

Применение байесовской регуляризации при обучении распознавания сигналов

А.Ф. Квасов, О.В. Красоткина

Факультет ВМК
МГУ им. М.В. Ломоносова

17-я Всероссийская конференция
Математические методы распознавания образов
Светлогорск, 19 - 24 сентября 2015

- Пусть каждый объект генеральной совокупности $\omega \in \Omega$ представлен точкой в линейном признаковом пространстве $\mathbf{x}(\omega) = (\mathbf{x}^1(\omega), \dots, \mathbf{x}^T(\omega))$, $\mathbf{x}^t \in \mathbb{R}^n$, а его скрытая фактическая принадлежность к одному из двух классов описывается значением индекса класса $y(\omega) \in \{1, -1\}$.
- В нашем распоряжении имеется обучающая выборка $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$, по которой необходимо восстановить принадлежность любого объекта генеральной совокупности к одному из двух классов.
- Будем рассматривать параметрическое решающее правило в виде гиперплоскости задаваемое дискриминантной функцией: $f(\mathbf{x}(\omega), \mathbf{a}, b) = \mathbf{a}^T \mathbf{x}(\omega) + b$, которая преимущественно > 0 , если $y(\omega) = 1$, и < 0 если $y(\omega) = -1$.

$$\mathbf{x} \in \mathbb{X}, x = (x_1, x_2, \dots, x_t, \dots), t = 1, \dots, T$$

$$x_t = (x_t^1, x_t^2, \dots, x_t^n)$$

$$x_t^i \in \mathbb{X}_i, i = 1, \dots, n$$

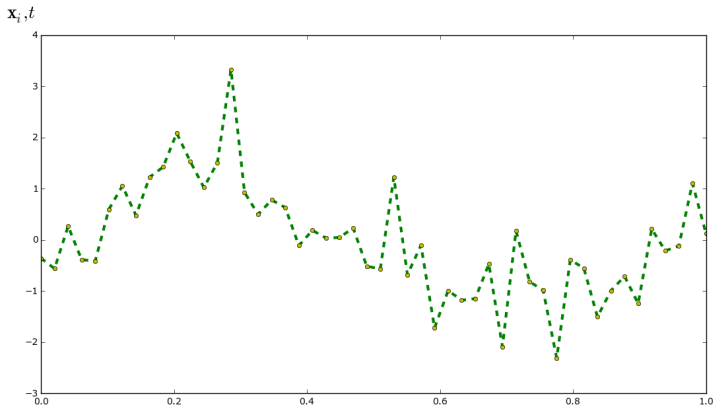
- Свойство упорядоченности признаков вдоль некоторой оси t
- Длина сигналов, соответственно, размерность признакового пространства может быть различной

Примеры объектов распознавания

Действительнозначный сигнал

Пример №1: Действительнозначный сигнал

$$n = 1, \mathbf{x}_i = (x_{i_t})_{t=1}^T, \\ x_{i_t} \in \mathbb{R}$$

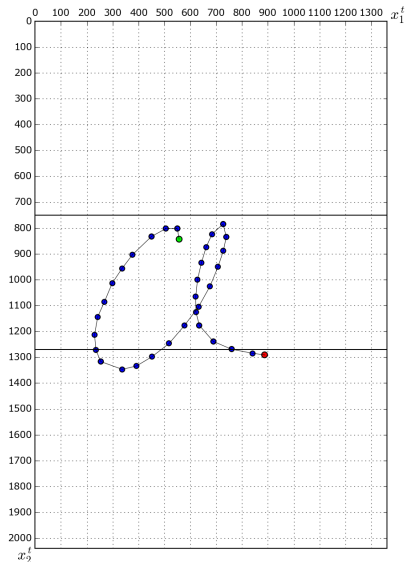


Примеры объектов распознавания

Рукописные символы (*online handwriting*)

Пример №2: рукописные символы (*online handwriting*)

$n = 2$, $\mathbf{x}_i = (x_{it})_{t=1}^{T_i}$, $x_{it} \in \mathbb{R}^2$



Критерий обучения для оценивания параметров решающего правила

Общая постановка

Метод максимума апостериорной вероятности

$$\begin{aligned}(\mathbf{a}_{opt}, b_{opt}) &= \operatorname{argmax}_{\mathbf{a}, b} P(\mathbf{a}, b \mid \{(\mathbf{x}_i, y_i)\}_{i=1}^N) = \\ &= \operatorname{argmax}_{\mathbf{a}, b} \Phi(\{(\mathbf{x}_i, y_i)\}_{i=1}^N \mid \mathbf{a}, b) \Psi(\mathbf{a}, b)\end{aligned}$$

$P(\mathbf{a}, b \mid \{(\mathbf{x}_i, y_i)\}_{i=1}^N)$ - апостериорное распределение параметров;

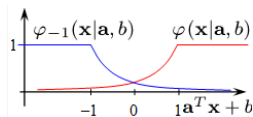
$\Phi(\{(\mathbf{x}_i, y_i)\}_{i=1}^N \mid \mathbf{a}, b)$ - функция правдоподобия выборки;

$\Psi(\mathbf{a}, b)$ - априорное распределение параметров;

Критерий обучения для оценивания параметров решающего правила

Пусть наблюдения независимы, тогда:

$$\Phi(\{(\mathbf{x}_i, y_i)\}_{i=1}^N | \mathbf{a}, b) = \prod_{i=1}^N \varphi_{y_i}(\mathbf{x}_i | \mathbf{a}, b)$$



Несобственное распределение функции правдоподобии

$$\varphi_1(\mathbf{x} | \mathbf{a}, b) = \begin{cases} 1, & \mathbf{a}^T \mathbf{x} + b > 1, \\ \exp[-c(1 - (\mathbf{a}^T \mathbf{x} + b))], & \mathbf{a}^T \mathbf{x} + b < 1, \end{cases}$$
$$\varphi_{-1}(\mathbf{x} | \mathbf{a}, b) = \begin{cases} 1, & \mathbf{a}^T \mathbf{x} + b < -1, \\ \exp[-c(1 + (\mathbf{a}^T \mathbf{x} + b))], & (\mathbf{a}^T \mathbf{x} + b) > -1. \end{cases}$$

Критерий обучения для оценивания параметров решающего правила

Априорное распределение параметров

Совместное несобственное (нет конечного интеграла по b) априорное распределение в пространстве параметров модели $\sim \mathcal{N}(\mathbf{0}, \sigma^2 B^{-1})$:

$$\Psi(\mathbf{a}, b | \sigma^2, B) \propto \exp\left(-\frac{1}{2\sigma^2} \cdot \mathbf{a}^\top B \mathbf{a}\right)$$

$$\forall \mathbf{a}, \mathbf{a}^\top B^{-1} \mathbf{a} > 0, (B^{-1})^\top = B^{-1}$$

Критерий обучения для оценивания параметров решающего правила

Вероятностная интерпретация регуляризатора

$$\ln P(\mathbf{a}, b|X, Y) = \ln \Psi(\mathbf{a}, b) + \sum_{i=1}^N \ln \varphi_{y_i}(\mathbf{x}_i|\mathbf{a}, b) \rightarrow \max_{\mathbf{a}, b}$$

Критерий обучения

$$\begin{aligned} Q(\mathbf{a}, b|X, Y) &= -\ln P(\mathbf{a}, b|X, Y) = \\ &= \frac{1}{2 \cdot \underbrace{\sigma^2}_C} \cdot \mathbf{a}^\top B \mathbf{a} + \sum_{i=1}^N \max(0, 1 - y_i(\mathbf{a}^\top \mathbf{x}_i + b)) \rightarrow \min_{\mathbf{a}, b} \end{aligned}$$

Выбор априорного распределения параметров в задаче распознавания сигналов

Пространство действительных признаков

Выберем априорное распределение параметров следующим образом:

$$\Psi(\mathbf{a}) \propto \exp\left[-\frac{1}{2}\mathbf{a}^T B \mathbf{a}\right], \text{ где } B = I + \alpha \tilde{B}$$

$$\tilde{B} = \begin{pmatrix} 1 & -1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ -1 & 2 & -1 & \ddots & & & & \vdots \\ 0 & -1 & 2 & -1 & \ddots & & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & -1 & 2 & -1 & 0 \\ \vdots & & & & \ddots & -1 & 2 & -1 \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & -1 & 1 \end{pmatrix} \in \mathbb{R}^{d \times d}$$

$$\text{Тогда } \exp\left[-\frac{1}{2}\mathbf{a}^T B \mathbf{a}\right] = \exp\left[-\frac{1}{2} \sum_{i=1}^T a_i^2 - \frac{1}{2} \sum_{i=1}^{T-1} (a_{i+1} - a_i)^2\right].$$

Выбор априорного распределения параметров в задаче распознавания сигналов

Пространство действительных признаков

Критерий гладкости и l^2 -регуляризатор

$$\tilde{B} \geq 0, B = I + \alpha \tilde{B} > 0$$

, где α - гиперпараметр гладкости

$$\forall \mathbf{v} \neq 0, \mathbf{v}^T (I + \alpha \tilde{B}) \mathbf{v} = \mathbf{v}^T \mathbf{v} + \underbrace{\alpha \mathbf{v}^T \tilde{B} \mathbf{v}}_{\geq 0} > 0$$

Вероятностная интерпретация критерия гладкости

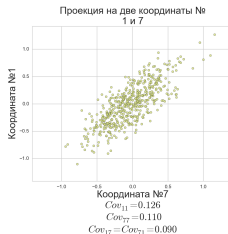
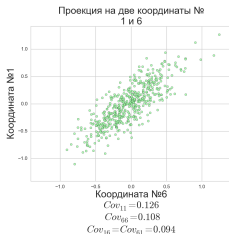
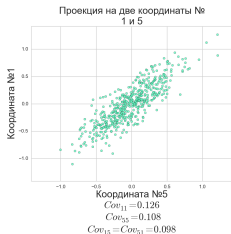
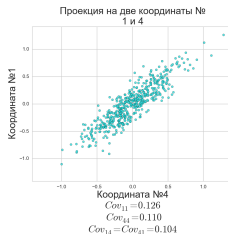
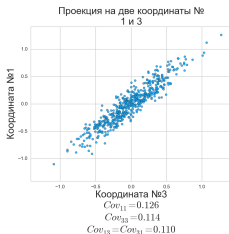
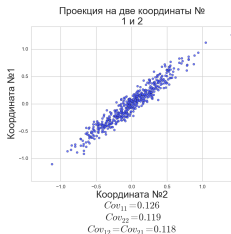
$$B > 0; B^T = B \Rightarrow$$

$$a \sim \mathcal{N}(\mathbf{0}, B^{-1})$$

Выбор априорного распределения параметров в задаче распознавания сигналов

Пространство действительных признаков

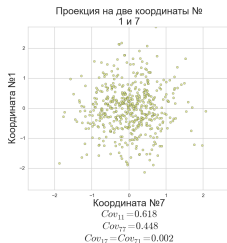
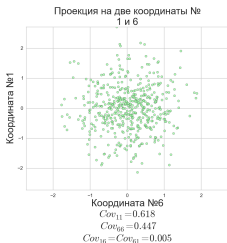
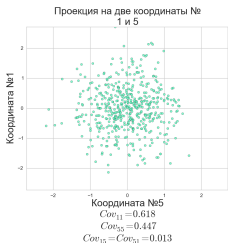
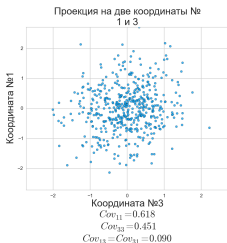
Визуализация двух координат случайной выборки в пространстве векторов параметров модели с ковариационной матрицей $Cov = B^{-1}$, $\alpha = 100$



Выбор априорного распределения параметров в задаче распознавания сигналов

Пространство действительных признаков

Визуализация двух координат случайной выборки в пространстве векторов параметров модели с ковариационной матрицей $Cov = B^{-1}$, $\alpha = 1$



Выбор априорного распределения параметров в задаче распознавания сигналов

Пространство действительных признаков

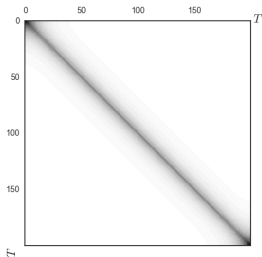


Рис. 1 : Матрица ковариации при $\alpha = 100$, $T = 200$

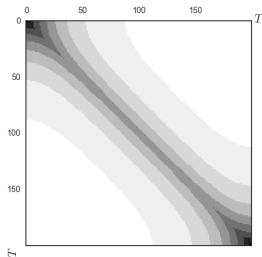
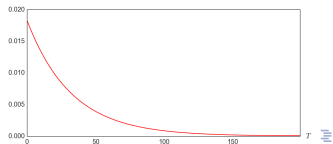
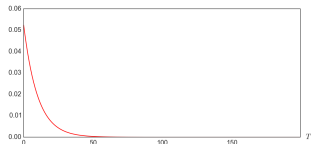


Рис. 2 : Матрица ковариации при $\alpha = 1000$, $T = 200$



Выбор априорного распределения параметров в задаче распознавания сигналов

Пространство метрических признаков

Метрическое пространство

Ω - пространство сигналов

Введем метрику $r(\omega_i, \omega_j) \geq 0$, где $\omega_i, \omega_j \in \Omega$

Дискриминантная функция в метрическом пространстве

$\mathbf{x}_i(\omega) = [-y_i r(\omega, \omega_j)]_{j=1}^N \in \mathbb{R}^N$, тогда дискриминантная функция

$$f(\mathbf{x}(\omega)) = \sum_{i:y_i=1} a_i r(\omega, \omega_i) - \sum_{i:y_i=-1} a_i r(\omega, \omega_i) = \sum_{i=1}^N -y_i a_i r(\omega, \omega_i)$$

Mottl V.V., Seredin O.S., Dvoenko S.D. Pattern Recognition in a Metric Space. Pattern Recognition and Image Analysis, Vol. 11, No. 1, 2001, pp. 65–68

Dynamic Time Warping (DTW)

$$\mathbf{a} = a_1, \dots, a_n; \mathbf{b} = b_1, \dots, b_m; d_{ij} = r(a_i, b_j);$$

$$D(i, j) = d_{ij} + \min[D(i-1, j-1), D(i-1, j), D(i, j-1)]$$

$$W = w_1, w_2, \dots, w_K; w_k = (i_k, j_k); d(w_k) = r(a_{i_k}, b_{j_k})$$

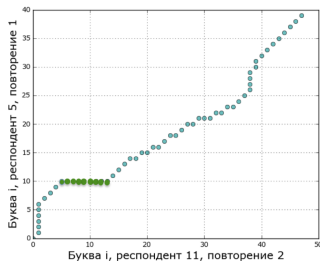
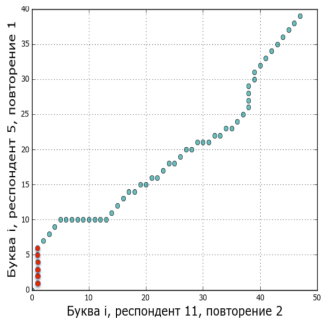
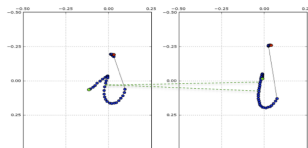
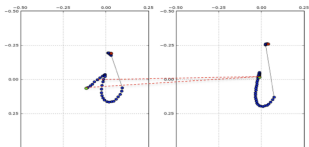
- Граничные условия: $w_1 = (1, 1); w_K = (m, n)$
- Непрерывность: $i_k - i_{k-1} \leq 1, j_k - j_{k-1} \leq 1, \forall k = 2, \dots, K$
- Монотонность: $i_k - i_{k-1} \geq 0, j_k - j_{k-1} \geq 0, \forall k = 2, \dots, K$

K - длина пути; $\max(m, n) \leq K < m + n$

$$\text{Расстояние между сигналами } DTW(\mathbf{a}, \mathbf{b}) = \min \left\{ \frac{\sum_{k=1}^K d(w_k)}{K} \right\}$$

Подсчет расстояний между объектами

Алгоритм динамической трансформации временной шкалы



Выбор априорного распределения параметров в задаче распознавания сигналов

Критерий близости в пространстве метрических признаков

Выберем априорное распределение параметров следующим образом:

$$\Psi(\mathbf{a}) \propto \exp\left[-\frac{1}{2}\mathbf{a}^T B \mathbf{a}\right], \text{ где } B = I + \alpha \tilde{B}$$

$$\tilde{B} = \begin{bmatrix} -\tilde{\beta}_{11} + \sum_{i=1}^N \tilde{\beta}_{1i} & \cdots & -\tilde{\beta}_{1N} \\ \vdots & \ddots & \vdots \\ -\tilde{\beta}_{N1} & \cdots & -\tilde{\beta}_{NN} + \sum_{i=1}^N \tilde{\beta}_{Ni} \end{bmatrix}, \text{ где } \tilde{\beta}_{ij} = \frac{1}{r(\omega_i, \omega_j)}$$

$$\text{Тогда } \exp\left[-\frac{1}{2}\mathbf{a}^T B \mathbf{a}\right] = \exp\left[-\frac{1}{2} \sum_{i=1}^T a_i^2 - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \tilde{\beta}_{ij} (a_i - a_j)^2\right].$$

Постановка SVM с критерием гладкости

$$\begin{cases} \frac{1}{2C} \mathbf{a}^T B \mathbf{a} + \sum_{i=1}^N \xi_i \rightarrow \min_{\mathbf{a}, \xi} \\ y_i (\mathbf{a}^T \mathbf{x}_i(\omega) + b) \geq 1 - \xi_i, i = 1, \dots, N \\ \xi_i \geq 0, i = 1, \dots, N \end{cases}$$

Постановка SVM с критерием близости

$$\begin{cases} \frac{1}{2C} \mathbf{a}^T B \mathbf{a} + \sum_{i=1}^N \xi_i \rightarrow \min_{\mathbf{a}, \xi} \\ y_i \left(\sum_{j=1}^N a_j (-y_j r(\omega, \omega_j)) + b \right) \geq 1 - \xi_i, i = 1, \dots, N \\ \xi_i \geq 0, a_i \geq 0, i = 1, \dots, N \end{cases}$$

Спрямяющая функция $\phi : \mathbb{R}^N \rightarrow \mathbb{R}^N$, использует матрицу гладкости B :

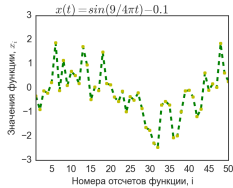
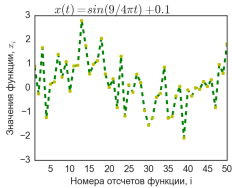
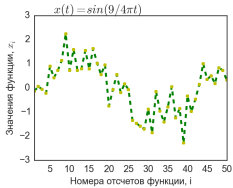
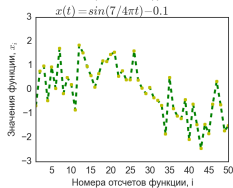
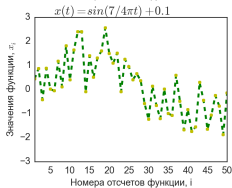
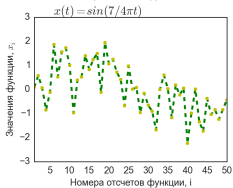
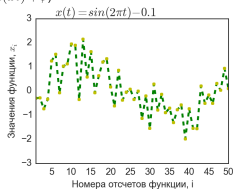
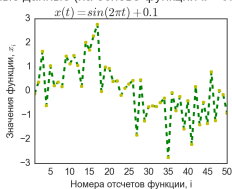
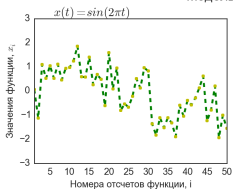
$$\begin{aligned}\mathbf{x}_i^\top B^{-1} \mathbf{x}_j &= (B^{-1} \mathbf{x}_j, \mathbf{x}_i) = ((B^{-1/2} \mathbf{x}_j, (B^{-1/2} \mathbf{x}_i)) = \\ &= (\phi(\mathbf{x}_j), \phi(\mathbf{x}_i)) \Rightarrow \phi(\mathbf{x}_i) = (B^{-1/2} \mathbf{x}_i).\end{aligned}$$

- В роли сигналов - зашумленные синусоиды длиной $T = 50$ отсчетов с различными частотами ω и сдвигами ϕ выбранные так, чтобы формы кривых синусоид отличались незначительно для 9 выбранных классов;
- Проводилась попарная классификация всех классов. Для этого настраивался SMSVM с ядром на подвыборках объектов различных классов
- Ошибки на 100 подвыборках усреднялись, для каждого α проводилось усреднение по парам классов
- Выбирались случайные подвыборки объектов двух классов (равномощные для обоих классов) размерами $N = 25, 50, 100, 150, 300$ для контроля и столько же для обучения

Экспериментальные исследования на модельных данных

Описание модельных данных

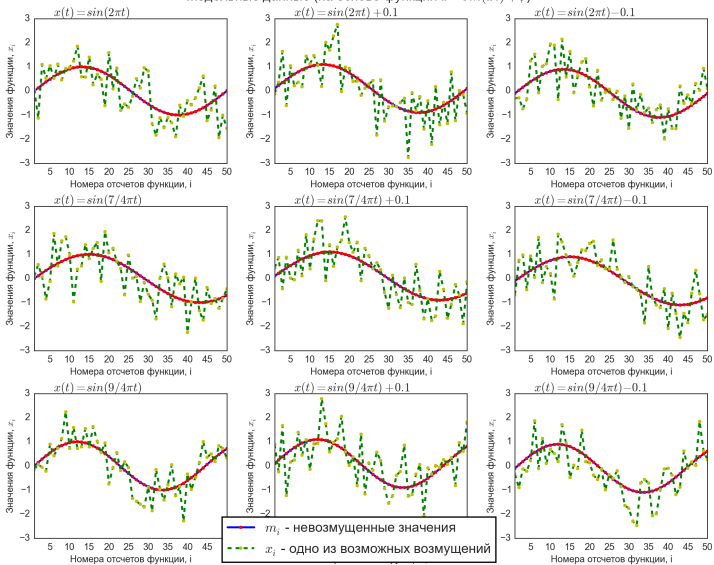
Модельные данные (на основе функции $x = \sin(\omega t) + \phi$)



Экспериментальные исследования на модельных данных

Представление модельных данных

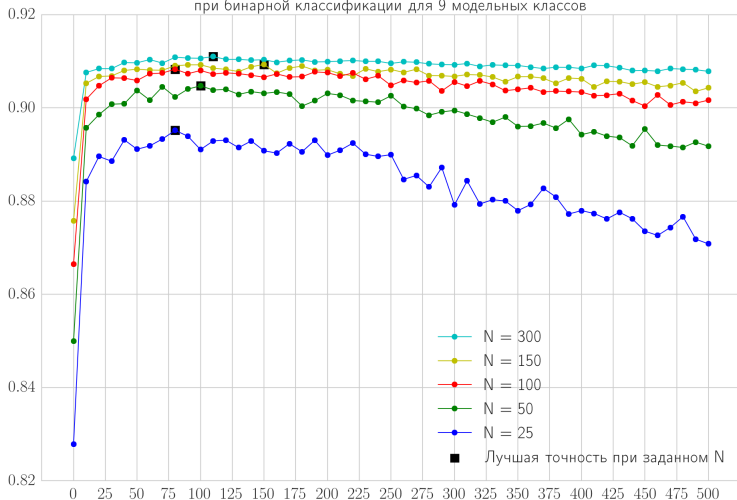
Модельные данные (на основе функции $x = \sin(\omega t) + \phi$)



Экспериментальные исследования на модельных данных

Результаты эксперимента

Усредненные точности на Hold Out с равномоощным разделением на обучение и контроль при бинарной классификации для 9 модельных классов



UJI Pen Characters Data Set (2008)

- 62 символа: 26 латинских букв, прописных и заглавных + 10 цифр
- Символы взяты у 11 респондентов по 2 варианта написания каждого символа;
- Данные собраны на Tablet PC с использованием беспроводного стилуса;
- Записаны только координаты точек (x, y) , данные о давлении и наклоне пера не регистрировались;
- Длина сигнала $T \in [14, 392]$;

<https://archive.ics.uci.edu/ml/datasets/UJI+Pen+Characters>

Структура эксперимента:

- Проводилась попарная классификация с целью исследования критерия близости на разделение символов таких как ('a', 'd'), ('i', 'l') и т.п.;
- 22 объекта: 11 объектов первого символа + 11 объектов второго символа; в обучающей 1ое повторение, а в тестовой 2ое повторение написания символов;
- С каждым сигналом осуществлялась следующая предобработка: перенос в центр масс, нормализация размеров букв поворот буквы вокруг центра масс. Наклон буквы вычисляется как среднее значение углов направления письма в каждой точке, попадающих в интревал от $[-50^\circ, 50^\circ]$;

- Для вычисления метрических признаков между каждой парой сигналов проводилась динамическая трансформация с вычислением расстояния.

$$DTW(\mathbf{a}, \mathbf{b}) = \frac{\sum_{k=1}^{K-1} \delta(\alpha_k) + 5\tau}{K}$$

, где $\delta(\alpha_k)$ - разница между углами сопоставленных точек, τ - количество не диагональных совпадений;

