

Metody szacowania zdolności kredytowej klientów indywidualnych

Mateusz Kobos
15.11.2006

Spis treści

- Czym jest „Credit Scoring” (CS)?
 - Analizowane dane
 - Zalety i ograniczenia CS
 - CS w praktyce
 - CS jako zagadnienie Machine Learning
 - Dane
 - Trendy w CS
- Metody klasyczne
 - Linear Discriminant Analysis (LDA)
 - Logistic Regression
- Inne

Czym jest „Credit Scoring”

- Definicja
 - Metoda służąca do oceny ryzyka związanego z udzieleniem pożyczki
 - Metoda służąca do przewidzenia, czy osoba składająca wniosek o przyznanie kredytu nie wywiąże się z płatności lub będzie zalegać z płatnością.
- Wprowadzono w 1950

Analizowane dane [4]

- Co bierze się pod uwagę?
 - np. przychód miesięczny, zadłużenie, posiadane aktywa, długość pracy w tym samym miejscu, czy były problemy ze spłatą poprzedniej pożyczki, czy kandydat posiada dom, typ konta bankowego itp.
- Przy tworzeniu modelu bierze się pod uwagę 50-60 zmiennych (a czasami i 1000), w końcowym modelu jest użytych:
 - 12 (Fair, Isaac and Co.)
 - 48 (First Data Resources)

Analizowane dane (c.d.) [4]

- Patrick McAllister i John Mingo szacują, że by stworzyć model predykcyjny do pożyczek komercyjnych, trzeba zebrać ok 20000-30000 aplikacji.

Zalety i ograniczenia CS [4]

- Zalety
 - zmniejszenie czasu analizy wniosku
 - Kredyt konsumencki: 9 dni -> 3 dni w jednym z kanadyjskich banków [4]
 - Kredyt dla małych firm: 3-4 tygodnie -> kilka godzin (Barnett Bank) [4]
 - zwiększona obiektywność
 - Przydaje się w przestrzeganiu „Equal Credit Opportunity Act” (implemented by the Federal Reserve Board’s Regulation B)
 - Zakaz dyskryminacji ze względu na: rasę, kolor skóry, wyznawaną religię, pochodzenie narodowe, płeć, stan cywilny, wiek

Zalety i ograniczenia CS [4]

- Ograniczenia i problemy
 - trzeba dbać, by „zbiór uczący” (przykłady) był zbliżony do „zbioru testującego” (właściwe zastosowanie)
 - kampania reklamowa może zmienić „zbiór testujący”
 - samo wprowadzenie nowego modelu zmienia „zbiór testujący”
 - trzeba często uaktualniać model (np. co 1-5 lat) przy pomocy nowych danych (zmieniające się warunki socjoekonomiczne)
 - przy tworzeniu modelu należy brać pod uwagę nie tylko osoby, którym został przyznany kredyt (i należą do zbioru uczącego), ale również te, które go nie otrzymały

Zalety i ograniczenia CS [4]

- Ograniczenia i problemy c.d.
 - dobry model powinien dobrze działać zarówno w dobrych jak i złych warunkach ekonomicznych
 - należy pamiętać, że nie ma modeli idealnych
 - Przypadek Lawrence Lindsey'a (Federal Reserve System Governor) i jego wniosku o kartę kredytową „Toys 'R' Us”

CS w praktyce

- Pod uwagę bierze się ankietę wypełnianą przez potencjalnych kredytobiorców oraz dane z biur kredytowych (*ang. credit bureaus*) (4 w USA)

CS jako zagadnienie Machine Learning

- Zagadnienie klasyfikacji (w klasycznej postaci CS)
- Należy wziąć pod uwagę różną wagę False Positives i False Negatives
- Najlepiej, by model potrafił wyjaśnić swoją decyzję

Dane

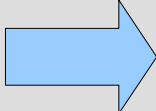
- B. mało danych
- Publicznie dostępne:
 - Standardowe zbiory danych :
 - Credit Approval (Australian Credit Approval)
 - 690 przykładów, 15 atrybutów (nominalne i liczbowe), 5% przykładów ma brakujące atrybuty,
 - German Credit
 - 1000 przykładów, 20 atrybutów (7 liczbowych, 13 nominalnych), w wersji numerycznej 24 atrybutów
 - Inne
 - Japanese Credit Screening Database
 - 125 przykładów, 9 atrybutów, przykłady w formacie LISP-owym
 - zawiera reguły decyzyjne stworzone przez ekspertów
 - PKDD'99 Discovery Challenge
 - 8 tabel z bazy danych Czeskiego banku (w sumie 67M danych)

Trendy w CS

- Przejście z szacowania zdolności kredytowej, do optymalizacji relacji z klientem (a nawet jeszcze dalej)
- Modelowanie ryzyka kredytowego portfela (zbioru) klientów a nie pojedynczego klienta (zgodnie z wytycznymi Basel II)
 - Podział portfela na segmenty

Linear Discriminant Analysis (Fisher 1936)

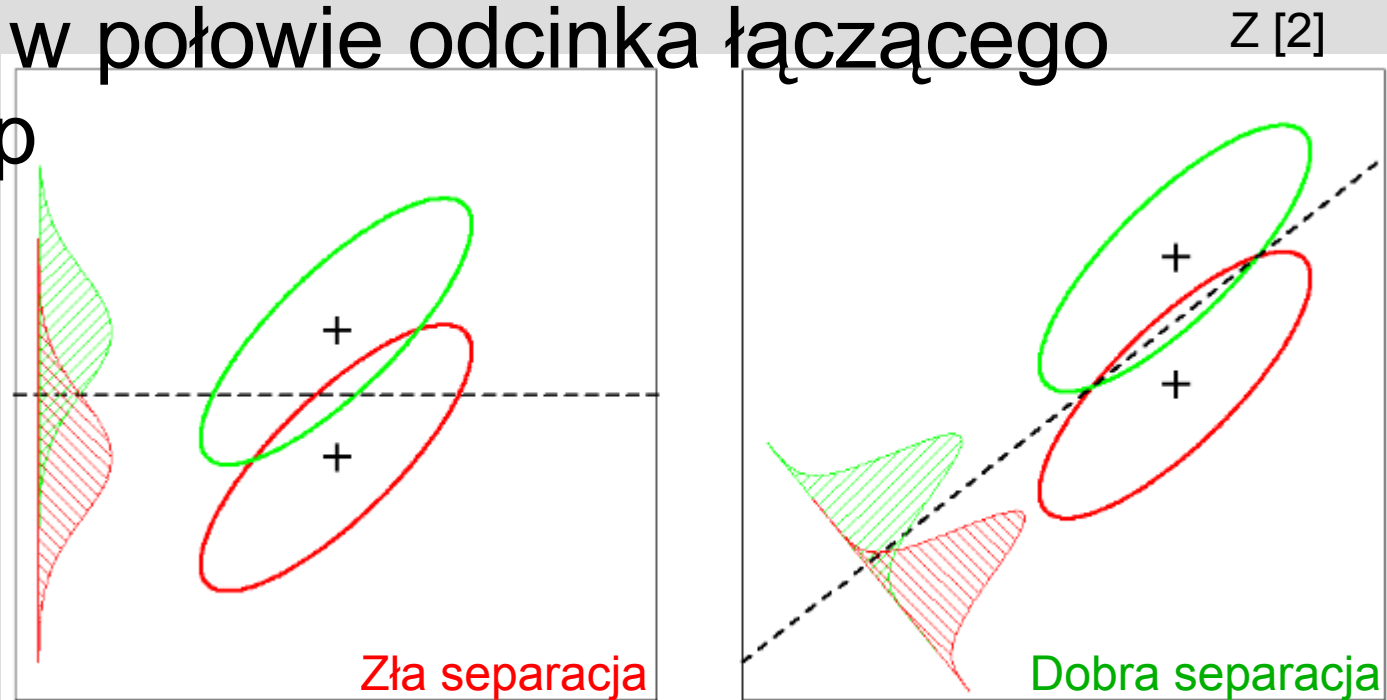
- Na jakie pytania może odpowiedzieć:
 - Jak wygląda (hiper-)płaszczyzna separująca grupy (klasy) przykładów?
 - Jak zredukować wymiary, by jak najlepiej była zachowana jakość klasyfikacji (z pomocą płaszczyzny decyzyjnej) ?
- Wady:
 - Dla obu zbiorów zakładamy:
 - Rozkład normalny
 - Równe macierze kowariancji



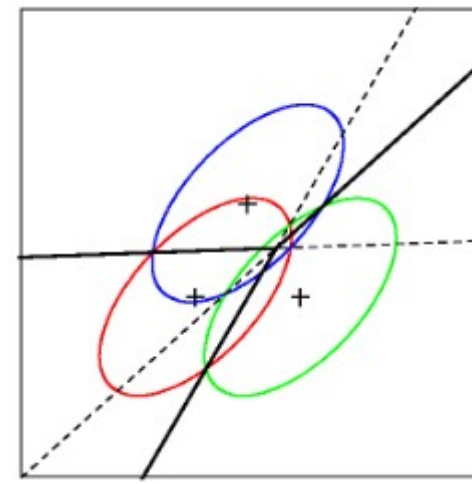
Intuicyjnie: rozkład
każdej z
grup to elipsoida
o tym samym kształcie

LDA - algorytm

- Szukamy wektora „kanonicznego”, po rzutowaniu na który grupy są jak najlepiej separowalne (w prosty sposób można to zrobić obliczając wektor własny o największej wartości własnej pewnej macierzy $W^{-1}M$)
- Bierzemy płaszczyznę prostopadłą do wektora „kanonicznego”, przechodzącą przez punkt znajdujący się w połowie odcinka łączącego środki obu grup



LDA - modyfikacje

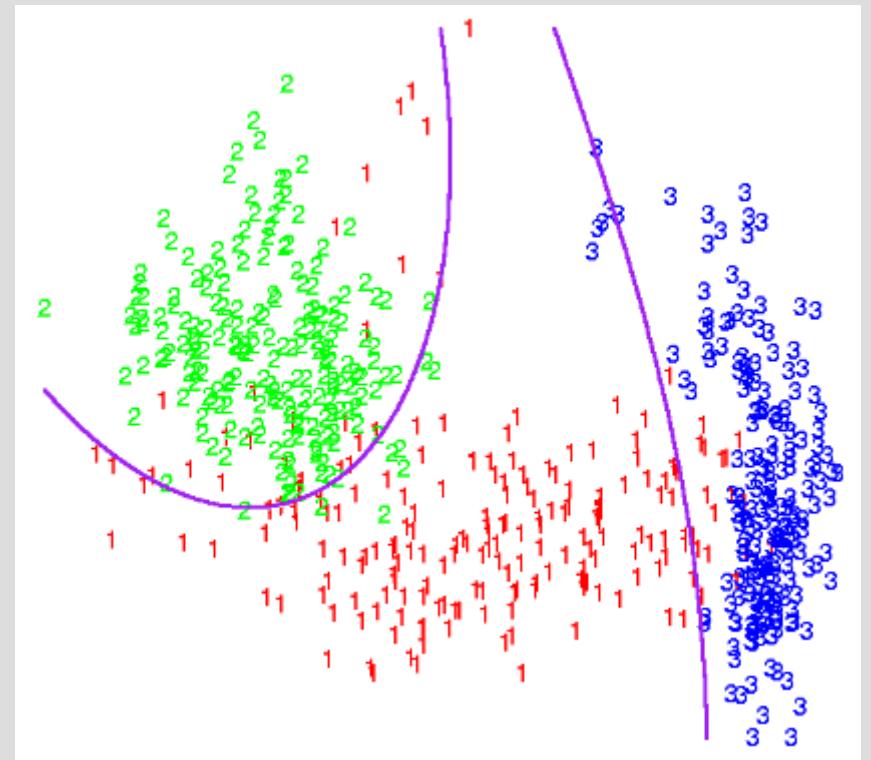


Z [2]

- >2 klasy:
 - 1. bierzemy pod uwagę nie tylko macierz kowariancji wewnątrzgrupowej, ale również międzygrupowej (gdzie pojedynczymi elementami są średnie grup, a wartością średnią - średnia dla wszystkich grup)
 - 2. obliczamy płaszczyzny separujące dla każdej pary grup i za ich pomocą „wykrajamy” odpowiednie części przestrzeni należące do każdej z grup

LDA – modyfikacje (c.d.)

- Wersja „kwadratowa”
 - zakładamy, że macierze kowariancji są różne => funkcja dyskryminacyjna staje się kwadratowa



Regresja liniowa

- Podstawowe wzory:

Wzór na prostą regresji:

$$x = w_0 + w_1 a_1 + w_2 a_2 + \dots + w_k a_k$$

Minimalizujemy
sumę kwadratów odchyłeń:

$$\sum_{i=1}^n \left(x^{(i)} - \sum_{j=0}^k w_j a_j^{(i)} \right)^2$$

- Za pomocą regresji liniowej można dokonywać klasyfikacji dla 2 grup:
 - Grupa 1 - zmienna objaśniana $x=0$
 - Grupa 2 – zmienna objaśniana $x=1$

Regresja liniowa (c.d.)

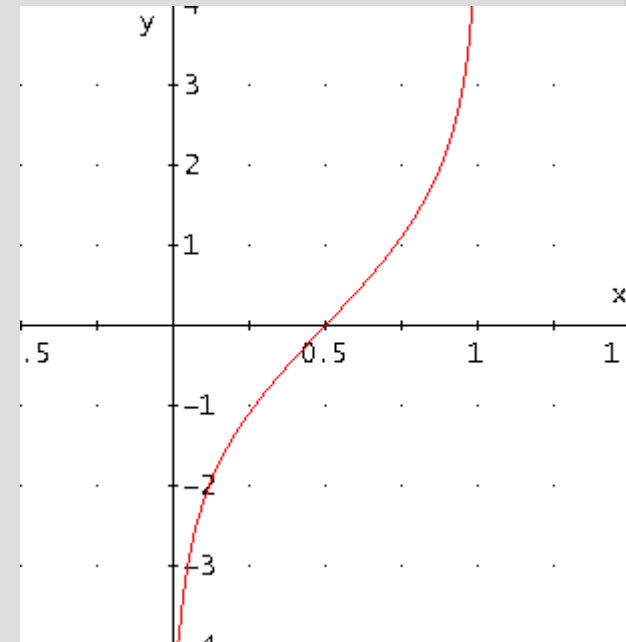
- Wady:
 - x nie może być interpretowana jako prawdopodobieństwo (wartości mogą być poza przedziałem $[0, 1]$)
 - Stosując metodę najmniejszych kwadratów zakładamy, że błędy mają rozkład normalny z o tej samej wariancji (przy binarnej klasyfikacji jest to niespełnione)

Logistic regression

- Regresja logistyczna to rozszerzona wersja regresji liniowej
- Podmieniamy zmienną wyjściową x :

$$\ln\left(\frac{p}{1-p}\right) = w_0 + w_1 a_1 + \dots + w_k a_k$$

- Gdzie:
 - p : prawdopodobieństwo należenia przykładu do 1 grupy
 - $1-p$: prawdopodobieństwo należenia przykładu do 2 grupy
- Wartość lewej strony równania jest teraz dowolna – a p jest z przedziału $(0, 1)$



Funkcja $\ln(x/(1-x))$:

Logistic Regression

- Przekształcona zmienna (x) jest aproksymowana przez funkcję liniową
- Z poprzedniego wzoru można obliczyć:

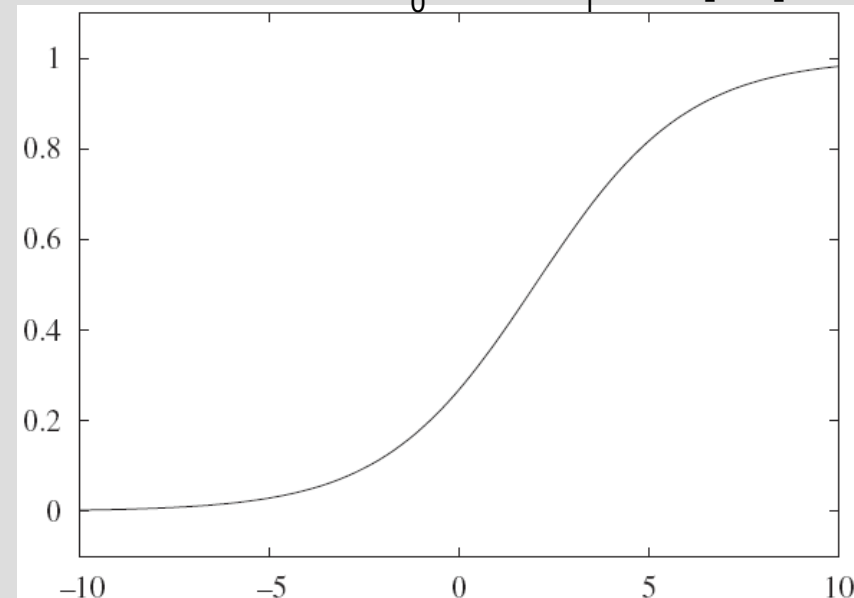
$$p = \frac{1}{1 + \exp(-w_0 - w_1 a_1 - \dots - w_k a_k)}$$

Przykład takiej funkcji dla 1 wymiaru z wagami $w_0 = 0.5$, $w_1 = 1$ z [X4]:

- Punkty spełniające $p=0.5$ znajdują się na krzywej decyzyjnej =>

$$-w_0 - w_1 a_1 - \dots - w_k a_k = 0$$

jest płaszczyzną decyzyjną



Logistic Regression

- By obliczyć parametry (wagi w_i) zgodnie z metodą największej wiarygodności maksymalizujemy: $\prod_{i=1}^N p^{(i)}$
- (maksymalizujemy np. za pomocą metod Newtonowskich)

Logistic Regression

- Zalety:
 - Parametry (wagi) modelu są interpretowalne
 - Nie wymaga założeń modelu LDA
- Wady
 - Parametry trudniejsze w obliczeniu od parametrów LDA
 - Duże błędy przy danych o zależnościach nieliniowych
- Ciekawostka:
 - Podobieństwo do modelu perceptronu
- Rozszerzenie
 - Model dla wielu klas

Inne metody

- Sieci neuronowe
- Drzewa decyzyjne
- k – Nearest Neighbor
- Programowanie liniowe
- Support Vector Machines
- Survival Analysis
- metaklasyfikatory

Przykładowe wyniki

- Wyniki dotyczące dokładności klasyfikacji z (zebrane z różnych publikacji) – testowane na zbiorze testowym

Table 1 Classification accuracy of different approaches

<i>Linear regression</i>	<i>Logistic regression</i>	<i>Class trees</i>	<i>LP</i>	<i>Neural nets</i>	<i>GA</i>
43.4	43.3	43.8	—	—	—
77.5	—	75	74.7	—	—
87.5	89.3	93.2	86.1	—	—
68.4	—	62.3	—	62.0	64.5
66.5	67.3	67.3	—	6.4	—

Przykładowe wyniki (c.d.)

- Wyniki porównujące wydajność sieci neuronowych z innymi metodami - podane są błędy mierzone za pomocą 10-fold

Cross Validation Good credit - „dobrzy” kredytobiorcy sklasyfikowani jako „źli”,
 Z [7] Bad crediti - analogicznie

	German credit data ^b			Australian credit data ^b		
	Good credit	Bad credit	Overall	Good credit	Bad credit	Overall
Neural models ^a						
MOE	0.1428	0.4775	0.2434	0.1457	0.1246	0.1332
RBF	0.1347	0.5299	0.2540	0.1315	0.1274	0.1286
MLP	0.1352	0.5753	0.2672	0.1540	0.1326	0.1416
LVQ	0.2493	0.4814	0.3163	0.1710	0.1713	0.1703
FAR	0.4039	0.4883	0.4277	0.2566	0.2388	0.2461
Parametric models						
Linear discriminant	0.2771	0.2667	0.2740	0.0782	0.1906	0.1404
Logistic regression	0.1186	0.5133	0.2370	0.1107	0.1409	0.1275
Non-parametric models						
<i>K</i> nearest neighbor	0.2257	0.5533	0.3240	0.1531	0.1332	0.1420
Kernel density	0.1557	0.6300	0.3080	0.1857	0.1514	0.1666
CART	0.2063	0.5457	0.3044	0.1922	0.1201	0.1562

Bibliografia

- [1] Dreiseitl S., Ohno-Machado L., „Logistic regression and artificial neural network classification models: a methodology review”, Journal of Biomedical Informatics 35, 352–359, 2002
- [2] Friedman J., Hastie T., Tibshirani R., „The elements of statistical learning”, Springer, 2003
- [3] Koronacki, Ówik, "Statystyczne systemy uczące się", WNT, 2005
- [4] Mester L.J., „What's the point in credit scoring”, Business review, Federal Reserve Bank of Philadelphia, September/October 1997
- [5] Thomas L.C., Oliver R.W., Hand D.J., „A survey of the issues in consumer credit modelling research”, Journal of the operational research society 56, 1006-1015, 2005
- [6] Vojtek M., Kočenda E. „Credit Scoring Methods”, Finance a Uver - Czech Journal of Economics and Finance, Volume 56, Issue 3-4, 152-167, 2006
- [7] West D., Neural network credit scoring models
- [8] Witten I.H., Eibe F. „Data mining. Practical machine learning tools and techniques”, Morgan Kaufmann, 2005

Dziękuję za uwagę!