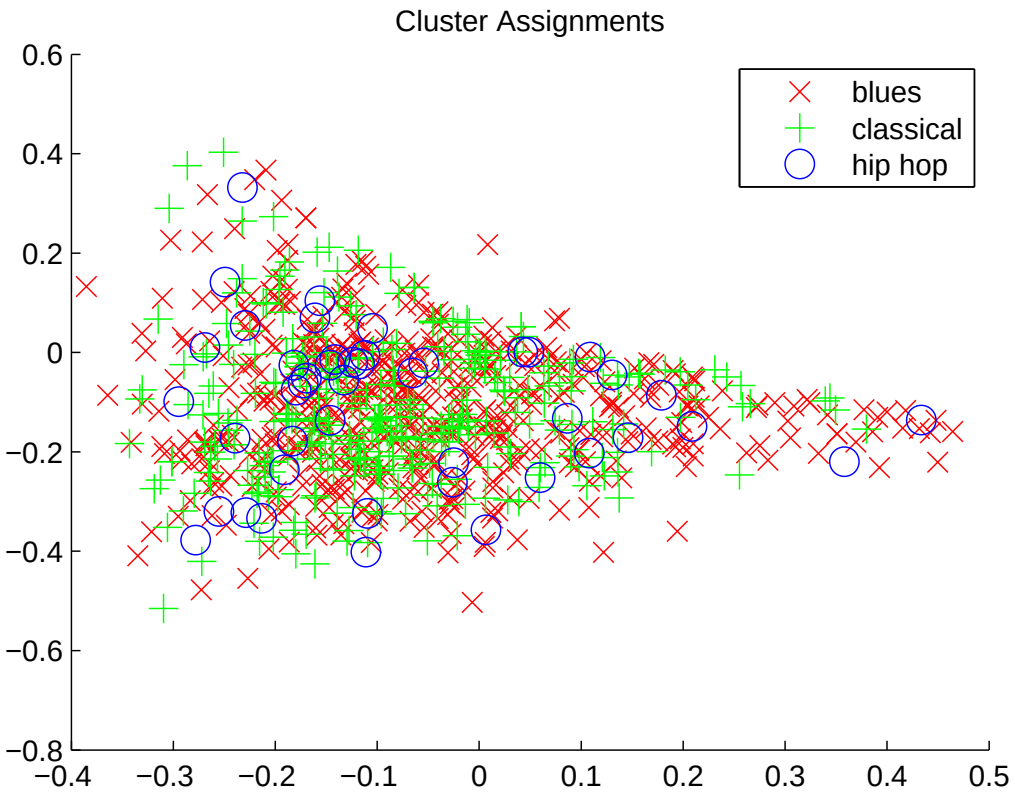
Вероятность элементов **E**

Карта распределения частот элементов **E**. Число событий — это число появлений N-грамма в Midi50k. Порядок строев (по горизонтали) — произвольный — результат представления аккорда в виде пары (s, r).



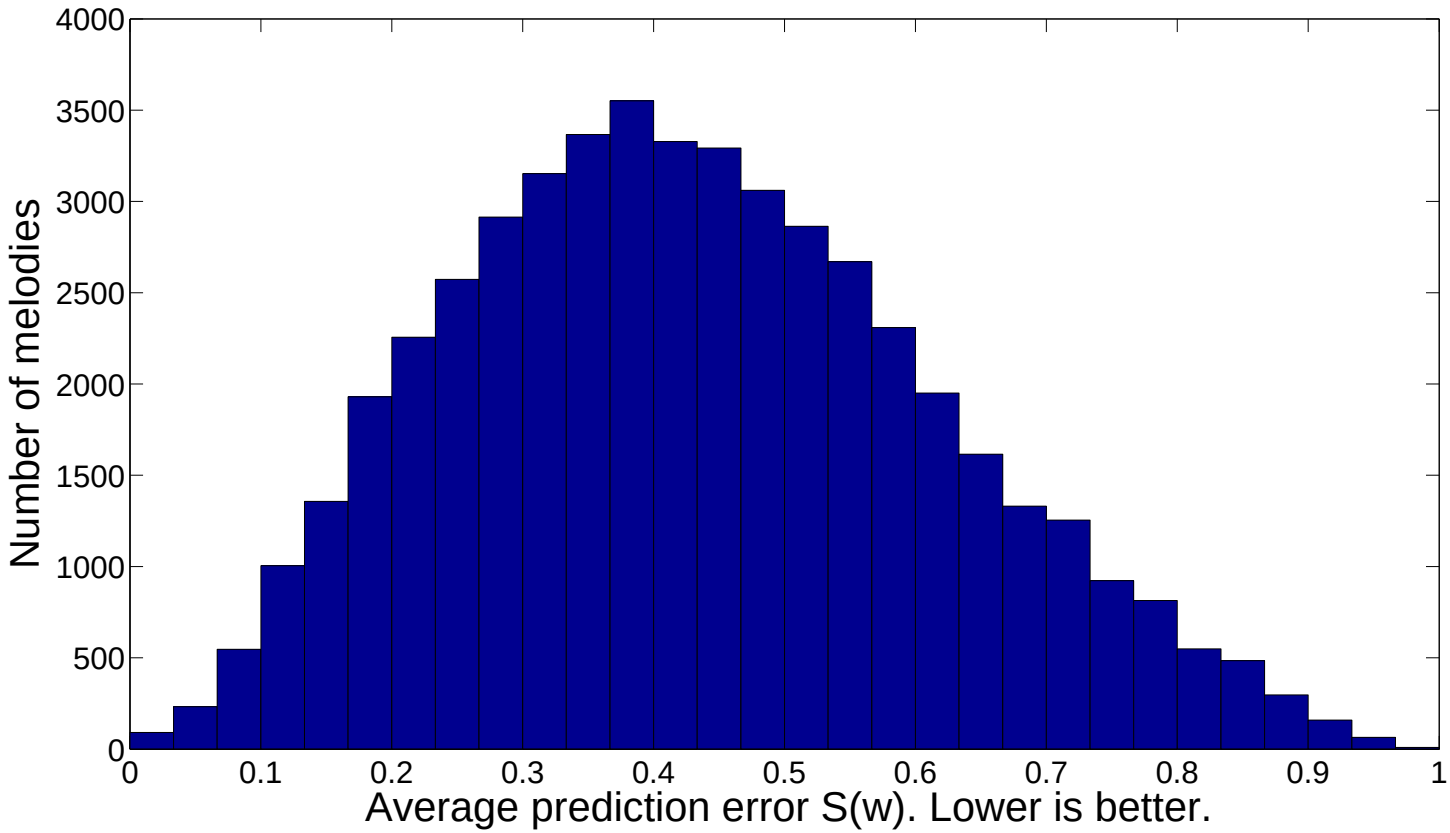
Пространство of 2-граммов

Распределение композиций по жанрам спроецированное из пространства биграммов аккордов на двумерную плоскость методом PCA.



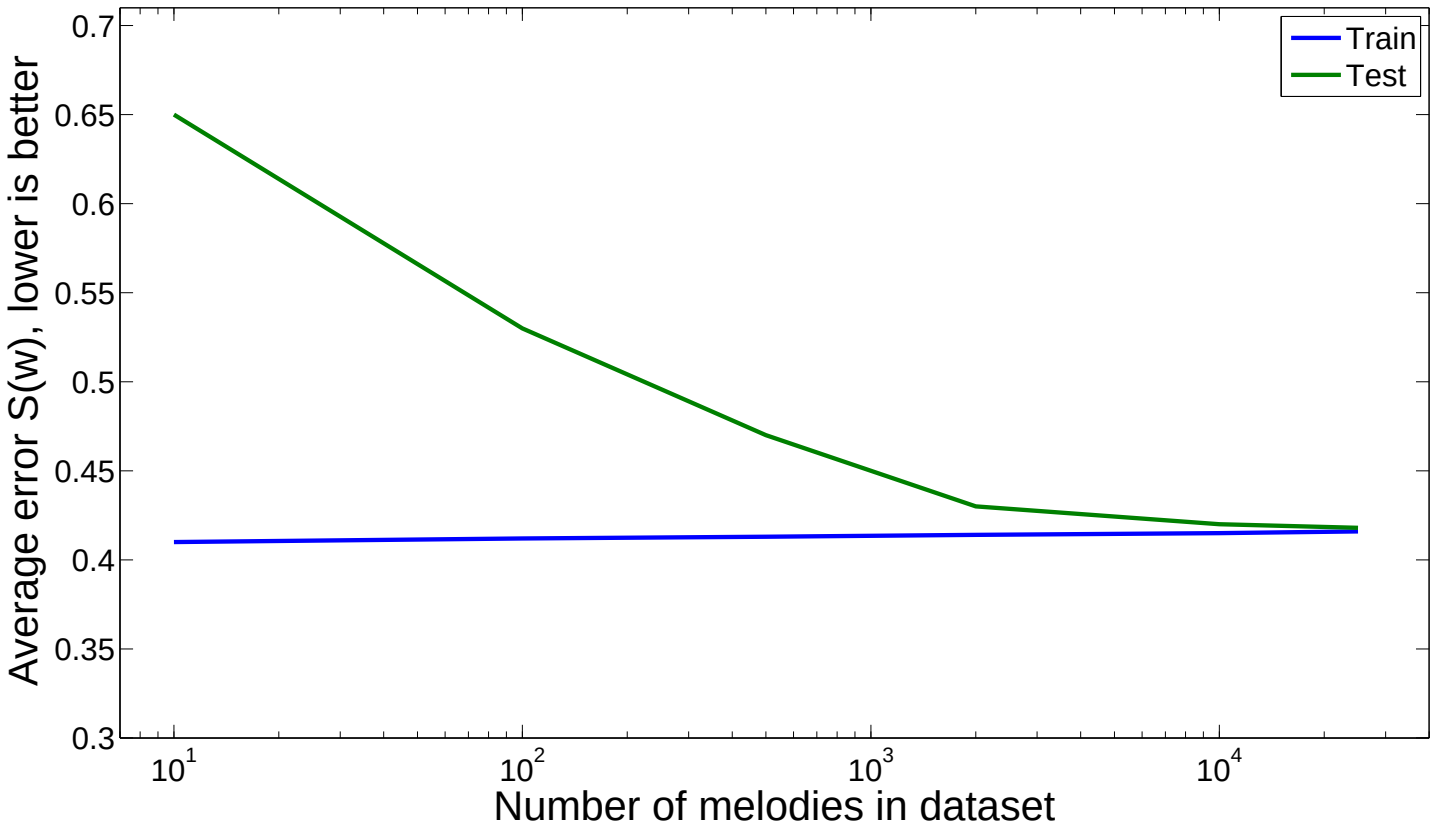
Качество предсказания

Число параметров равно 16, тренировочный набор — весь Midi50k. Средняя величина ошибки равна **0.42** (означает 58% успешно предсказанных элементов $x \in \mathbf{E}$). Существуют мелодии, предсказанные на 100%, так же как и мелодии предсказанные плохо (<5%).



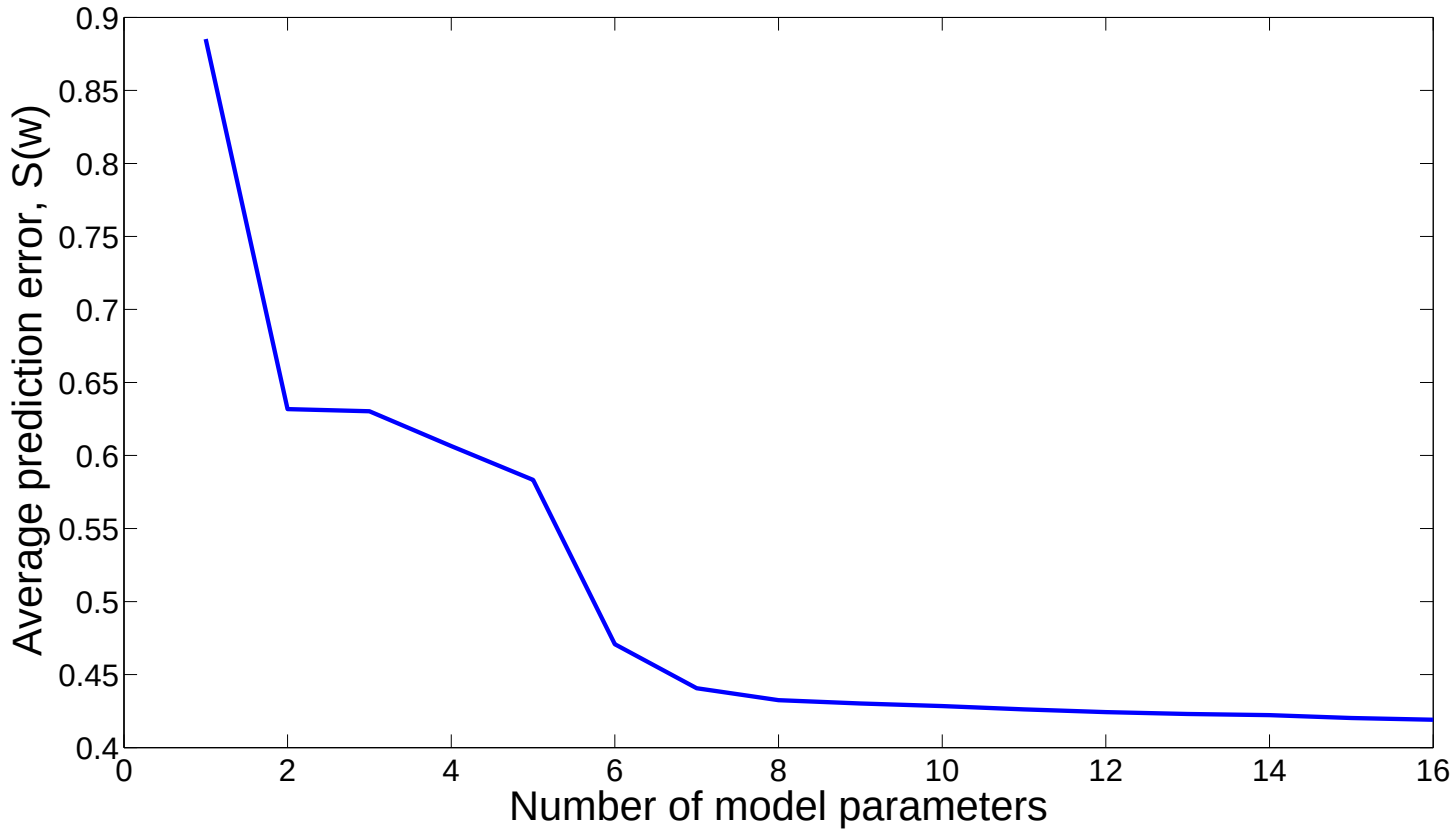
Размер тренировочных данных

$$S(\mathbf{w}, \mathbb{X}) = \sum_{\mathbf{x} \in \mathbb{X}} \sum_{i=1}^{N_{\mathbf{x}}-1} \left[x_{i+1} \neq f(\mathbb{X}, \underbrace{\{x_1, \ldots, x_i\}}_{\text{Prev. part of the melody}}, \mathbf{w}) \right].$$



Сложность модели

Функция ошибки $S(\mathbf{w}, \mathbb{X})$ и число параметрв K ($\mathbf{w} \in \mathbb{R}^{2K}$), проверочная выборка:



Сравнение методов

Quality	Mozer[1]	Conklin[2]	Proposed
Main idea	Neural network	Music patterns	Bayes classifiers
Chords	—	40%	**58.0%**
Pitches	93%	**95%**	92.5%
Durations	**90%**	75%	—
Datasize	20	4500	50 000

---

[1] *Neural network music composition by prediction* — M. Mozer, Connection Science, 1994.

[2] *Multiple viewpoint systems for music prediction* — D. Conklin, I. Witten, Journal of New Music Research, 1995, rev. 2002.

Результаты вычислительного эксперимента

- Оптимальная сложность модели (максимальная длина N-граммов) равна 8, но чем больше тем лучше.
- Число мелодий в тренировочном наборе желательно больше 1000.
- Качество предсказания 58% (по-аккордно, 0.024% если наугад), расстояние Хэминга равно 0.075 (92.5% совпадающих полутонов, 50% для случайного угадывания).

Защищаемые утверждения

- Представление последовательности аккордов как поледовательности элементов специально сконструированной группы вполне обосновано.
- Разработан алгоритм предсказания одного следующего элемента музыкальной последовательности.
- Алгоритм протестирован на наборе Midi50k и проведено его сравнение с другими работами (Mozer, Conklin) с точки зрения качества предсказания.

Публикации

- M. Matrosov, V. Strijov, A. Matrosov. Short-term forecasting of musical compositions using chord sequences. Conference of International Federation of Operational Research Societies, — July 2014, Barcelona, Spain.
- M. Matrosov, V. Strijov. Short-term forecasting of musical compositions using chord sequences. 57-th Scientific Conference of MIPT. — November 2014, Dolgoprudny, Russia.
- Seminar “Music and Science”, Moscow State Conservatory named for P. I. Tchaikovsky. — January 20, 2015, Moscow, Russia.