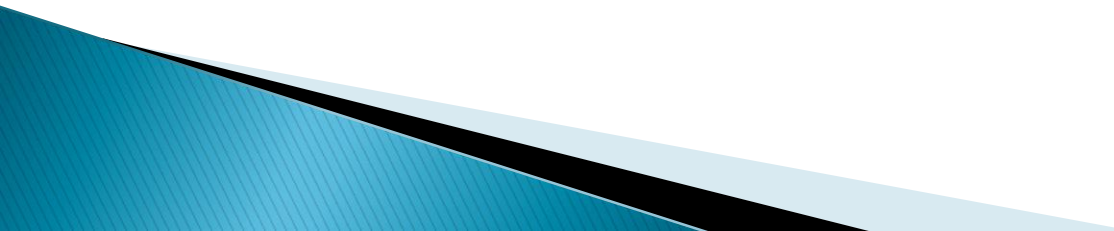


Credit Scoring & Logistic regression

Alžbeta Švaňová

Seminár z finančnej matematiky

Obsah prezentácie

- ▶ Kredit skóring
 - ▶ Skóre karta
 - ▶ Logistická regresia
 - ▶ Logitová transformácia
 - ▶ Odhad parametrov logistickej regresie
- 

Kredit skóring

- ▶ podstata kredit skóringu je na základe odhadu pravdepodobnosti zlyhania ohodnotiť úverovú spôsobilosť (bonitu) svojich potenciálnych klientov
- ▶ pomáha veriteľom pri rozhodovacom procese medzi poskytnutím a zamietnutím žiadosti o úver (dobrý/zlý žiadateľ)
- ▶ využíva sa pri stanovení úrokovej sadzby na základe odhadnutého rizika, prípadne úverového stropu (limitu)

Delenie kredit skóringu

- ▶ aplikačný – rozhoduje o prijatí / neprijatí nového klienta s využitím socio-demografických charakteristík žiadateľa, prípadne z informácií o danom klientovi z úverového registra
- ▶ behaviorálny – u existujúcich klientov; sleduje správanie sa individuálnych účtov, ich vývoja, využíva sa na správu účtu, stanovenie limitov, schvaľuje navýšenie limitu, ...

Zber informácií

- ▶ fáza prípravy dát je najnáročnejšia časť
- ▶ dáta prechádzajú čistením, prípadne modeláciou chýbajúcich údajov
- ▶ zlúčenie dát do 1 tabuľky
- ▶ napr.: priebeh skóringového procesu spoločnosti AdastraGroup :



Skóre karta

- ▶ reprezentácia kredit skóringového modelu
- ▶ pozostáva zo skupiny vybraných otázok a rôzne odpovede k nim
- ▶ každej odpovedi je pridelený určitý počet bodov
- ▶ súčet bodov určuje klientove kreditné skóre
- ▶ ak je skóre vyššie ako bankou stanovená prahová hodnota, tzv. cut-off, úver bude klientovi poskytnutý

Príklad skóre karty

Variable	Points
Age	
18...37	0
37...57	7
57...76	25
Cards	
None	0
paycard	150
credit card	102
Children	
One	60
None	0
no information	30
Two and more	98
Credit History	
None	182
positive	356
average	155
with negative factors	0

- ▶ výsledok súčtu bodov = klientove kreditné skóre

Výhody kredit skóringu

- ▶ zlepšenie úrovne portfólia – zamietnutí s nízkym skóre zlepšujú úroveň portfólia
- ▶ vytvára lepšie podmienky pre dobrých dlžníkov – žiadateľom s vysokým skóre je možné znížiť úrokovú sadzbu a naopak, čím sa v podstate poisťuje riziko

Výhody kredit skóringu

- ▶ zníženie nákladov – najmä na manuálnych procesoch; model posudzovania sa stále zlepšuje
- ▶ rozhodnutia bez osobných a subjektívnych názorov – výsledok kredit skóringového modelu je jednoduchšie obhájitelný

Logistická regresia

- ▶ najčastejšie využívaná technika pri tvorbe skóringových modelov
- ▶ metóda zameraná na odhad pravdepodobnosti výskytu udalosti na základe pozorovaných údajov
- ▶ uplatnenie – bankovníctvo, poisťovníctvo, telekomunikácie,
- ▶ v oblasti riadenia rizika pomáha spoločnosti predikovať splatenie/nesplatenie poskytnutého úveru

Ciele logistickej regresie

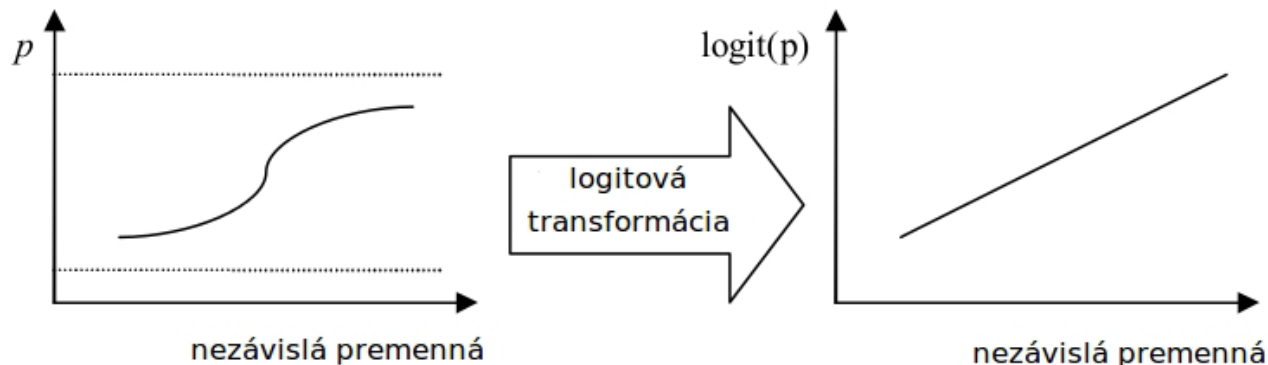
- ▶ identifikácia premenných X , ktoré významne ovplyvňujú príslušnosť objektu ku skupine
- ▶ predikcia príslušnosti objektu do skupiny na základe jeho hodnôt premenných X

Logitová transformácia

- ▶ vzťah medzi pravdepodobnosťou p a vysvetľujúcou premennou X nie je lineárny, preto transformácia; vyjadríme log podielu, tzv. logit

$$\text{logit}(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

- ▶ i indexy všetkých udalostí (pozorovaní)
- ▶ p_i pravdepodobnosť udalosti



Spätaná logitová transformácia

- ▶ spätanou transformáciou vzniká exponenciálna funkcia

$$p = \frac{e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)}}{1 + e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)}} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)}}$$

- ▶ logistické regresné koeficienty $\beta_0, \beta_1, \dots, \beta_k$

- ▶ ak $\beta_i > 0$ tak vyššie hodnoty premennej X_i

pri rovnakých hodnotách ostatných premenných X zvyšujú šancu zaradenia objektu do skupiny

Binárna logistická regresia

- ▶ závislá premenná je v binárnom tvare a jej 2 jediné možné hodnoty sú
 - splatenie úveru y_1
 - nesplatenie úveru y_2

$$p = P(Y = y_1) = 1 - P(Y = y_2)$$

- ▶ podiel $\frac{p}{1-p}$ vyjadruje pravdepodobnosť výskytu javu p ku pravdepodobnosti, že jav p nenastane, t.j. pomer šancí *odds*

Výsledky binárneho logistického modelu

- ▶ výpočet pomeru šancí *odds* $\frac{p}{1-p}$
- ▶ predpoklad: jednorozmerný model
- ▶ jediná vysvetľujúca premenná x , kt. je binárna

$$y_i = \beta_0 + \beta_1 x_i$$

- ▶ logit $\log it(p) = \ln\left(\frac{p}{1-p}\right) = \ln\left(\frac{P(Y = y_1)}{P(Y = y_2)}\right) = \beta_0 + \beta_1 x$

▶ pre $x=0$ šanca $odds(0) = \frac{P(y = 1 | x = 0)}{P(y = 0 | x = 0)} = \frac{e^{\beta_0}}{1 + e^{\beta_0}} = e^{\beta_0}$

▶ pre $x=1$ šanca $odds(1) = \frac{P(y = 1 | x = 1)}{P(y = 0 | x = 1)} = \frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}} = e^{\beta_0 + \beta_1}$

▶ logaritmus šancí

$$\ln(odds) = \ln\left(\frac{odds(1)}{odds(0)}\right) = \ln\left(\frac{e^{\beta_0 + \beta_1}}{e^{\beta_0}}\right) = \ln(e^{\beta_1}) = \beta_1$$

▶ hodnota pomeru šancí – odlogaritmovanie

▶ odds ratios = e^{β_1}

Odhad parametrov logistickej regresie

- ▶ predpoklad: pozorovania náh. veličiny $y_i, i = 1, 2, \dots, n$ ktorá vystupuje v logistickej regresii ako vysvetľujúca premenná sú nezávislé
- ▶ parametre modelu $\beta_0, \beta_1, \dots, \beta_k$ udávajú váhy jednotlivých vysvetľujúcich premenných $x_i, i = 1, 2, \dots, n$
- ▶ úlohou je teda nájsť čo najlepší odhad týchto parametrov, teda vektoru β

Metóda maximálnej vierohodnosti

- ▶ konštrukcia vierohodnostnej funkcie

$$P(y_1 = m) = p^m \cdot (1 - p)^{1-m}, m = 0, 1$$

- ▶ funkcia parametru β , ktorý je odhadovaný

$$l(\beta) = \prod_{i=1}^n p^{y_i} \cdot (1 - p)^{1-y_i}$$

- ▶ získanie takej hodnoty parametru, ktorý maximalizuje $l(\beta)$
- ▶ v praxi maximalizácia funkcie $\ln[l(\beta)]$

- dostávame

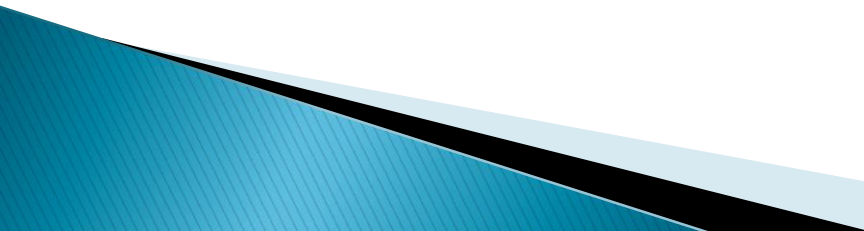
$$L(\beta) = \ln[l(\beta)] = \sum_{i=1}^n (y_i \ln p + (1 - y_i) \ln[1 - p])$$

- pri hľadaní vektoru β , ktorý maximalizuje $L(\beta)$ spočítame parciálne derivácie funkcie $L(\beta)$ podľa jednotlivých parametrov $\beta_0, \beta_1, \dots, \beta_k$ a položíme rovno 0
- dostaneme vierohodnostné rovnice a z nich nelineárnu sústavu rovníc
- riešením je potom odhad vektoru β
- odhad smerodatnej odchyľky odhadov parametrov pomocou matice druhých derivácií funkcie $L(\beta)$

System SAS

- ▶ procedúra LOGISTIC
- ▶ linková funkcia logit prípadne probit

Výhody logistickej regresie

- ▶ málo parametrov
 - ▶ jednoduché použitie a interpretácia
 - ▶ možné jednoduché začlenenie disktrétnych prediktorov
 - ▶ funguje dobre na dátach, ktoré sa značne líšia od Gaussovských zmesí
 - ▶ väčšinou dobre funguje pokiaľ venujeme dostatočnú pozornosť príprave dát
 - ▶ z praxe: v 4 prípadoch z 5 je log.regresia na dátach, ktoré analyzuje, buď najlepšia alebo rovnako dobrá ako iné metódy
- 

Literatúra

- ▶ THOMAS, L. C. a David B. EDELMAN a Jonathan N. CROOK. *Credit scoring and its applications*. Philadelphia, Pa.: Society for Industrial and Applied Mathematics, 2002. xiv, 248 p. ISBN 0-89871-483-4.
- ▶ ŘEZÁČ , M.. Data Mining 1 [online], Brno: Masaryk University, 2013.
- ▶ Diplomová práce, *Skóringové modely hodnotenia úverovej spôsobilosti*, Ekonomicko-správí fakulta MUNI
- ▶ <http://www.plug-n-score.com/learning/what-is-scoring-and-why-you-need-it-part2.htm>
- ▶ http://www.adastra.cz/784_kreditni-skoring.aspx

Ďakujem za pozornosť