

Skóringové a klasifikační metody v bankovníctví

Petr Marhoun

Obhajoba diplomové práce

Motivace

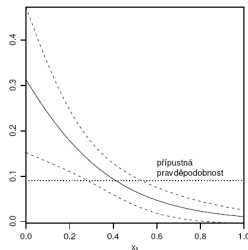
- Poskytnutí úvěru
- Analýza a řízení portfolia
- Kapitálová přiměřenost a interní ratingy - **Basel II**

Data

- Pozorování - veličina \mathbf{X} , bod v p -dimenzionálním prostoru
- Třída - veličina Y , hodnoty 0 (splácený úvěr) a 1 (default)
- Trénovací množina - pro konstrukci modelu
- Validační množina - pro zhodnocení kvality modelu

Problémy

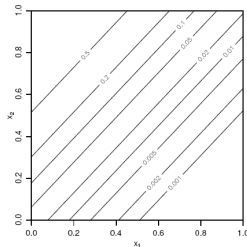
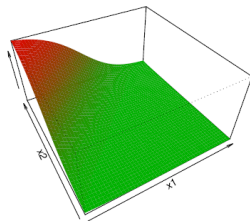
- Podmíněná pravděpodobnost defaultu - $p(\mathbf{x})$, skóre - $\mathbf{w}^T \mathbf{x}$
- Klasifikace - $p(\mathbf{x}) > \frac{c_0}{c_0+c_1}$ (c_1 - ztráta z nesprávného zařazení pozorování z třídy 1)
- Ratingové modely - pozorování je přiřazen rating



Lineární diskriminační analýza

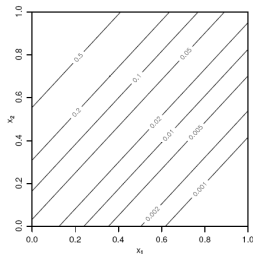
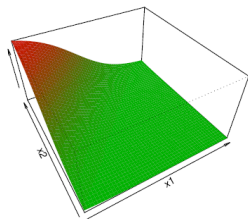
Předpoklady

- Splácené úvěry a defaulty - výběry z dvou různých normálních rozdělení
- Společná varianční matice
- Známé apriorní pravděpodobnosti
- Normalita není zcela nutná
- Lineární metoda - skóre $\mathbf{w}^T \mathbf{x}$



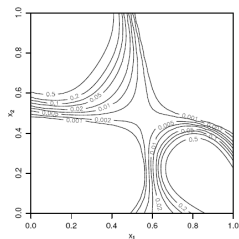
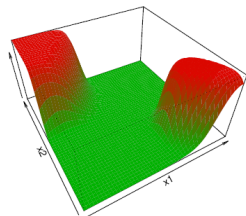
Předpoklady

- Skóre ztransformováno logistickou funkcí - $p(\mathbf{x}) = \frac{\exp(\mathbf{w}^T \mathbf{x})}{1 + \exp(\mathbf{w}^T \mathbf{x})}$
- Logaritmus poměru pravděpodobností lineární - $\log \frac{p(\mathbf{x})}{1-p(\mathbf{x})} = \mathbf{w}^T \mathbf{x}$
- Splněno rozděleními exponenciálního typu s disperzním parametrem nezávislým na třídě



Předpoklady

- Tři vrstvy
- $p(\mathbf{x}) = f_2(\sum_{k=1}^K w_{h_k}^{out} f_1(\sum_{j=1}^P w_{in_j}^{h_k} x_j))$
- f_1 a f_2 - logistické funkce
- Síť s konečným počtem skrytých uzlů stejnoměrně aproximují libovolnou spojitou funkci
- Nelineární metoda



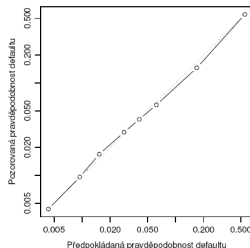
Tvorba ratingového modelu

Tvorba modelu

- Na základě trénovací množiny
- Ratingová funkce - zobrazuje množinu všech možných pozorování do množiny ratingů ($\{1, 2, \dots, R\}$)
- Určeny předpokládané pravděpodobnosti defaultu

Kalibrace

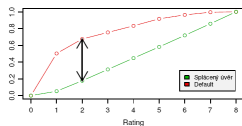
- Na základě validační množiny
- Test shod pravděpodobností defaultu



Diskriminační schopnost

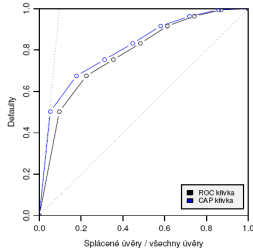
Křivky

- ROC křivka - splácené úvěry a defaulty
- CAP křivka - všechny úvěry a defaulty



Statistiky

- AUC - založena na ROC křivce
- AR - založena na CAP křivce
- $AR = 2 AUC - 1$
- Kolmogorov-Smirnovův test - měří maximální vzdálenost

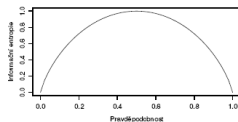


Informační entropie

- Kolik informace třeba k plné znalosti
- Pro úvěry nepřesná

Simulace

100	1000	10000	Skutečnost
0.2973	0.3398	0.3443	0.3449

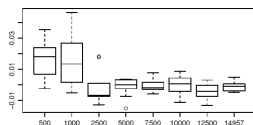


Implementace

- Výpočet statistickým jazykem R
- Výsledky přes webové rozhraní
- <http://diplomka.pm13.org/>

Simulace

- Na základě vygenerovaného portfolia
- 320 modelů, 960 validací, 3200 predikcí, desítky tisíc obrázků
- Podle AUC hodnoceno množství dat



Přínosy

- Podrobné a přesto relativně jednoduché zpracování daného problému
- Výsledky statisticky podložené
- Postupy prakticky vyzkoušeny
- Odhalena nepřesnost informační entropie

Možné otázky

- Velikost datových souborů
- Nepoužití validační množiny

Regularizovaná diskriminační analýza

- “Lineární kombinace” lineární a kvadratické diskriminační analýzy
- $\hat{\Sigma}_l(\alpha) = \alpha \hat{\Sigma} + (1 - \alpha) \hat{\Sigma}_l, \quad \alpha \in [0, 1], \quad l = 0, 1$

Konvergence algoritmu pro logistickou regresi

- Logistická regrese - zobecněný lineární model s kanonickým linkem \Rightarrow konkávní účelová funkce
- Newtonův algoritmus
- Algoritmus by měl konvergovat

Přínosy

- Podrobné a přesto relativně jednoduché zpracování daného problému
- Výsledky statisticky podložené
- Postupy prakticky vyzkoušeny
- Odhalena nepřesnost informační entropie

Možné otázky

- Velikost datových souborů
- Nepoužití validační množiny