

Lokalne klasyfikatory jako narzędzie analizy i klasyfikacji sygnałów

Wit Jakuczun

25 listopada 2005

Część I

Hierarchiczne biortogonalne bazy dyskryminacyjne

Sformułowanie problemu

Opracować metodę do wyznaczania cech o następujących własnościach:

Lokalność Każda cecha jest wyliczana na podstawie *części* oryginalnego sygnału.

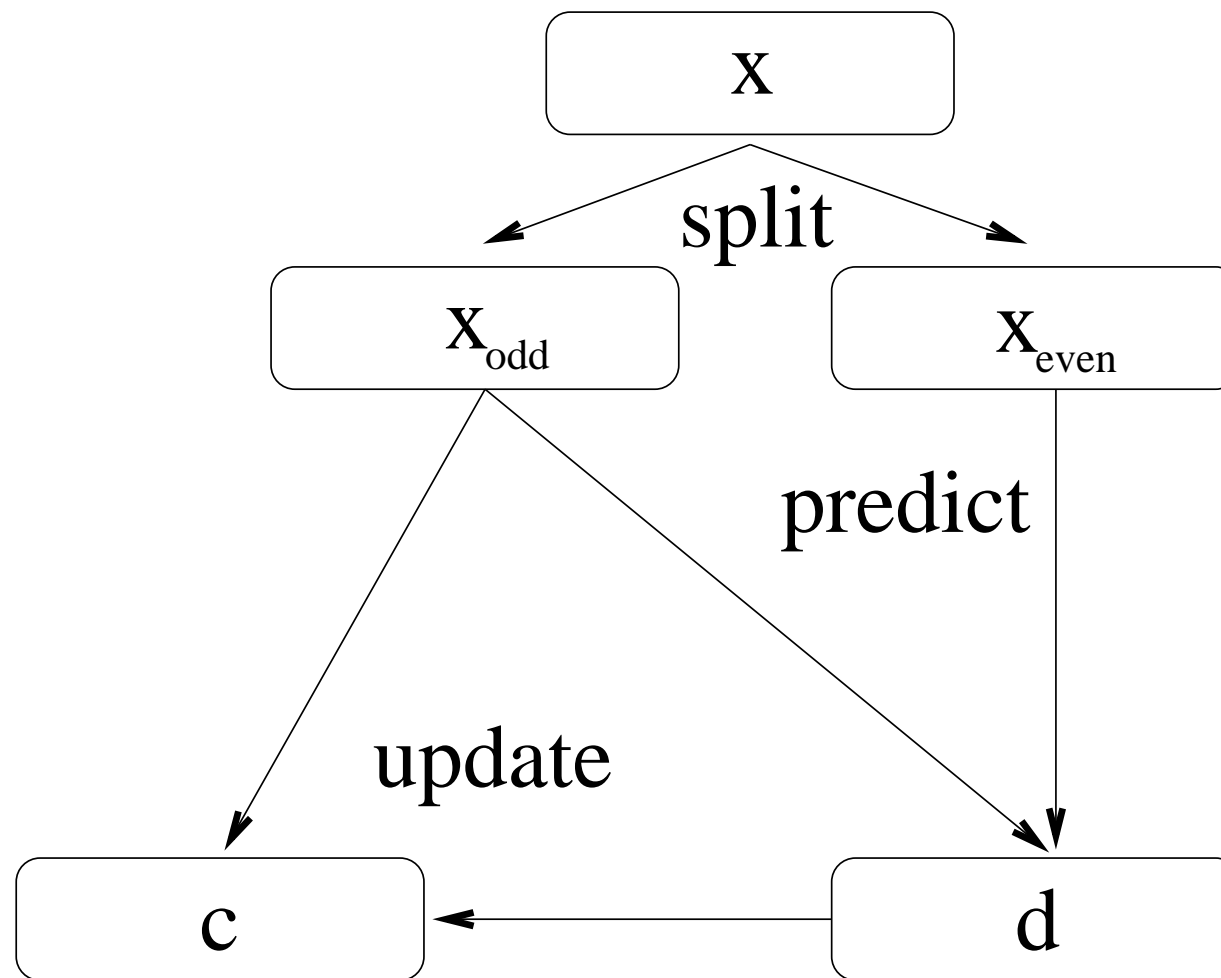
Dyskryminacja Przy wyliczaniu cech brana jest pod uwagę informacja o przynależności do klasy decyzyjnej.

Linowość Własność ta umożliwia rozpisanie analizowanego sygnału w specjalnej bazie zwanej *lokalną bazą dyskryminacyjną*.

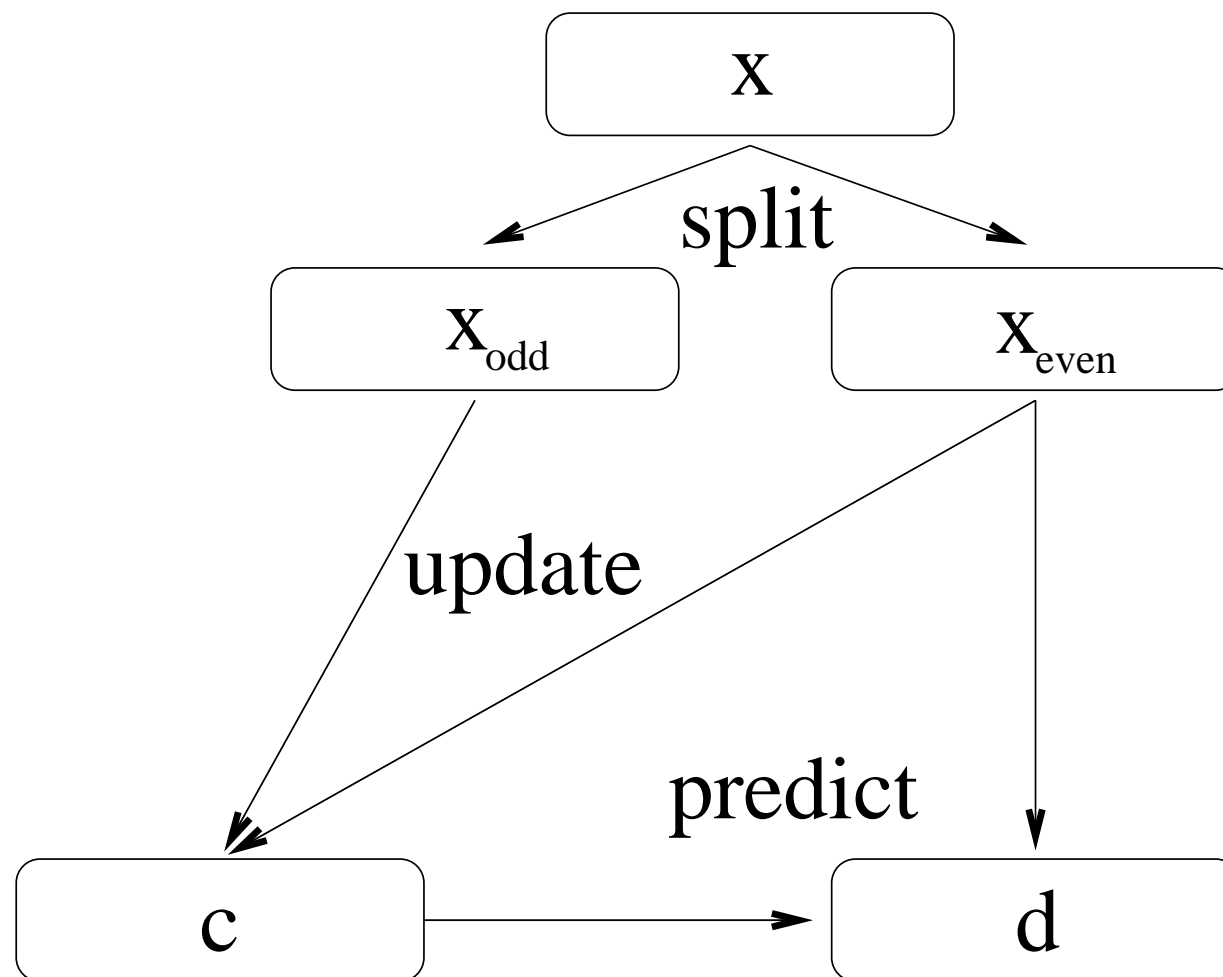
Lifting Scheme

- Metoda konstrukcji falek zaproponowana przez Wima Sweldensa.
- Bardzo prosty i naturalny sposób konstrukcji falek. Nie wymaga znajomości transformaty Fouriera. Cała konstrukcja odbywa się w dziedzinie czasu (przestrzeni).
- Brak problemu z dziedzinami zwartymi (np. odcinkami) oraz z sygnałami, których długość nie jest potęgą dwójki.

Lifting Scheme - schemat



Lifting Scheme - update-first var



Lifting Scheme - operatory

Niech $\mathbf{x} \in \mathbb{R}^T$ for $T = 2^t$ dla pewnego $t \in \mathbb{Z}$

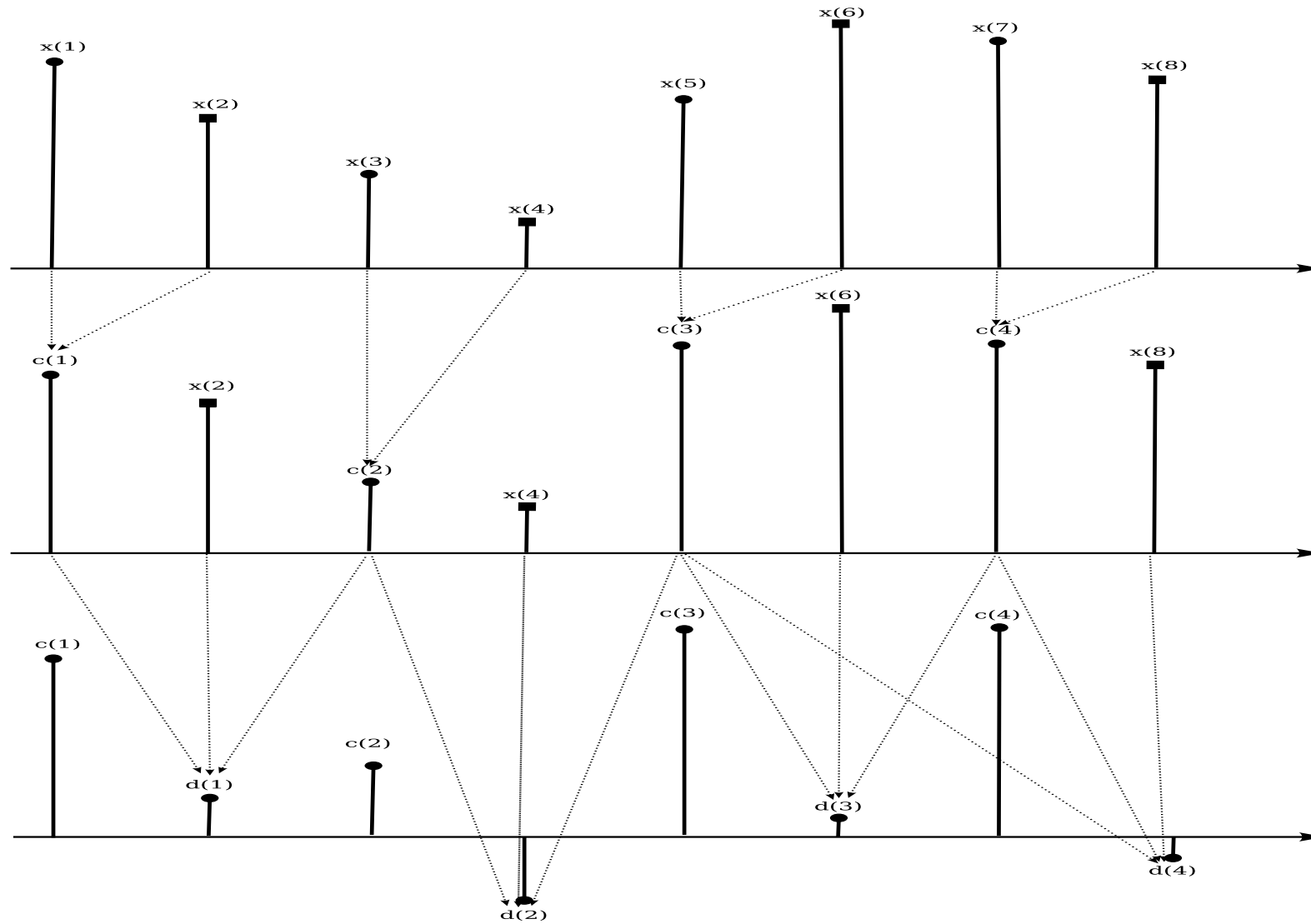
- **SPLIT** - rozbij sygnał \mathbf{x} na dwa podsygnały, składające się odpowiednio z próbek o indeksach nieparzystych (\mathbf{x}_o) i parzystych (\mathbf{x}_e).
- **UPDATE** - wylicz *zgrubną* (ang. *coarse*) aproksymację \mathbf{c} sygnału \mathbf{x}

$$\mathbf{c}(k) = \frac{\mathbf{x}_e(k) + \mathbf{x}_o(k)}{2}, \quad k = 1, 2, \dots, T/2$$

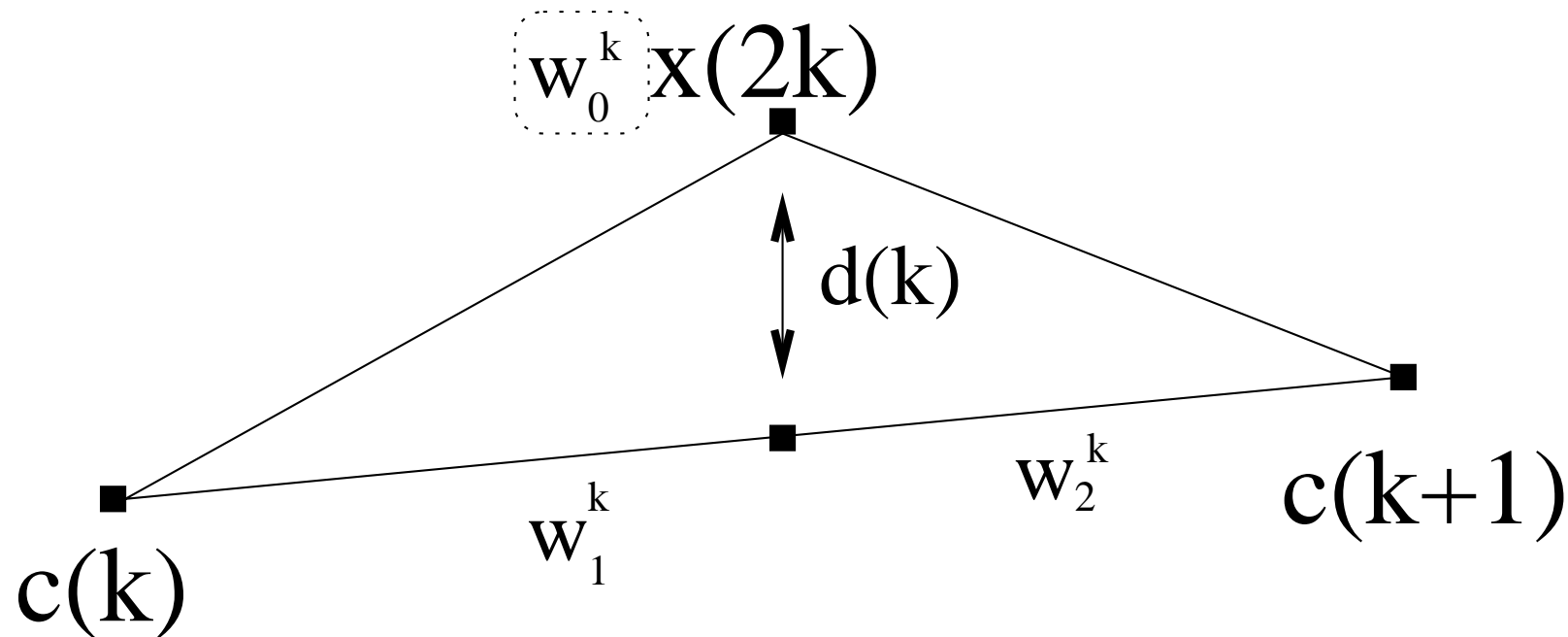
- **PREDICT** - wylicz współczynniki falkowe \mathbf{d}

$$\mathbf{d}(k) = \text{PREDICT}_k(\mathbf{x}_e(k), \mathbf{c}), \quad k = 1, 2, \dots, T/2$$

Lifting Scheme - przykład



Liniowa wersja operatora PREDICT



Współczynniki $d(k)$ można wyznaczyć następująco:

$$d(k) = \mathbf{x}_e(k) - (w_1 \mathbf{c}(k) + w_2 \mathbf{c}(k+1)) \quad \text{częściowa regularyzacja}$$

$$d(k) = w_0 \mathbf{x}_e(k) - (w_1 \mathbf{c}(k) + w_2 \mathbf{c}(k+1)) \quad \text{pełna regularyzacja}$$

Liniowa wersja operatora PREDICT

Zadaniem jest znalezienie takich współczynników $d(k)$, które będą **różne** dla przykładów z **różnych** klas.

- Pełna regularyzacja

$$d(k) = w_0 x_e(k) - (w_1 c(k) + w_2 c(k + 1))$$

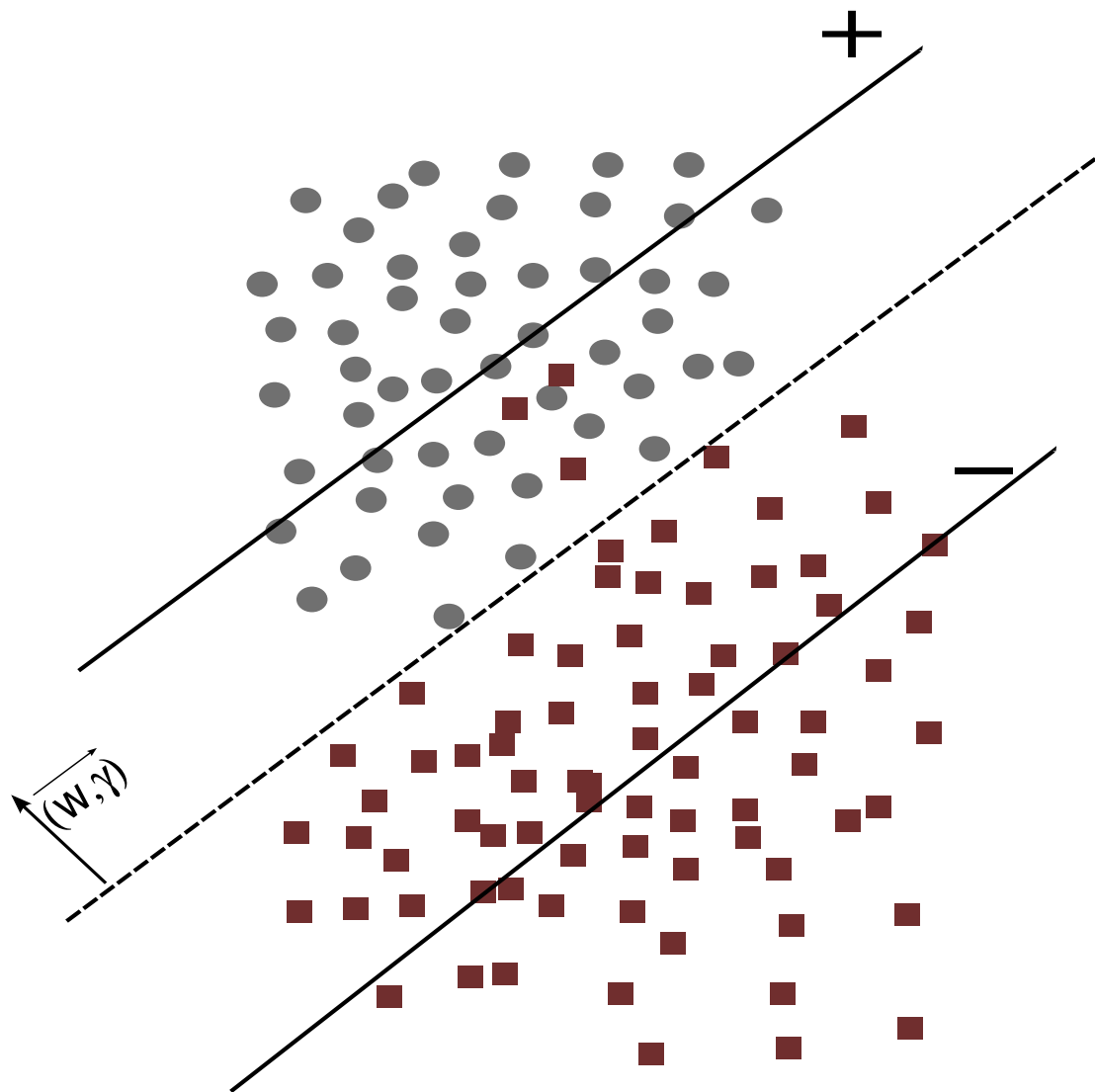
- Niepełna regularyzacja

$$d(k) = x_e(k) - (w_1 c(k) + w_2 c(k + 1))$$

Proximal Support Vector Machines

- Autor Olvi Mangasarian (bardzo płodny w dziedzinie SVM). Polecam lekturę jego strony [www](http://www.cs.wisc.edu/~olvi/)
`http://www.cs.wisc.edu/~olvi/`
- Jedna z wielu wersji powszechnie znanej metody opracowanej przez Vapnika.
- Łatwa w implementacji (wystarczy umieć rozwiązywać układy równań). Wspaniała jak korzysta się z Matlaba/Octave (kilka linijek kodu).
- Równie skuteczna co klasyczne SVM.

PSVM - obrazek poglądowy



PSVM - wersja liniowa

Założmy, że dysponujemy zbiorem treningowym postaci $X = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_l, y_l)\}$ gdzie $\mathbf{x}_i \in \mathbb{R}^N$, $y_i \in \{+1, -1\}$.

Reguła klasyfikacyjna jest postaci:

$$h(\mathbf{x}) = \eta(\langle \mathbf{w}, \mathbf{x} \rangle + \gamma)$$

gdzie

$$\eta(t) = \begin{cases} +1 & \text{dla } t > 0 \\ -1 & \text{dla } t \leq 0 \end{cases}$$

PSVM - wersja liniowa

Optymalny wektor \mathbf{w} jest dany poprzez rozwiązanie następującego problemu optymalizacyjnego

$$\min_{\mathbf{w}, \gamma, \boldsymbol{\xi}} \frac{1}{2} (\|\mathbf{w}\|^2 + \gamma^2) + \frac{\nu}{2} \|\boldsymbol{\xi}\|^2$$

przy ograniczeniach

$$\mathbf{Y} (\mathbf{A}\mathbf{w} - \mathbf{e}\gamma) + \boldsymbol{\xi} = \mathbf{e}$$

Proste rachunki prowadzą do następujących wzorów:

$$\mathbf{u} = \left[\frac{1}{\nu} \mathbf{I} + \mathbf{Y} (\mathbf{A}\mathbf{A}^T + \mathbf{e}\mathbf{e}^T) \mathbf{Y} \right]^{-1} \mathbf{e}$$

$$\mathbf{w} = \mathbf{A}^T \mathbf{Y} \mathbf{u}$$

$$\gamma = -\mathbf{e}^T \mathbf{Y} \mathbf{u}$$

PSVM - dodatkowe ograniczenia

$$\min_{\mathbf{w}, \gamma, \boldsymbol{\xi}} \frac{1}{2} (\|\mathbf{w}\|^2 + \gamma^2) + \frac{\nu}{2} \|\boldsymbol{\xi}\|^2$$

przy ograniczeniach

$$\mathbf{Y} (\mathbf{A}\mathbf{w} - \mathbf{e}\gamma) + \boldsymbol{\xi} = \mathbf{e}$$

$$\mathbf{B}\mathbf{w} = \mathbf{f}$$

Proste rachunki prowadzą do następujących wzorów:

$$\mathbf{u} = \left\{ \frac{1}{\nu} \mathbf{I} + \mathbf{Y} \left[\mathbf{A}\mathbf{A}^T + \mathbf{e}\mathbf{e}^T - \mathbf{A}\mathbf{B}^T \left(\mathbf{B}\mathbf{B}^T \right) \mathbf{B}\mathbf{A}^T \right] \mathbf{Y} \right\}^{-1} \left[\mathbf{e} - \mathbf{Y}\mathbf{A}\mathbf{B}^T \left(\mathbf{B}\mathbf{B}^T \right)^{-1} \mathbf{f} \right]$$

$$\mathbf{w} = \mathbf{A}^T \mathbf{Y} \mathbf{u} + \mathbf{B}^T (\mathbf{B}\mathbf{B}^T)^{-1} \mathbf{f} - \mathbf{B}^T \left[\left(\mathbf{B}\mathbf{B}^T \right) \mathbf{B}\mathbf{A}^T \mathbf{Y} \right] \mathbf{u}$$

$$\gamma = -\mathbf{e}^T \mathbf{Y} \mathbf{u}$$

Szybki algorytm dla PSVM

Do rozwiązywania układów równań postaci

$$\left(\mathbf{G} + \mathbf{U}\mathbf{V}^T \right) \mathbf{x} = \mathbf{b}$$

można wykorzystać formułę *Shermana-Morrisona-Woodbury'ego*

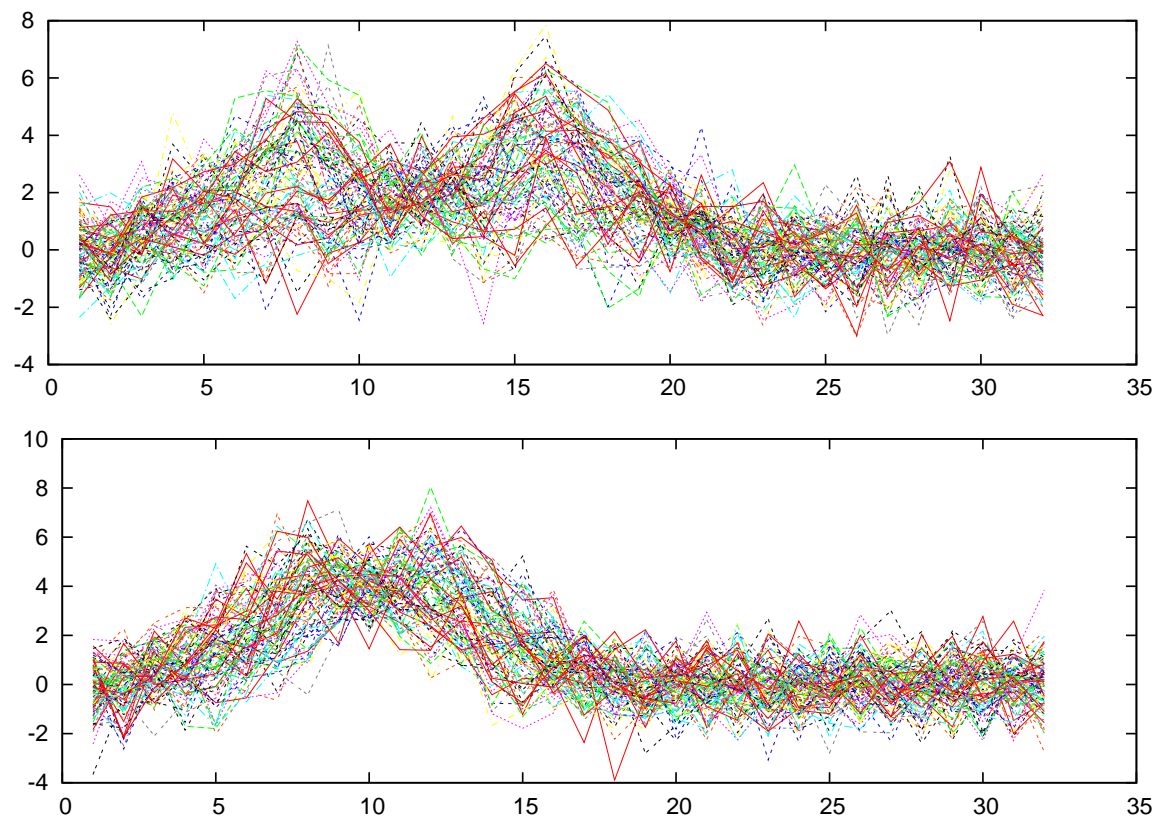
$$\left(\mathbf{G} + \mathbf{U}\mathbf{V}^T \right)^{-1} = \mathbf{G}^{-1} - \mathbf{G}^{-1}\mathbf{U} \left(\mathbf{I} + \mathbf{V}\mathbf{G}^{-1}\mathbf{U} \right)^{-1} \mathbf{V}^T \mathbf{G}^{-1}$$

- Rozwiązanie PSVM jest dane tego typu układem (także z dodatkowymi ograniczeniami)
- Redukcja kosztu obliczeniowego jest znacząca. (Przypominam, że koszt rozwiązywania układu równań jest rzędu $\mathcal{O}(n^3)$).
- W przypadku zastosowania funkcji jądrowych, zysk nie będzie duży, chyba, że zastosuje się podejście zwane Reduced Support Vector Machines (pomysł Mangasariana).

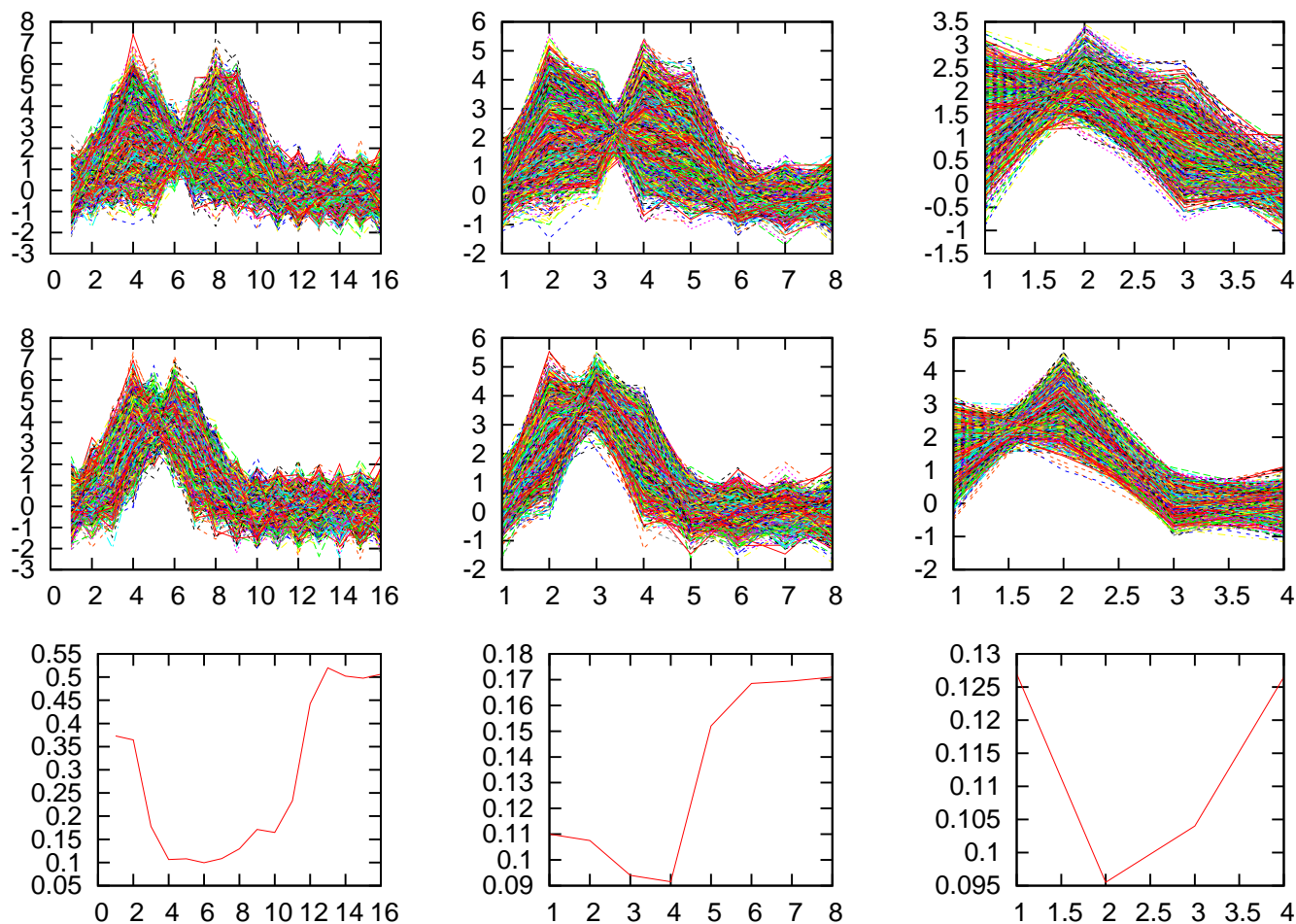
Po co wersja z dodatkowymi ograniczeniami?

- Klasyczny *Lifting Scheme* narzucał na operator PREDICT taki warunek, aby w przypadku gdy sygnał x jest *lokalnie* wielomianem, współczynniki falkowe $d(k)$ były równe 0.
- Konsekwencje są takie, że liczba otrzymanych współczynników jest (w przypadku odpowiednio gładkich sygnałów) jest dużo mniejsza niż liczba próbek (**Kompresja**)
- Stosując wersję algorytmu z *niepełną regularyzacją* oraz PSVM z dodatkowymi ograniczeniami możemy otrzymać bazę, która będzie spełniała dwa warunki:
 - Współczynniki będą równe 0 jeśli sygnał będzie lokalnie wielomianem.
 - Współczynniki będą różne dla różnych przykładów z różnych klas.

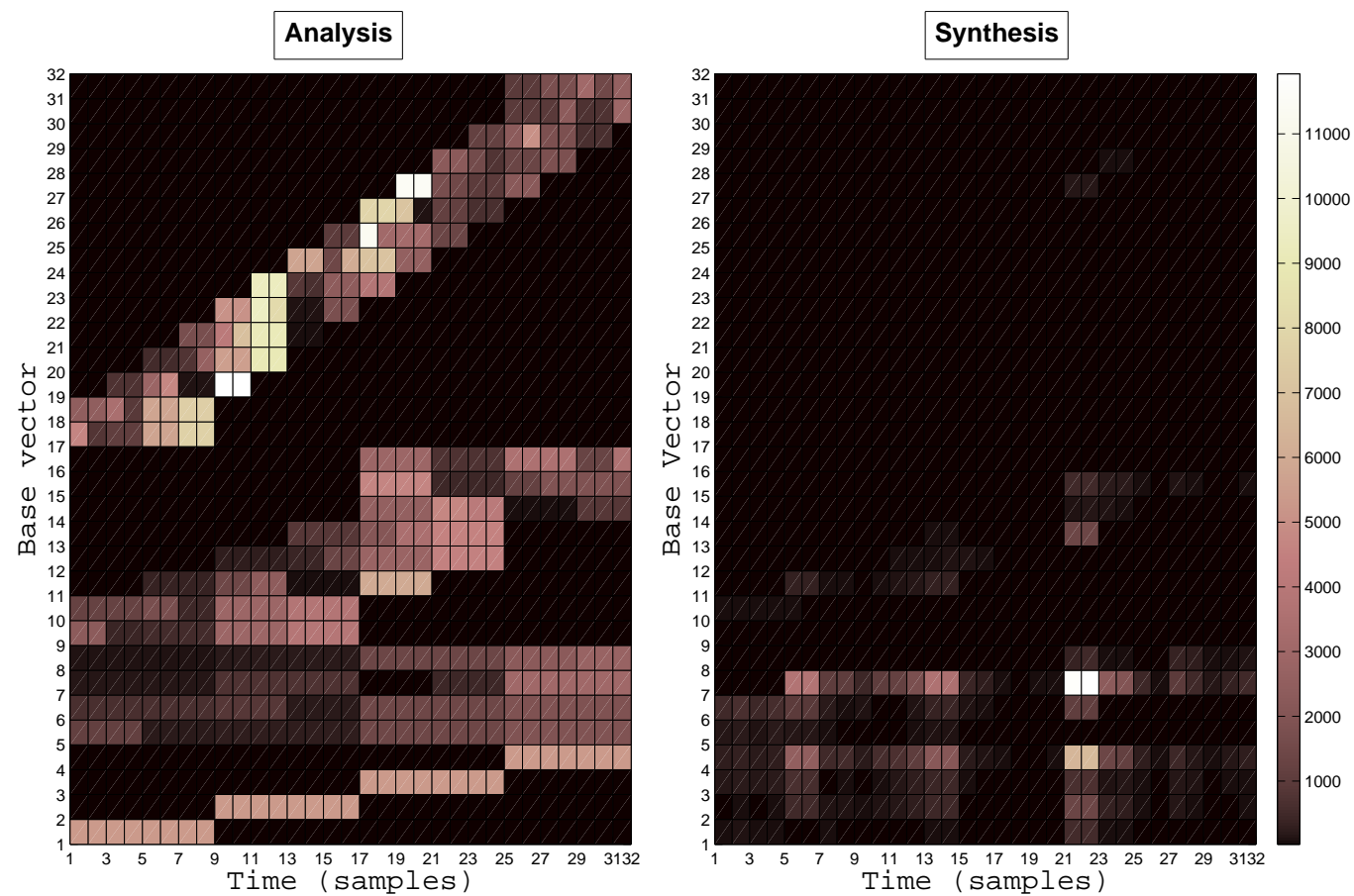
Przykład (dane *Waveform*)



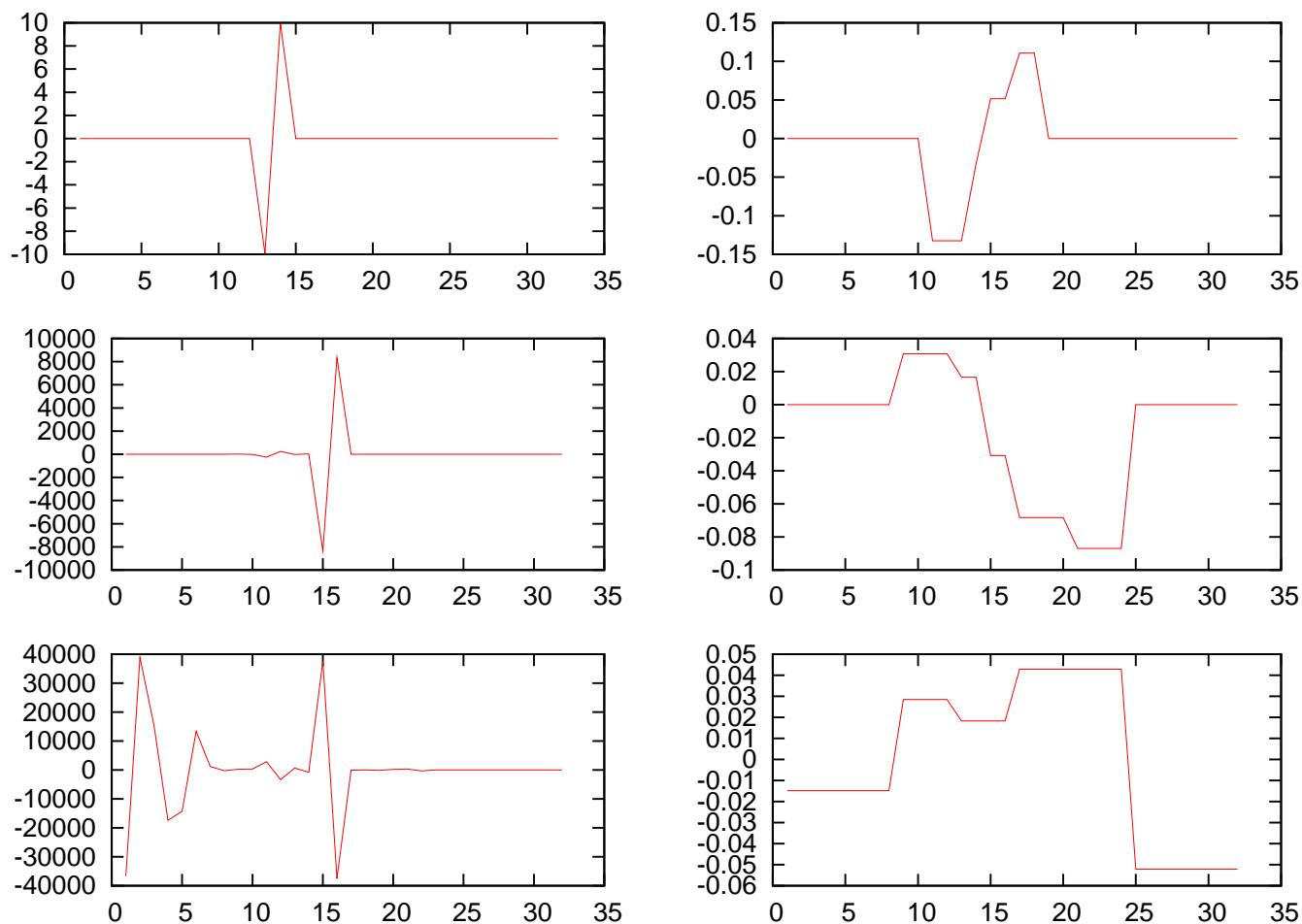
Rysunek 1: Zbiór treningowy (dwie klasy z trzech)



Rysunek 2: Zgrubna aproksymacja oraz błąd klasyfikacji dla trzech poziomów dekompozycji (błąd dla danych oryginalnych wynosi 0.10)



Rysunek 3: Nośniki dyskryminacyjnej bazy biortogonalnej



Rysunek 4: Wektory bazowe: synteza (lewa), analiza (prawa) dla trzech poziomów dekompozycji.

Część II

Zastosowania

Różne scenariusze

Zespół najlepszych Zespół klasyfikatorów składa się z kilku najlepszych współczynników (ocena poprzez walidację krzyżową)

Bagging Zespół klasyfikatorów jest konstruowany na pseudo zbiorach treningowych powstałych poprzez losowanie ze zwracaniem ze zbioru oryginalnego.

Arcing dla wielu klas Jak *bagging* z tym, że losowanie odbywa się według odpowiednio zmienianego rozkładu na zbiorze treningowym.

Różne scenariusze

Arcing + stacking Zespół klasyfikatorów jest tworzony jak w *arcing* ale ostateczna klasyfikacja jest przy pomocy dodatkowego klasyfikatora (PSVM).

Las drzew Wersja bardzo ciekawego algorytmu. Modyfikacja polega na tym, że w każdym węźle budowanego drzewa wybierany jest *losowy* współczynnik falkowy.

Wiele klas Bazowe klasyfikatory są binarne, dlatego należy zastosować odpowiednio *zmodyfikowane* wersje powyższych algorytmów.

Algorytm został przetestowany na zestawie uzyskanym od Eamonna Keogha

Cylinder-Bell-Funnel 3 klasy, 128 próbek

Waveform 3 klasy, 32 próbki

Tracedata 4 klasy, 275 próbek

Gunx 2 klasy, 150 próbek

Control 6 klas, 60 próbek

Auslan 10 klas, 30 próbek, sygnał wielokanałowy
dodatkowo zmierzyłem się z następującymi danymi

USPS 10 klasy, 256 próbek

EEG Dane z konkursu BCI (2003 i 2005)

Wyniki

CBF	0.053	0.029	0.029	0.04	0.019	0.0
Waveform	0.15	0.15	0.156	0.15	0.15	0.14
Trace	0.165	0.15	0.0415	0.415	0.019	0.0
Gunx	0.25	0.105	0.255	0.10	0.03	0.005
Twopattern	-	0.15	0.13	-	-	0.0256
Control	-	-	0.02	0.112	0.005	0.0017
Auslan	-	-	-	0.01	-	0.02
USPS	-	-	-	0.10	0.08	0
EEG(2003)	0.10	-	-	-	-	0.113
EEG(2005)	0.41	-	-	-	-	0.09

Część III

Zakończenie

Wnioski

- Algorytm jest kosztowny, ale można z tym walczyć poprzez wykorzystanie wbudowanej *równoległości* oraz wydajnych bibliotek do algebry liniowej.
- Uzyskane cechy wyliczane są na podstawie fragmentu sygnału. Można to wykorzystać do identyfikacji potencjalnie ciekawych części sygnału. Podejście takie zostało zastosowane przy analizie danych z Instytutu Nenckiego.
- Istnieje możliwość stosowania tzw. *kernel trick* do wyliczania cech. Traci się wtedy własność liniowości.

Propozycje rozszerzeń

- Uogólnienie metody na sygnały wielowymiarowe poprzez tzw. nieseparowalne bazy falkowe. Bardzo komplikuje się implementacja.
- Bardziej stabilny algorytm dla PSVM z dodatkowymi ograniczeniami.
- Sygnały o różnych długościach.
- Algorytm uczenia wykorzystujący dane bez etykietek.
- Warto by było się zastanowić i przebadać dlaczego czasami jakość jest bardzo kiepska!
- Testy, testy, testy