

Uczenie częściowo nadzorowane

wykład 13

Jerzy Stefanowski

Instytut Informatyki PP
2021

Akademia Innowacyjnych Zastosowań Technologii Cyfrowych (AI-TECH)
projekt finansowany z środków Programu Operacyjnego Polska Cyfrowa
POPC.03.02.00-00-0001/20



**Fundusze
Europejskie**
Polska Cyfrowa



**Rzeczpospolita
Polska**

Unia Europejska
Europejski Fundusz
Rozwoju Regionalnego



Plan wykładu

- Motywacje do uczenia częściowo nadzorowanego (ang. semi – supervised learning; SSL)
- Wybrane metody SSL
 - Cluster-and-label
 - Self training
 - Co-training oraz multi-view learning
 - Podejścia grafowe z propagacją etykiet
 - Rozszerzania SVM
- Podsumowanie

Różne podejścia do uczenia się z przykładów

- Uczenie w pełni nadzorowane (ang. supervised) – etykietowane przykłady

$$L = \{(x_i, y_i)\}_{i=1}^n$$

- Uczenie nienadzorowane (unsupervised) – nieetykietowane $U = \{(x_i)\}_{i=1}^m$

- Uczenie częściowo etykietowane (ang. semi-supervised)
Etykietowane dane L oraz (część) nieetykietowanych U , na ogół $m \gg n$

Nie możemy pozyskać etykiet dla choć wybranych przykładów z U – różnica wobec aktywnego uczenia się!

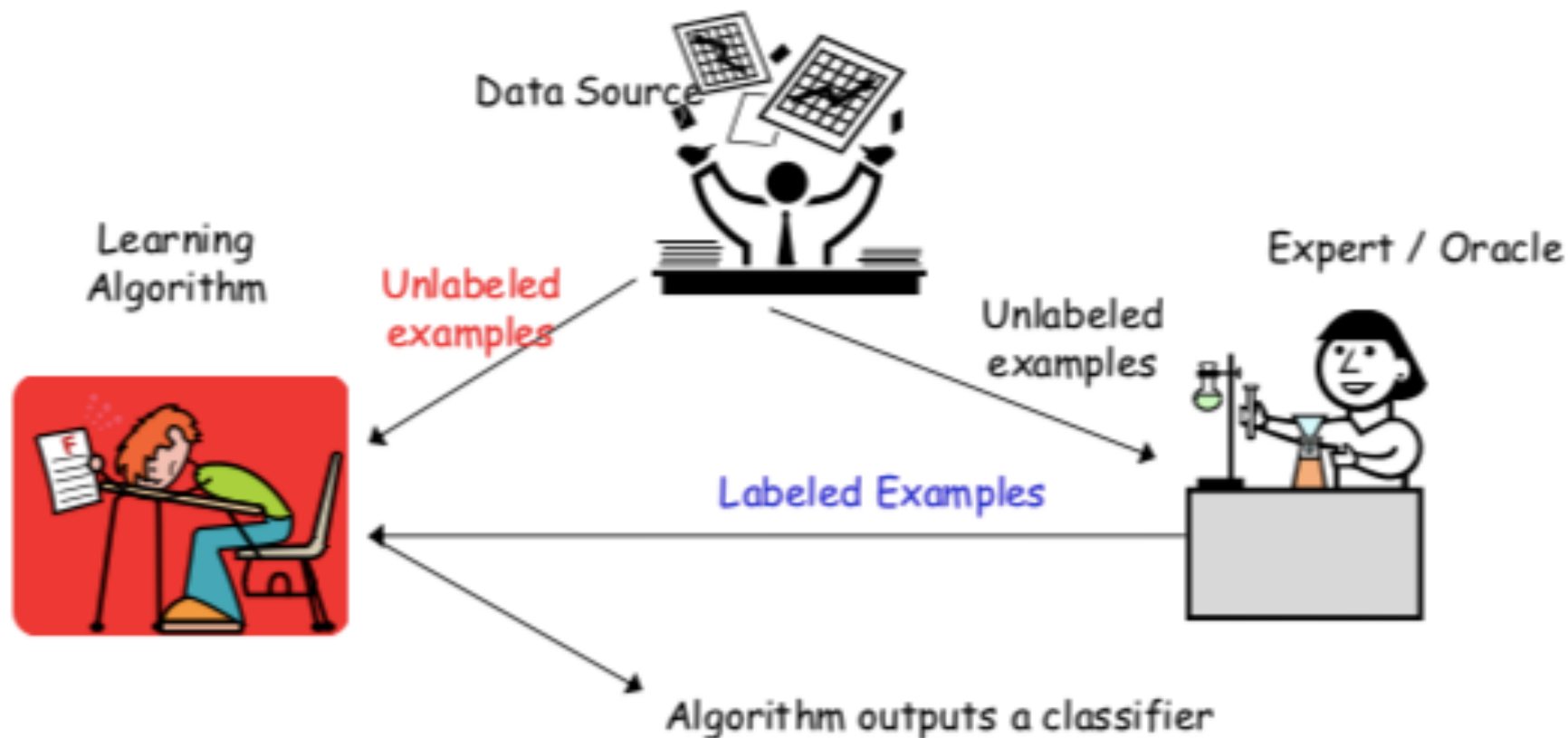
Cel: skonstruowanie lepszego klasyfikatora z L i U niż z samych przykładów etykietowanych L

Różne podejścia do uczenia się z przykładów

Obserwacje motywacyjne:

- Uczenie nadzorowane wymaga wielu etykietowanych przykładów
 - Etykietowanie = kosztowne, czasochłonne i męczące, nierealne dla wielkich kolekcji przykładów
 - Ograniczony dostęp do osób adnotujących przykłady
 - Wymóg wiarygodności (ekspert lepszy niż ochotnicy; niektóre dane trudne do etykietyzacji, obciążenie – bias osób)
- Nieetykietowane przykłady
 - Potencjalnie łatwo dostępne w dużej ilości
 - Często możliwe do automatycznej rejestracji, np.
 - Web, pomiary sensorów, pozyskiwane przykłady z repozytoriów danych
 - „Tańsze” w przygotowaniu

Ogólne motywacje – skorzystajmy z nieetykietowanych przykładów do ulepszenia klasyfikatora



za wykładem N. Balcan nt semi-supervised learning

Przykład anotacji zapisu mowy

- Tzw. Switchboard data – analiza mowy dla zapisu rozmów między ludźmi
- System rozpoznawania mowy -> narzędzia do tworzenia surowej transkrypcji
- Oszacowanie 400 godzin pracy specjalisty anotującego dla 1 godziny zapisu i transkrypcji

Przykłady L i U

Strony WWW, obrazy, dokumenty

Etykietowanie:

- Konieczność czytania lub oglądu
- Wymaga namysłu osoby
- Czasami wspomagane dodatkowymi źródłami

Nieetykietowane:

Potencjalne dostępne w dużej ilości, automatycznie pozyskiwane z minimalnym kosztem

Dane bio-medyczne

Przykładowo predykcja struktur genetycznych

Adnotacja wymaga wysiłku eksperta, dodatkowej wiedzy, niekiedy potwierdzenia (po długim czasie)

Nowe urządzenia diagnostyczne (np. sekwencjonowanie DNA) może generować dużo danych eksperymentalnych

Związek SSL z ludzkim poznawaniem świata

Ludzie często poznają obiekty ze świata z ograniczonym wskazywaniem kategorii, a wiele obiektów później sami klasyfikują :

Przykład uczenia dzieci pojęć

X =zwierzę (cechy) y = gatunek, np. pies

Nauczyciel-rodzic wskazuje i wyjaśnia “to jest pies”

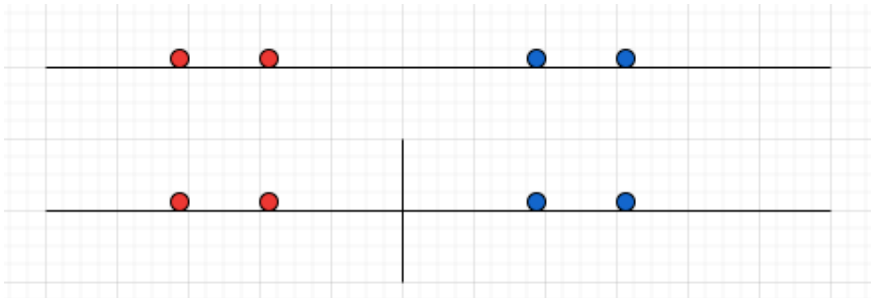
Dzieci obserwują inne zwierzęta (nie objaśnione) i samodzielnie klasyfikują i nazywają



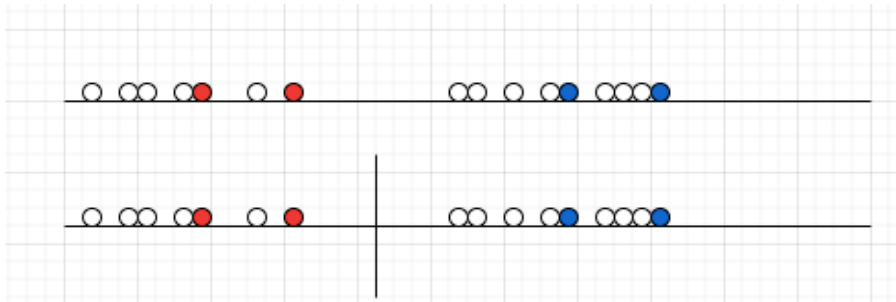
Za Barnabas Poczós

Kiedy SSL może pomóc?

Etykiety : Czerwone klasa 1, niebieskie klasa -1



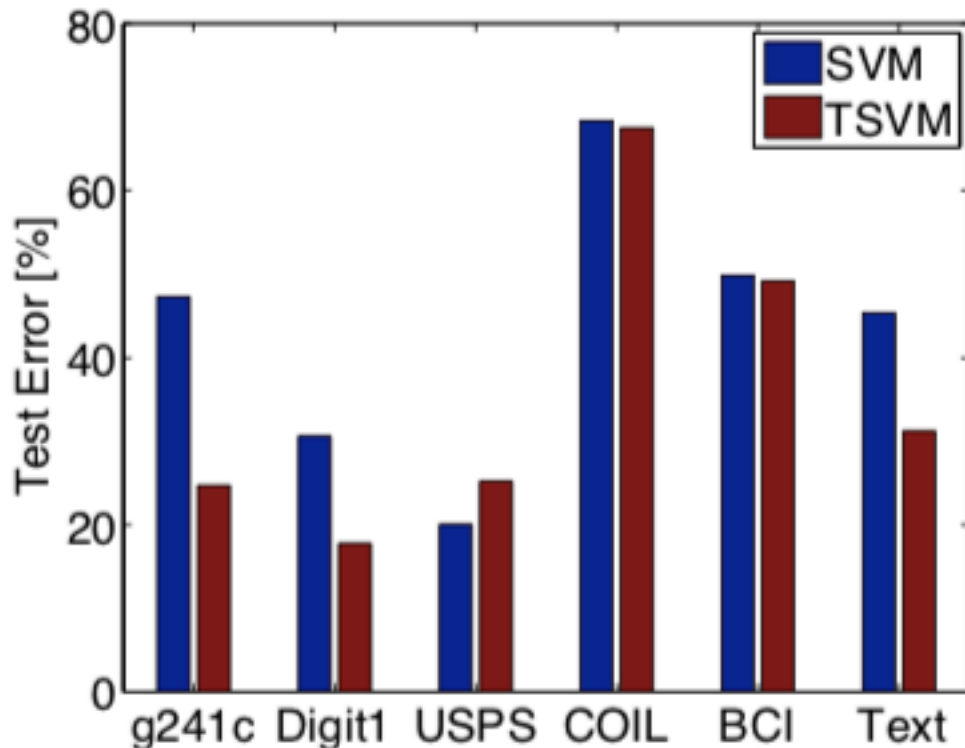
Dodajmy rozkład przykładów nieetykietowanych (szare)



Zmiana granicy decyzyjnej na dokładniejszą!

Założenie: przykłady z różnych klas tworzą spójne rozkłady

Inny przykład użycia Transductive SVM dla różnych zbiorów danych



10 labeled points
~1400 unlabeled
points

SVM: supervised
TSVM:
semi-supervised

Wzrost zainteresowania SSL

W “świecie naukowym – konferencyjnym” widoczny od początku wieku, nowe metody i zastosowania, często w przykładach / obszarach:

- Rozpoznawania obrazów i wizji komputerowej
- Analizy zapisów mowy i przetwarzania dokumentów tekstowych
- Eksploracji zasobów internetu (strony WWW, sieci użytkowników, itd.)

Także widoczne w „przemysłowym” ML i DM / np. patrz wystąpienia na GHOSTDay

Podstawowe metody SSL

Zgodnie z podziałem Zhu:

- Cluster and label (wykorzystane algorytmów grupowania)
 - Generatywne podejścia probabilistyczne (EM, ...)
- Self-learning / samo uczenie się
- Co-training / uczenie się wzajemne
- Multi-view ensembles
- Transductive SVM
- Graph based models

Więcej w pracy przeglądowej po ang. : Zhu, Xiaojin, Semi-Supervised Learning University of Wisconsin-Madison [online pdf].

Pewne założenia i hipotezy o danych

W celu efektywnego wykorzystania nieetykietowanych przykładów U – powinno się poczynić pewne założenia i oczekiwania

A.Blum pisał o potrzebie przekonania, że:

Unlabeled data useful if we have beliefs not only about the form of the target, but also about its relationship with the underlying distribution.

Wstępne hipotezy badawcze dla SSL

W celu wykorzystania nietykietowanych przykładów w procesie uczenia konieczne są założenia co do rozkładów danych. Najczęstsze dwa to:

- **Założenie ciągłości** (Continuity): Najbliższe punkty najprawdopodobniej mają taką samą etykietę / powiązane z preferencją do granicy decyzyjnej w rzadszych obszarach
- **Zgrupowanie w skupiskach** (cluster assumption): dane mogą tworzyć skupiska i punkty wewnątrz skupisk najprawdopodobniej mają te same etykiety (lecz przykłady z tej samej klasy mogą tworzyć wiele skupisk).
- **Manifolds** – punkty często należą do tzw. kształtu geometrii rozmaitości punktów w nowej przestrzeni a nie w oryginalnej reprezentacji.

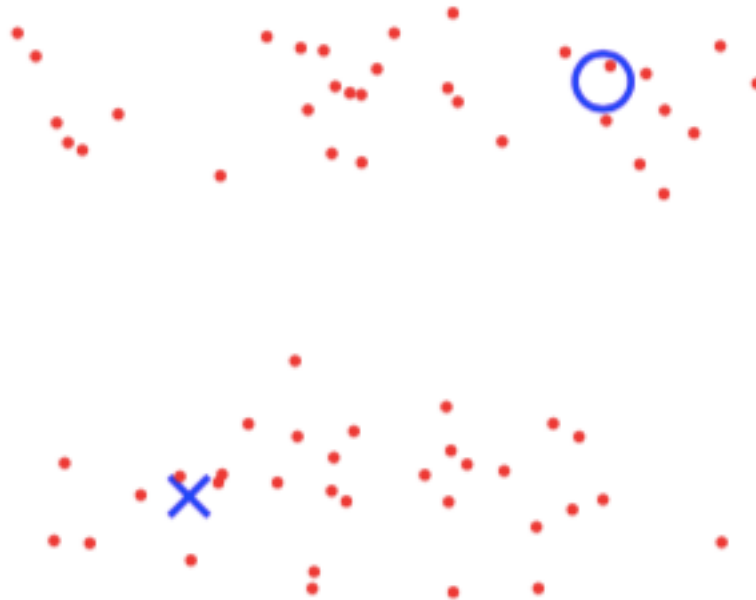
Ilustracja założenia w przypadku skupisk

Punkty położone bardzo blisko siebie mają podobne etykiety

Założenie nt. skupień w danych

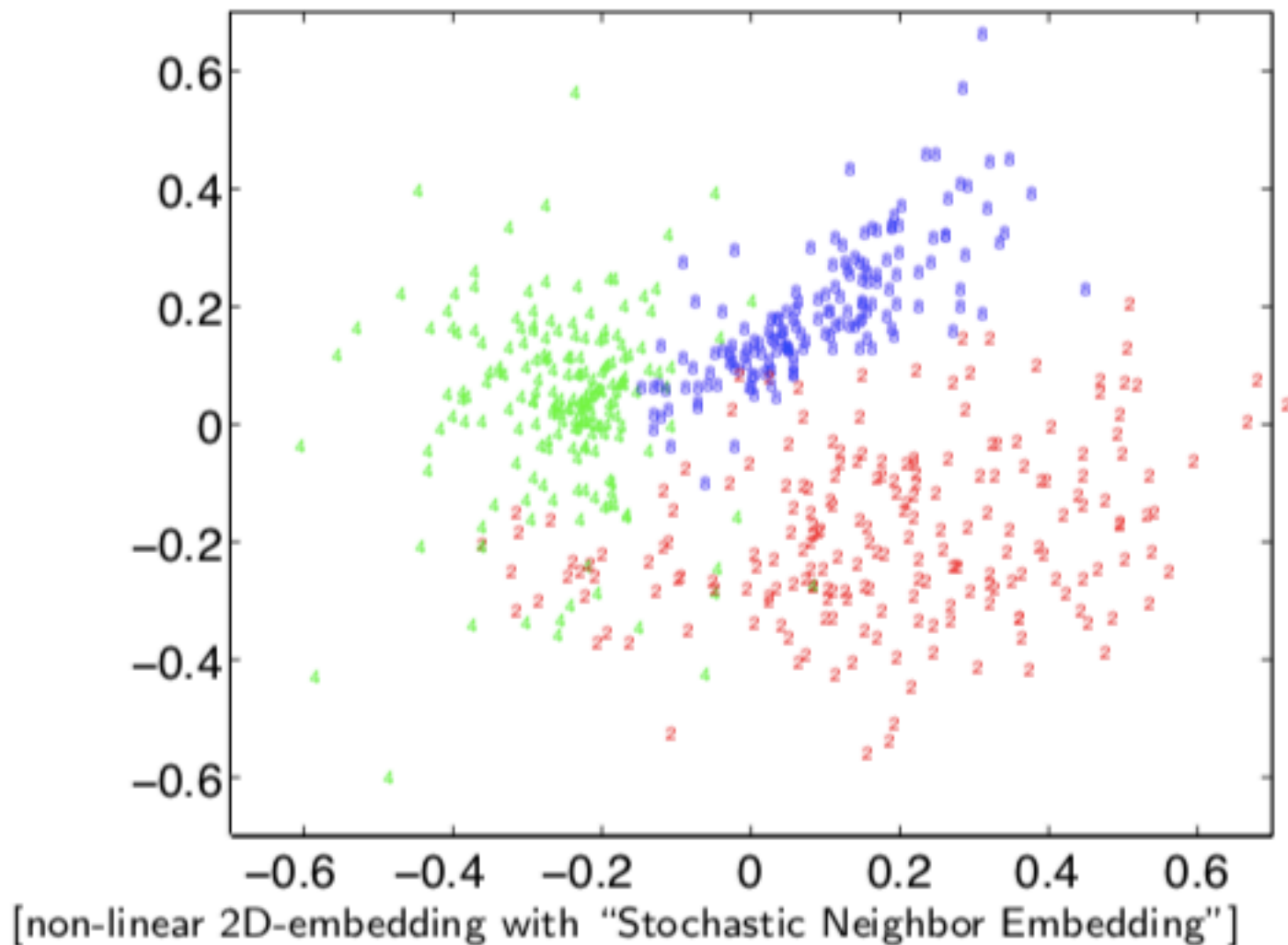
- Dane naturalnie tworzą skupiska
- Punkty w jednym skupisku powinny mieć tę samą etykietę

Rozważmy zbiory nieetykowanych punktów w dwóch skupiskach wokół niewielu etykietowanych:



Przykład rozpoznawania liter

Example: 2D view on **handwritten digits** 2, 4, 8



Grupuj i etykietuj /Cluster and label

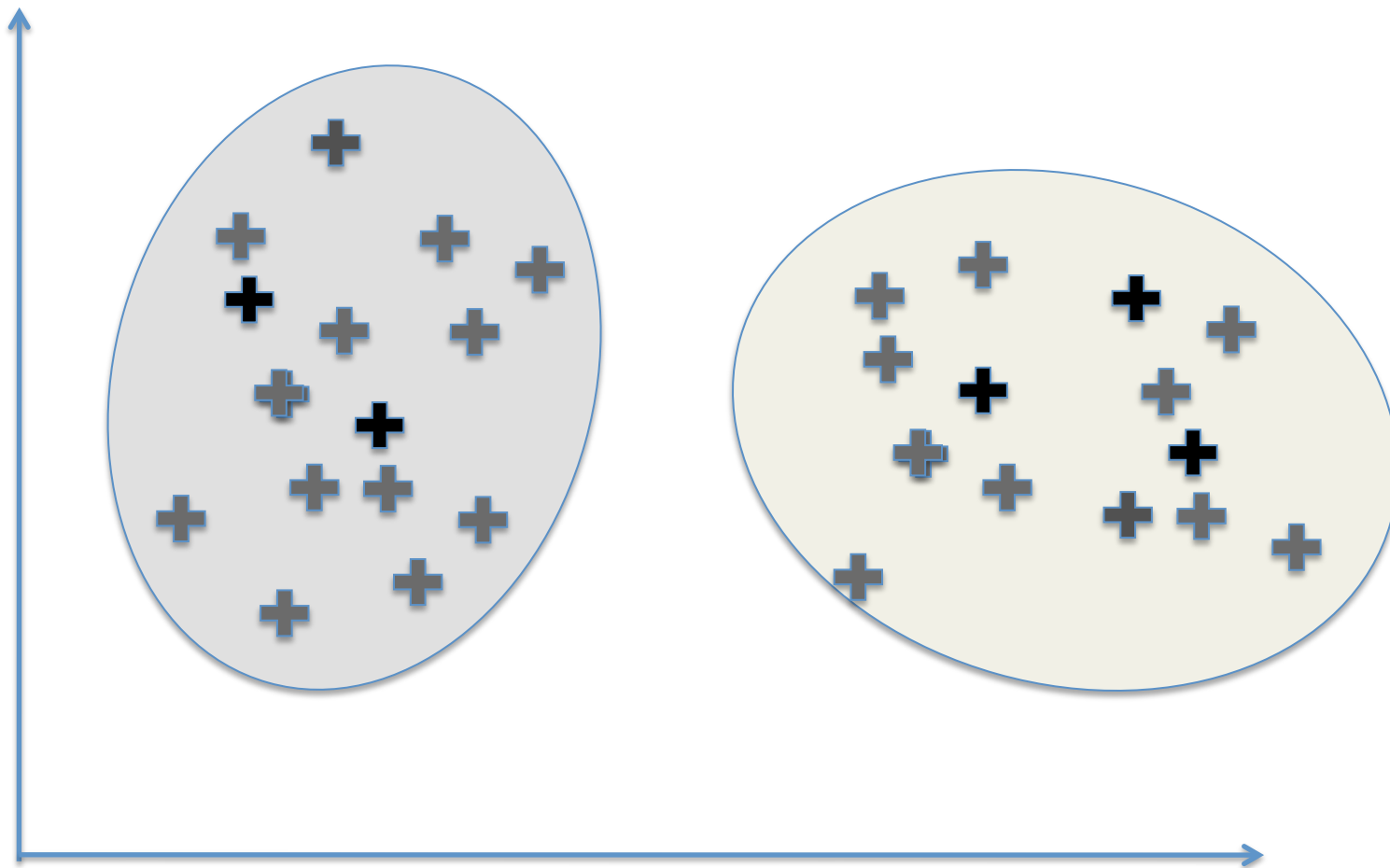
Input: etykietowane przykłady $L=\{x,y\}$ / n / oraz nieetykietowane $U=\{x\}$ / m / oraz algorytmy A grupowania i uczenia klasyfikatora LK

1. Znajdź skupienia w zbiorze nienadzorowanych x ($m+n$)
2. Dla każdego ze skupień, znajdź zbiór wszystkich etykietowanych przykładów S
3. Naucz algorytmem LK klasyfikator f na podstawie zbioru S
4. Zastosuj f do predykcji etykiet na pozostałych nieetykietowanych przykładów w skupieniu

Na koniec nauczyć klasyfikator na całym zbiorze przykładów (obejmującym zaetykietowane przykłady ze skupisk) -> **output**

Obserwacja: zakłada się, że granice skupisk są zgodne z granicami decyzyjnymi

SSL – grupowanie = założenia nt.
“kształtu” rozkładów



Pytanie o dobór algorytmu

- Liczba skupisk nie powinna być mniejsza niż liczba klas (ta jest znana z uwagi na zbiór L)
- Regularne sferyczne rozkłady = niektóre podejścia wykorzystując k-means / typowe cluster and label; w niektórych propozycjach także algorytmy gęstościowe
- Bardziej złożone kształty rozkładów = adaptacja modeli generatywnych, w szczególności **algorytmu EM** (przykłady etykietowane wykorzystane do początkowej identyfikacji składników mieszaniny rozkładów)

Basic Algorithm

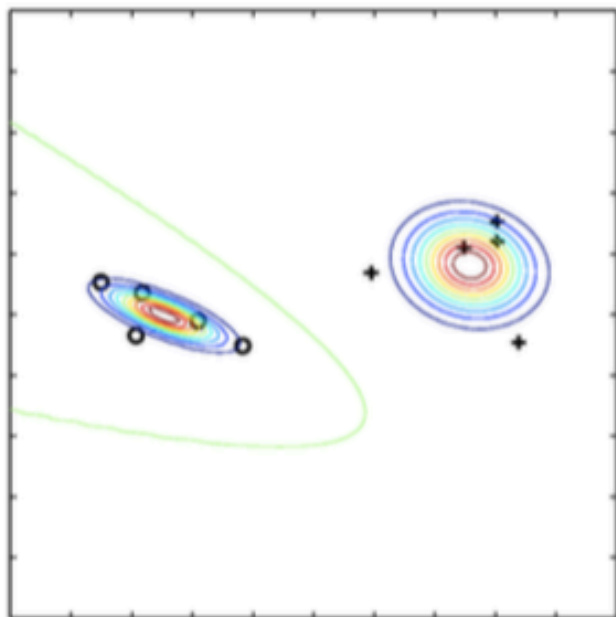
Algorithm EM(L, U)

- 1 Learn an initial naïve Bayesian classifier f from only the labeled set L (using Equations (27) and (28) in Chap. 3);
 - 2 **repeat**
 - // E-Step
 - 3 **for** each example d_i in U **do**
 - 4 Using the current classifier f to compute $\Pr(c_j|d_i)$ (using Equation (29) in Chap. 3).
 - 5 **end**
 - // M-Step
 - 6 learn a new naïve Bayesian classifier f from $L \cup U$ by computing $\Pr(c_j)$ and $\Pr(w_i|c_j)$ (using Equations (27) and (28) in Chap. 3).
 - 7 **until** the classifier parameters stabilize
- Return the classifier f from the last iteration.

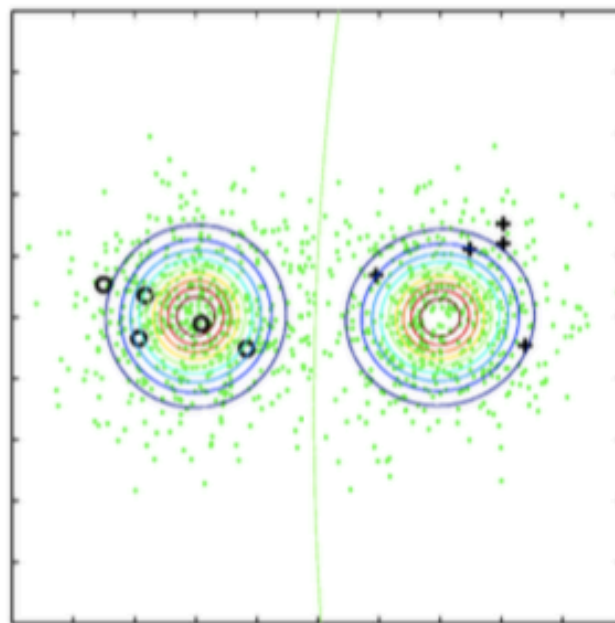
Fig. 5.1. The EM algorithm with naïve Bayesian classification

Wykorzystanie modeli generatywnych - algorytm EM w SSL z nieetykietowanymi danymi

only labeled data



with unlabeled data



from [Semi-Supervised Learning, ICML 2007 Tutorial; Xiaojin Zhu]

Podejścia Samo-uczenie się (ang. Self-learning)

Ucz się f z L i klasyfikuj przykłady U oceniając pewność predykcji

Wykorzystaj część najpewniejszych predykcji do rozszerzania zbioru uczącego:

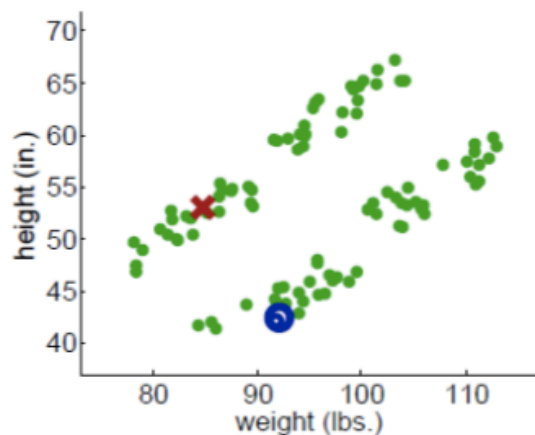
Algorithm 1 Self-training

```
1: repeat  
2:    $m \leftarrow \text{train\_model}(L)$   
3:   for  $x \in U$  do  
4:     if  $\max m(x) > \tau$  then  
5:        $L \leftarrow L \cup \{(x, p(x))\}$   
6: until no more predictions are confident
```

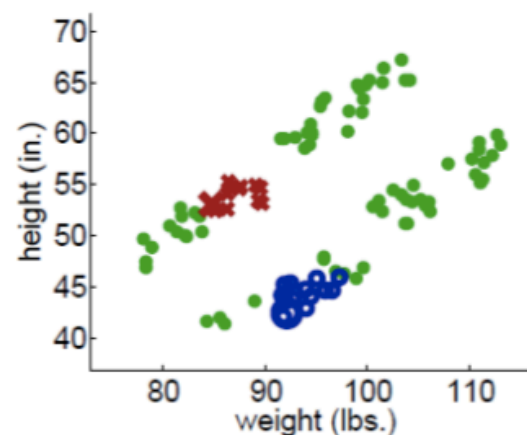
Podejście dość ogólne (typu wrapper) stosowalne wokół klasycznych algorytmów / Yarowsky, 1995; McClosky et al., 2006

Pomimo ciekawych zastosowań, złożone dane i niewłaściwe predykcje mogą źle ukierunkować kolejne iteracje

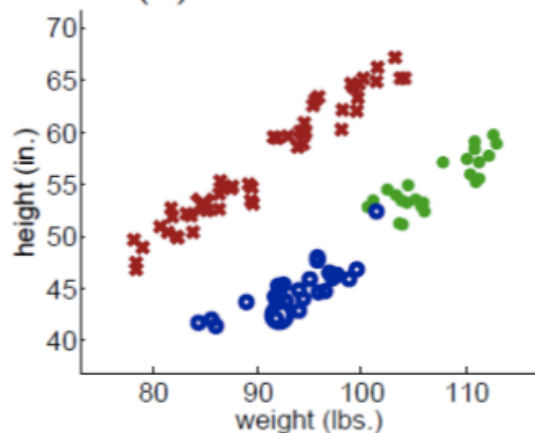
Przykład self-learning



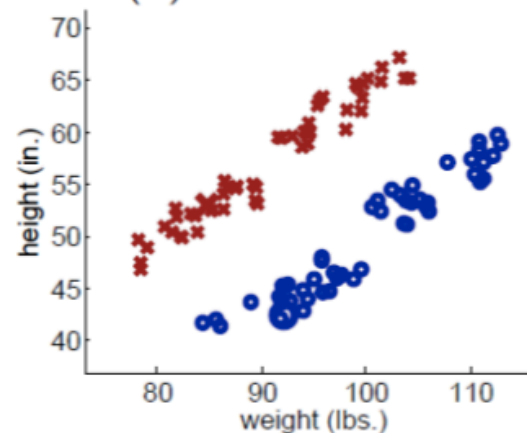
(a) Iteration 1



(b) Iteration 25



(c) Iteration 74



(d) Final labeling of all instances

Propagowanie etykiet z wykorzystaniem 1-NN

Samo-uczenie, cd.

Otwarte pytania:

Q1: tzw. Wrapper approach – zastanów się dlaczego

Q2: Parametryzacja – co i jak dobieramy?

Q3: Kiedy ma szanse na działanie (tzw. bootstrap classifier)?

Przygody barona Munchausena ->



Podejścia Self-learning - pytania

Podejście dość ogólne (typu wrapper) stosowalne wokół różnych klasyfikatorów – lecz w miarę skutecznych (mały błąd) z tzw. dobrym oszacowaniem marginesu pewności

Wybór miary pewności predykcji (wiele możliwości), także wybór warunku zatrzymania

Złożone dane i niewłaściwe predykcje mogą źle ukierunkować kolejne iteracje

“The main downside of self-training is that the model is unable to correct its own mistakes. If the model's predictions on unlabelled data are confident but wrong, the erroneous data is nevertheless incorporated into training and the model's errors are amplified.” S.Ruder

Inne znane zastosowania

Przetwarzanie języka naturalnego, np.

- Yarowsky - ujednoznacznienie sensu słowa w zależności od kontekstu / np. jakie jest znaczenia słowa ang. plant
- Rilof – identyfikacja tzw. subiektywnych rzeczowników w tekście
- Maeirezio i in: ocena emocjonalnego wydźwięku rejestrowanych dialogów pomiędzy ludźmi

Rozpoznawanie obrazów, np.

Rosenberg i in.: rozpoznawanie obiektów na zdjęciach

Inny paradygmat -> co-training

Co-training wprowadzony przez (Blum & Mitchell, 1998) (Mitchell, 1999) wykorzystuje:

- tzw. **dwa różne, komplementarne spojrzenia** (**views**) na dane uczące, które są przydatne do budowy klasyfikatorów
- zbiór atrybutów X podzielony na **dwa rozłączne** podzbiory $\{X_1, X_2\}$, każdy z nich dostarcza wystarczająco dużo informacji aby wytrenować poprawny klasyfikator przy odpowiednio dużej ilości danych

Hipoteza zgodności -> klasyfikatory wyuczone z części

c_1, c_2 s.t. $c_1(X_1)=c_2(X_2)=C^*(X)$ mogą być potencjalnie zgodne dla wystarczająco dużej liczby przykładów

Dwa niezależnie zbiory atrybutów

Naturalne w części zastosowań:

Obrazy: - różne metody przetwarzania, np. view 1 - pixel features; view 2 - Fourier coefficients

Teksty/emaile: nagłówek (header) vs. wnętrze (tekst)

Strony WWW = oryginalne zadania Blum & Mitchell (zawartość tekstowa strony vs. linki i ich opisy anchor text of hyperlinks pointing to the webpage)

Prof. Avrim Blum My Advisor




x - Link info & Text info

Prof. Avrim Blum My Advisor



x_1 - Text info

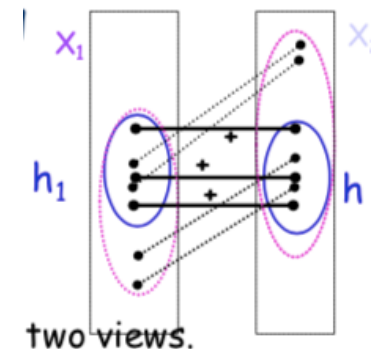
Prof. Avrim Blum My Advisor



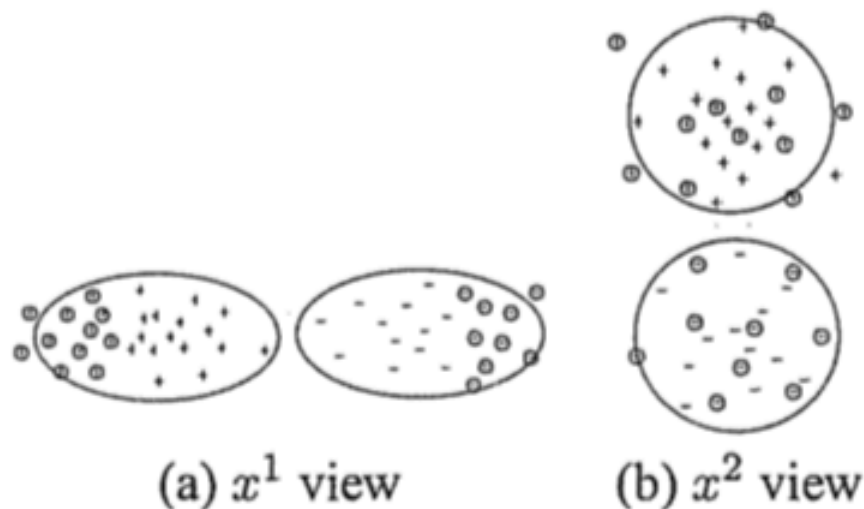
x_2 - Link info

Idea iteracyjnego uczenia się wzajemnego

- Każdy klasyfikator jest uczony niezależnie na etykietowanych przykładach ze swoim zbiorem atrybutów (L_1 oddzielnie L_2)
- Następnie każdy klasyfikator jest użyty do predykcji klasyfikacji puli nieetykietowanych przykładów U
- Jeden klasyfikator przekazuje kilka najpewniejszych “swoich predykcji” dla nieetykietowanych przykładów drugiemu klasyfikatorowi
- Zbiory uczące L_1 i L_2 są poszerzone i można ponownie wyuczyć klasyfikatory (z oddzielnymi zbiorami atrybutów)
- Idea: “Each view teaching (training) the other view”



Klasyfikatory powinny być odpowiednio skuteczne i pewne w etykietowaniu



Rysunek z pracy Zhu

Figure 3: Co-Training: Conditional independent assumption on feature split. With this assumption the high confident data points in x^1 view, represented by circled labels, will be randomly scattered in x^2 view. This is advantageous if they are to be used to teach the classifier in x^2 view.

Pożyteczne – jeśli najbardziej pewne predykcje jednego klasyfikatora odpowiadają tym przykładom dla których drugi klasyfikator nie był zbyt pewien

Co-training – uwagi metodyczne

Uwagi badaczy:

- Views – niezależne zbiory atrybutów [formalnie pełna niezależność w sensie statystycznym]
- Klasyfikatory - weakly-useful predictor, tzn. ma wyższe prawdopodobieństwo predykcji klasy pozytywnej na rzeczywistym przykładzie pozytywnym niż analogicznej predykcji pozytywnej na rzeczywistym przykładzie negatywnym.
- Posiadamy wystarczającą liczbę przykładów startowych – z poprawnymi etykietami (Labeled)
- Ciekawe powiązania z modelem teoretycznym tzw. PAC learning [Arvin Blum]

Więcej w badaniach M.Balcan praca nt : Direct Optimization of Agreement between h_1 and h_2

Przykłady eksperymentów [Blum, Mitchell]

- Startowy zbiór L zawierający 12 etykietowanych stron WWW (Faculty vs not)
- Ponadto 1,000 nie zaetykietowanych przykładów
- average error: learning from labeled data 11.1%;
- average error: co-training 5.0%

	Page-base classifier	Link-based classifier	Combined classifier
Supervised training	12.9	12.4	11.1
Co-training	6.2	11.6	5.0

Inne zastosowania - obrazy

Results: images [Levin-Viola-Freund '03]:

- Visual detectors with different kinds of processing

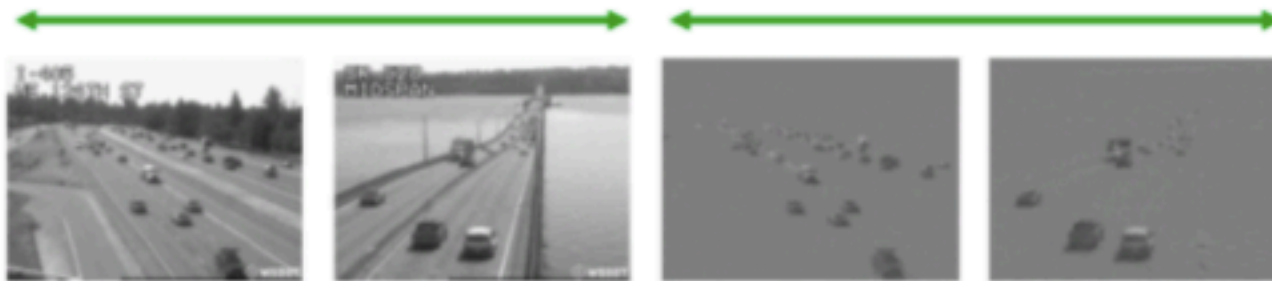
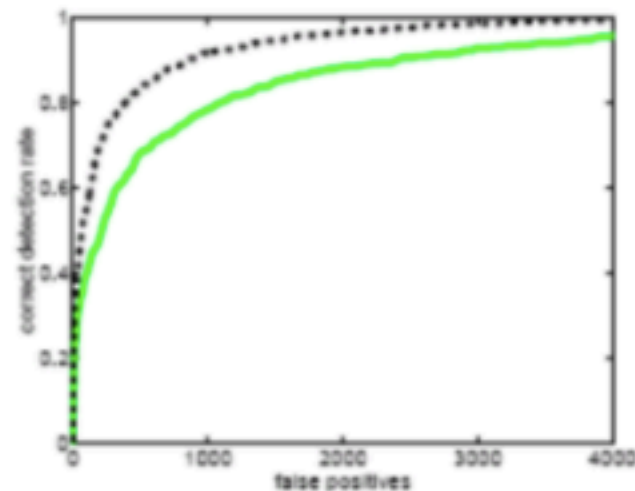


Figure 1: Example images used to test and train the car detection system. On the left are the original images. On the right are background subtracted images.

- Images with 50 labeled cars. 22,000 unlabeled images.
- Factor 2-3+ improvement.

From [LVF03]



Multi-view learning- rozszerzenie

- Więcej niż dwa spojrzenia na dane (czyli niezależne zbiory atrybutów)
- Naucz niezależne klasyfikatory na oddzielnych zbiorach (views – zbiory atrybutów)
- Modus operandi: predykcje klasyfikatorów wystarczająco zgodne na przykładach nieetykietowanych
- Dla przykładów testowych – agregacja **głosowania** jak w zespole klasyfikatorów
- Co-training może być widziany specjalny pod-przypadek

Spojrzyć do: J.Zhao et al: Multi-view Learning Overview: Recent Progress and New Challenges.

Podejścia inspirowane

Spójrz do bloga Sebastian Ruder'a [An overview of proxy-label approaches for semi-supervised learning](#)

Powiązane z multi-view training

- Democratic co-training
- Tri-training
- Tri-training with disagreement
- Asymmetric tri-training
- Multi-task tri-training
-

Democratic co-training

Wielokrotne uczenie zróżnicowanych klasyfikatorów (z tzw. Innymi inductive bias) z L i ocena ich predykcji na U

M – zbiór klasyfikatorów z tą samą predykcją j -tej klasy dla przykładu x

w – ocena pewności predykcji klasyfikatora

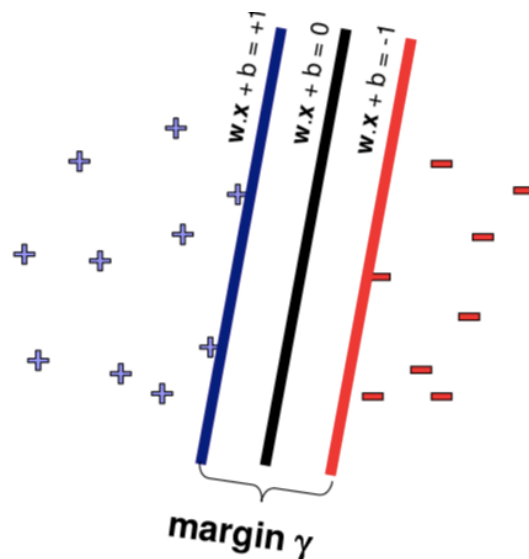
Przykłady nieetykietowane dla które większość klasyfikatorów osiąga wysoką zgodność mogą być dodane do zbioru uczącego

Algorithm 3 Democratic Co-learning

```
1: repeat
2:   for  $i \in \{1..n\}$  do
3:      $m_i \leftarrow \text{train\_model}(L)$ 
4:   for  $x \in U$  do
5:     for  $j \in \{1..C\}$  do
6:        $M \leftarrow \{i \mid p_i(x) = j\}$ 
7:       if  $|M| > n/2$  and  $\sum_{i \in M} w_i > \sum_{i \notin M} w_i$  then
8:          $L \leftarrow L \cup \{(x, j)\}$ 
9: until none of  $m_i$  changes
10: apply weighted majority vote over  $m_i$ 
```

SVM w obecności częściowo etykietowanych

Przypomnijmy zasady klasycznego SVM



W trudniejszych nakładających się rozkładach przykładów element regularyzacyjny ze zmiennymi osłabiającymi

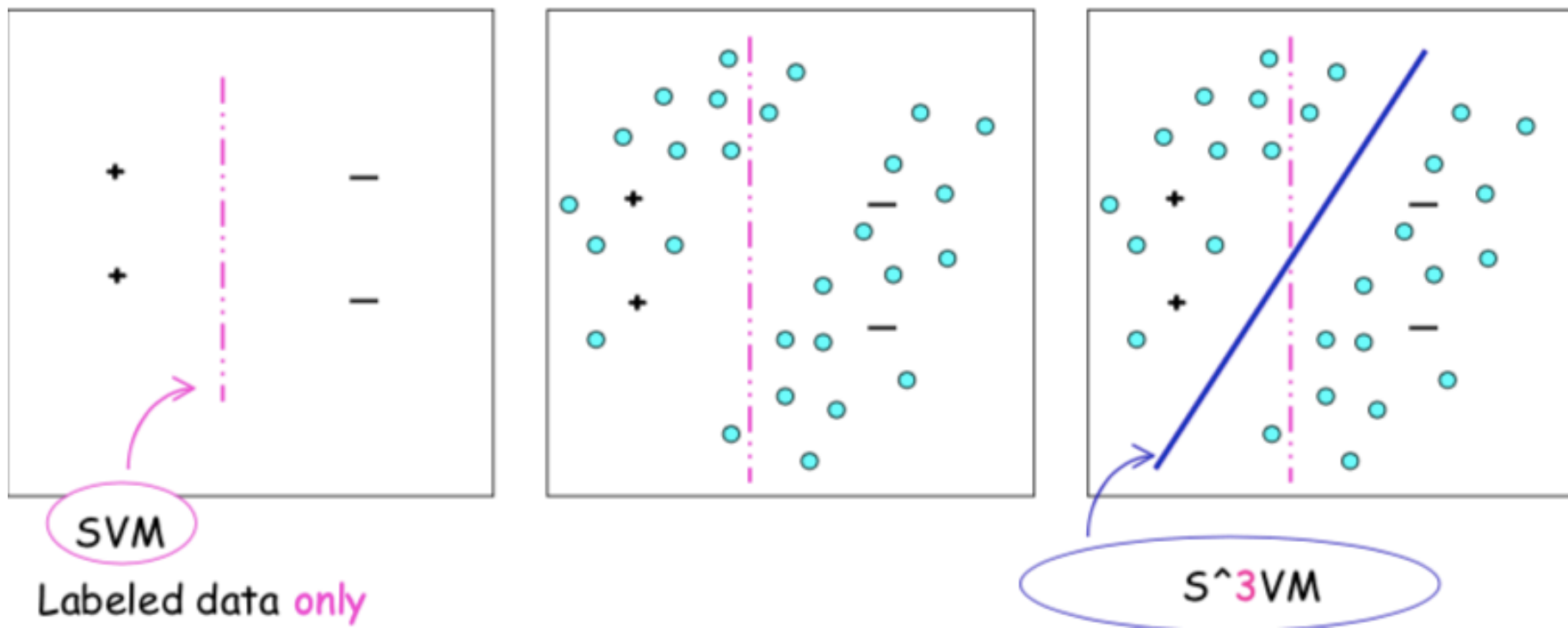
$$L(w) = \frac{\|\vec{w}\|^2}{2} + C \left(\sum_{i=1}^N \xi_i^k \right)$$

$$f(\vec{x}_i) = \begin{cases} 1 & \text{if } \vec{w} \cdot \vec{x}_i + b \geq 1 - \xi_i \\ -1 & \text{if } \vec{w} \cdot \vec{x}_i + b \leq -1 + \xi_i \end{cases}$$

Semi-supervised SVM

Granica decyzyjna powinna przejść przez **obszar o małej gęstości przykładów** (etykietowanych i nieetykietowanych) pomiędzy dwoma klasami (-1;1)

Rozszerzenie sformułowania zadania programowania mat. SVM, aby uwzględnić obecność nieetykietowanych przykładów



SVM -> S3VM

Zadanie optymalizacji SVM

$$L(w) = \frac{\|\vec{w}\|^2}{2} + C \left(\sum_{i=1}^l \xi_i \right)$$

$$y_i(w x_i + b) \geq 1 - \xi_i \quad i = 1, \dots, l$$

Funkcja celu jest przetworzona z wykorzystaniem funkcji straty (Hinge loss), która zastępuje zmienną osłabiającą (ξ)

$$\min \left(\frac{1}{2} \|\mathcal{W}\|^2 + C \sum_{i=1}^l \max \left(1 - y_i |w \cdot x_i + b|, 0 \right) \right)$$

za: Ding. S. et al. An overview on semi-supervised support vector machine

S3VM

Przykłady nieetykietowane – analogiczna funkcja straty

$$\max(1 - |f(x)|, 0)$$

- ma wartość dodatnią, gdy $-1 < f(x) < 1$ oraz 0 poza tym.

Odpowiednik kary za naruszenie separowalności marginesu

Dla wszystkich przykładów nieetykietowanych w funkcji celu pojawi się element

$$\frac{1}{\mathcal{U}} \sum_{i=l+1}^{l+u} \max(1 - |f(x)|, 0)$$

Co prowadzi do re-definicji zadania programowania matematycznego

Trudne do rozwiązania – nieciągłe zadanie programowania kwadratowego

Przeformułowanie S3VM w pracy [Bennett and Demiriz]

Transductive Reasoning

Komentarz:

"Transductive" means here "reasoning from particular to particular" as opposed to "Inductive" - "reasoning from particular to general". So the motivation is that the performance of the model outside the current set of labeled and unlabeled points is not of interest

lub [Wikipedia]

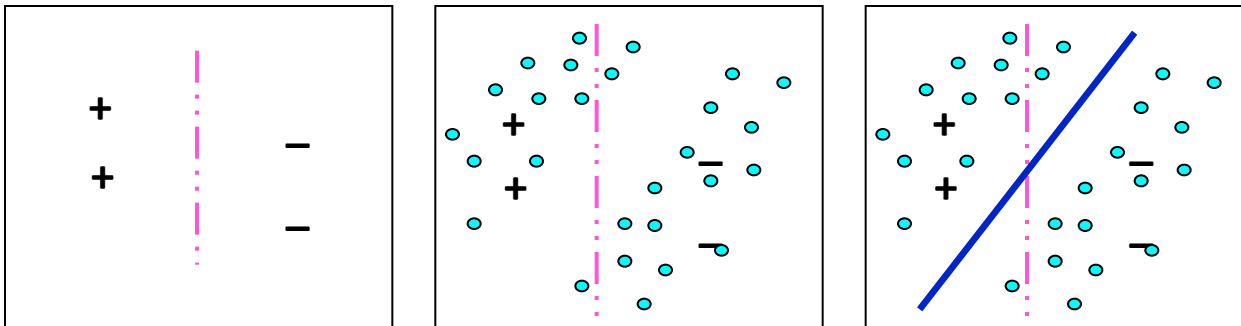
Transduction or transductive inference is reasoning from observed, specific (training) cases to specific (test) cases.

jako przeciwieństwo klasycznej zasady indukcji :

induction is reasoning from observed training cases to general rules, which are then applied to the test cases

SVM [Joachims]

- Hiperpłaszczyzna powinna przejść przez obszar małej gęstości przykładów (L+U) starając się utrzymać możliwie szeroki margines
- W ogólności trudny problem do optymalizacji
- Alternatywne podejście (z transdukcją)
 - Zbuduj szeroki margines z etykietowanymi przykładami L. Etykietuj przykłady z U.
 - Spróbuj przełączyć etykiety niektórych z przykładów z U jeśli jest to korzystane.



Transductive SVM [Joachims]

Wykorzystuje optymalizację na podzbiorze przykładów L i U i próbuje zaetykietować część przykładów U (tak jakby były testowe)

Następnie przełącza się etykiety niektórych przykładów, jeśli powiększają margines

1. Train SVM on labeled data
2. Classify the unlabeled data using the SVM, assign q examples with highest $f(x)$ examples to the positive class, rest to negative class (q is user-set parameter)
3. Initialize $C^* = 1E - 5$ (small value, almost disregarding unlabeled data points)
4. Iterate while increasing slack coefficient C^* of unlabeled data points:
 - 4.1 Find two unlabeled examples that have different predicted labels, and a total Hinge loss > 2
 - 4.2 Swap the predicted labels of the examples \implies decrease Hinge loss
 - 4.3 Increase slack parameter C^* and retrain SVM

Transductive SVM

Formalizacja Joachims [99] wykorzystuje założenia etykietyzacji podobnych obiektów i podobny problem optymalizacyjny

Input: $S_l = \{(x_1, y_1), \dots, (x_{m_l}, y_{m_l})\}$

$S_u = \{x_1, \dots, x_{m_u}\}$

$$\operatorname{argmin}_w ||w||^2 + C \sum_i \xi_i + C \sum_u \widehat{\xi}_u$$

- $y_i w \cdot x_i \geq 1 - \xi_i$, for all $i \in \{1, \dots, m_l\}$
- $\widehat{y}_u w \cdot x_u \geq 1 - \widehat{\xi}_u$, for all $u \in \{1, \dots, m_u\}$
- $\widehat{y}_u \in \{-1, 1\}$ for all $u \in \{1, \dots, m_u\}$

Zmienne osłabiające ξ – funkcje straty Hinge loss – jak poprzednio

Parę uwag nt. implementacji SVM

Sklearn Python – tylko wersje label propagation (grafowe)

Indywidualne projekty – semisup-learn for Python - patrz listy dla mat. dodatkowe

SVM – dostępny w nowych bibliotekach R -> RSSL Package

Package ‘RSSL’

February 4, 2020

Version 0.9.1

Title Implementations of Semi-Supervised Learning Approaches for Classification

Depends R(>= 2.10.0)

Imports methods, Rcpp, MASS, kernlab, quadprog, Matrix, dplyr, tidyr, ggplot2, reshape2, scales, cluster

LinkingTo Rcpp, RcppArmadillo

Suggests testthat, rmarkdown, SparseM, numDeriv, LiblineaR

Description A collection of implementations of semi-supervised classifiers and methods to evaluate their performance. The package includes implementations of, among others, Implicitly Constrained Learning, Moment Constrained Learning, the Transductive SVM, Manifold regularization, Maximum Contrastive Pessimistic Likelihood estimation, S4VM and WellSVM.

License GPL (>= 2)

URL <http://www.github.com/jkrijthe/RSSL>

BugReports <http://www.github.com/jkrijthe/RSSL>

Inne zagadnienia

Podejścia grafowe i propagacja etykiet – pomijamy w tym wykładzie

Inne metody, nieomawiane w tej edycji przedmiotu

- Zaawansowane metody generatywne (np. wykorzystujące Hidden Markov Models)
- Sieci neuronowe
 - w szczególności: GAN - Generative Adversarial Networks - w uczeniu częściowo nadzorowanym (z małą liczbą etykietowych obrazów)
 - oraz inne DANN, i tzw. Restricted Boltzmann Machines

Literatura i materiały dodatkowe

Artykuły:

J.Engelen, H.Hoos: A survey on semi-supervised learning. Machine learning 2020.

Xiaojin Zhu: Semi-Supervised Learning Literature Survey.

Jing Zhao, Xijiong Xie, Xin Xu, Shiliang Sun: Multi-view Learning Overview: Recent Progress and New Challenges

Avrim Blum, Tom Mitchell: Combining Labeled and Unlabeled Data with Co-Training. COLT 1998

Książka: Chapelle, Olivier; Schölkopf, Bernhard; Zien, Alexander (2006). Semi-supervised learning. Cambridge, Mass.: MIT Press

Wykłady:

P.Rai: CS5350 / 6350 Semi-supervised Learning

N.Balcan: Semi-supervised Learning [inspiracja dla obecnego wykładu]

A.Zien: Semi-supervised learning: Tutorial at ANN summer school

Barnabas Póczos: Semi-supervised learning

S.Ruder: An overview of proxy-label approaches for semi-supervised learning [blog]

Software:

Głównie indywidualne projekty

np. semisup-learn for Python, RSSL dla R

Przeglądowa strona WWW

Soft Computing and Intelligent Information Systems




A University of Granada research group

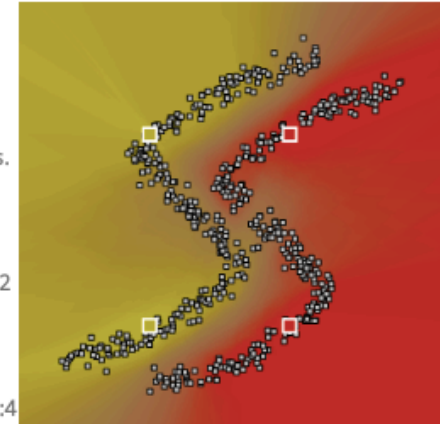
Research Publications Teaching Thematic Sites Software Awards In the Press

[Home](#) » [Thematic Sites](#) » Semi-supervised Classification: An Insight into Self-Labeling Approaches

Semi-supervised Classification: An Insight into Self-Labeling Approaches

This Website contains SCI²S research material on Semi-Supervised Classification. This research is related to the following SCI²S papers published recently:

- **I. Triguero, S. García, F. Herrera**, *Self-Labeled Techniques for Semi-Supervised Learning: Taxonomy, Software and Empirical Study*. Knowledge and Information Systems, COMPLEMENTARY MATERIAL to the paper [here](#): datasets, experimental results and source codes. doi: 10.1007/s10115-013-0706-y, in press (2014). 
- **I. Triguero, José A. Sáez, J. Luengo, S. García, F. Herrera**, *On the Characterization of Noise Filters for Self-Training Semi-Supervised in Nearest Neighbor Classification*. Neurocomputing 132 (2014) 30-41, doi: 10.1016/j.neucom.2013.05.055 
- **I. Triguero, S. García, F. Herrera**, *SEG-SSC: A Framework based on Synthetic Examples Generation for Self-Labeled Semi-Supervised Classification*. IEEE Transactions on Cybernetics 45:4 (2015) 622-634, doi: 10.1109/TCYB.2014.2332003. 



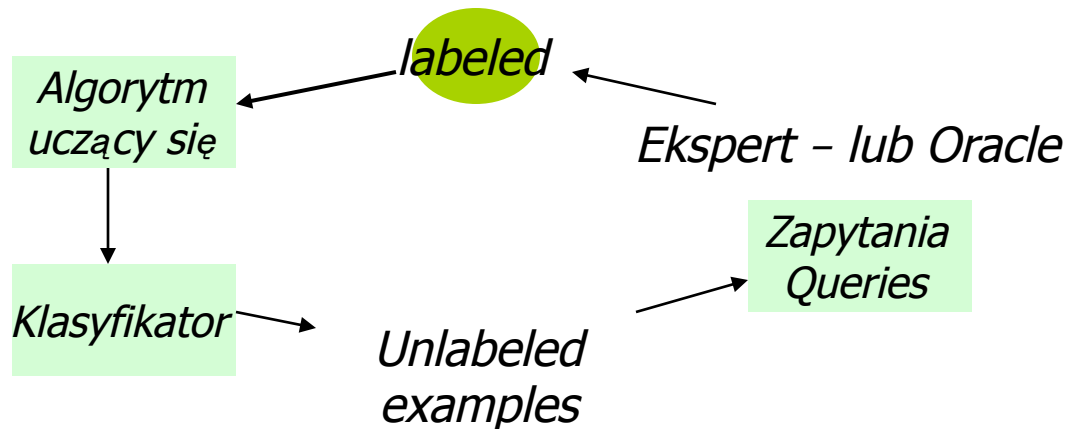
The web is organized according to the following Summary:

1. [Introduction to Semi-Supervised Classification](#)
2. [Self-Labeled Classification](#)
 - A. [Background](#)
 - B. [Taxonomy](#)

<https://sci2s.ugr.es/ssl> - repozytorium Univ. Granada

SSL nie jest aktywnym uczeniem się

- Także zbiór etykietowanych i niezaetykietowanych przykładów
- **Active Learning:**
 - Algorytm na wpływ na etykietowanie przykładów
- Ma możliwości wyboru części niezaetykietowanych przykładów (najbardziej informatycznych dla klasyfikacji)
- Jest w stanie zadać pytanie ekspertowi / wyroczni o tzw. prawdziwą wartość etykiety wybranych przykładów
- Jak wybierać przykłady?



Pytanie i komentarze?

Dalszy kontakt:

jerzy.stefanowski@cs.put.poznan.pl

<http://www.cs.put.poznan.pl/jstefanowski/>



**Fundusze
Europejskie**
Polska Cyfrowa



**Rzeczpospolita
Polska**

Unia Europejska
Europejski Fundusz
Rozwoju Regionalnego

