Wprowadzenie

- Termin regresja trochę mylący
  - Zmienna wyjściowa nie jest liczbą / nie stosuje się typowego modelowania regresji
- Zadanie dotyczy klasyfikacji binarnej
  - Poszukujemy zależności między binarną zmienną y {0,1} a zmiennymi objaśniającymi x_1,x_2,…x_p.
- Często występuje w zastosowaniach
  - Medycyna – badanie wpływu różnych atrybutów diagnostycznych pacjentów na występowanie choroby *y*
  - Finanse – wpływ atrybutów opisujących charakterystykę klienta na ocenę prawdopodobieństwo spłaty kredytu
  - Predykcja odpowiedzi klienta na akcje marketingową
- Zamiast samej wartości y bardziej oceniamy p(y|x) prawdopodobieństwo zajścia sytuacji zerojedynkowej

Inne spojrzenie na prawdopodobieństwo

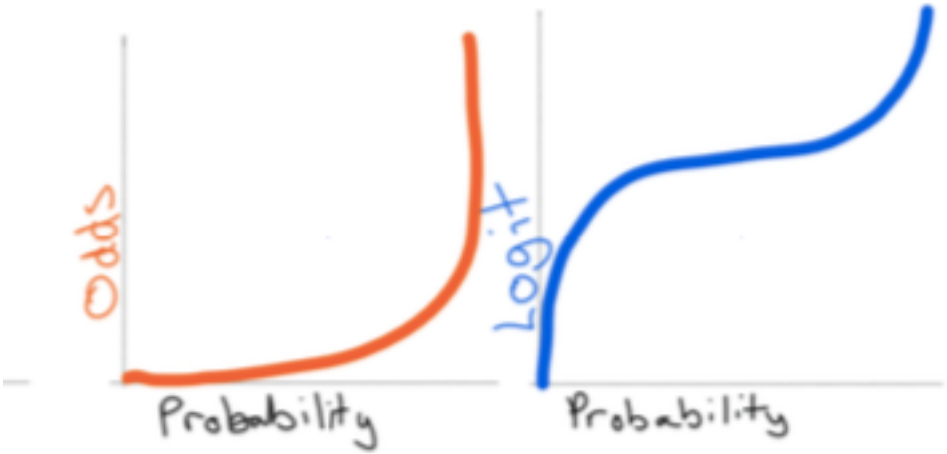
- **Iloraz szans (ang. Odds) na podstawie p**
  - Zakładamy ze p dotyczy sytuacji zerojedynkowej O1 i O2 – z rozkładu dwumianowego $B(p)$
- **Definicja**
  - Sytuacja $p(\mathrm{O1}) = p$
  - Sytuacja $p(\mathrm{O2}) = 1 - p = q$

Odds

$$Odds = \frac{p}{1-p} = \frac{p}{q}$$

Dalsze przekształcenia logit

	notation	range equivalents		
standard probability	p	0	0.5	1
odds	p / q	0	1	+ ∞
log odds (logit)	log(p / q)	- ∞	0	+ ∞



Postać modelu regresji logistycznej

Prawdopodobieństwo warunkowe przyjęcia przez *y* wartości 1, pod warunkiem, że zmienne ***x*** przyjęły wartości $(x_1,x_2,..,x_p)$

$$p = P(Y = 1 \mid X = x) = \frac{e^z}{1+e^z} = \frac{1}{1+e^{-z}}$$

Gdzie $z = \mathbf{x\beta} = x_1\beta_1+x_2\beta_2+...+x_p\beta_p$

Przykład ilustracyjny

Pacjenci − zależność pomiędzy zmienną wiek pacjenta (x) a obecnością (1) lub brakiem choroby y [D.Larose book]

TABLE 4.1 Age of 20 Patients, with Indicator of Disease

Patient ID	Age, x	Disease, y	Patient, ID	Age, x	Disease, y
1	25	0	**11**	50	0
2	29	0	**12**	59	1
3	30	0	**13**	60	0
4	31	0	**14**	62	0
5	32	0	**15**	68	1
6	41	0	**16**	72	0
7	41	0	**17**	79	1
8	42	0	**18**	80	0
9	44	1	**19**	81	1
10	49	1	**20**	84	1

Przykład - wykres

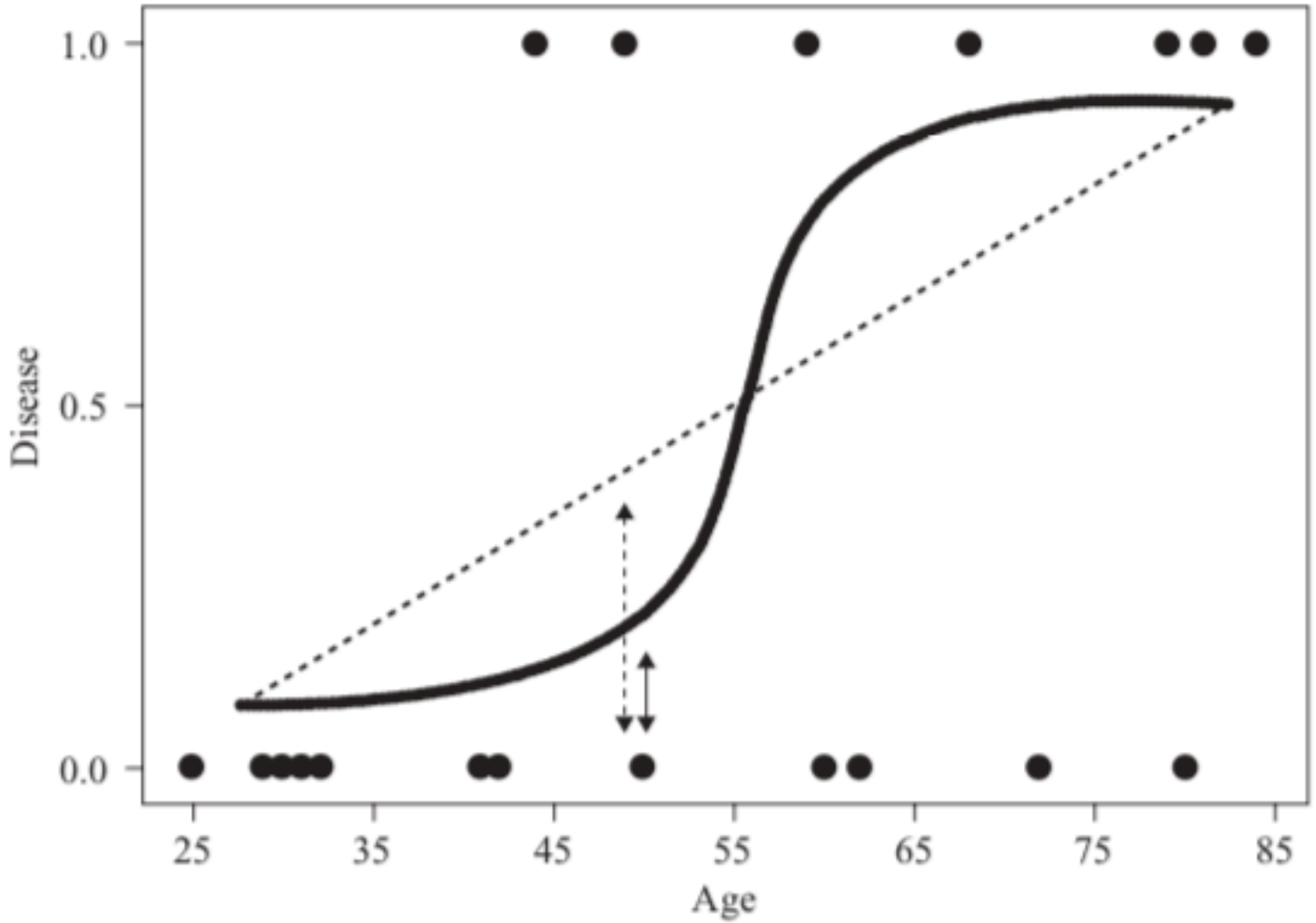


Figure 4.1 Plot of *disease* versus age, with least-squares and logistic regression lines.

Przekształcenie regresji logistycznej

Równanie na prawd p w modelu logistycznym można przekształcić logarytmując *odds*

$$\text{logit}(p) = \ln \frac{p}{1-p} = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p$$

oraz

$$\frac{p}{1-p} = e^{\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p}$$

Interpretacja współczynników

- Interpretacji podlega iloraz szans odds

$$\frac{p}{1-p} = e^{\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p}$$

- Jeżeli zwiększymy xj o jednostkę, to przy ustalonych wartościach innych zmiennych, szansa jest zwiększana poprzez iloczyn z wartością $e^{\beta j}$
- a logit-u szansy o βj
- Jeżeli $e^{\beta j} > 1$, to zmienna x_j działa stymulująco na możliwość wystąpienia badanego zjawiska, w przeciwnym razie działa ograniczająco (jeżeli $e^{\beta j} = 1$, to zmienna x_j nie ma wpływu na badane zjawisko).

Estymacja paramterów

Konieczność oszacowania współrzędnych wektora $\boldsymbol{\beta}=(\beta_1,\beta_2,\ldots,\beta p)^T$ na podstawie danych uczących $(\mathbf{x},y)_{n\times p}$

- W odróżnieniu od regresji liniowej – nie można otrzymać analitycznej postaci dla rozwiązania optymalnej wartości β
- Stosuje się metodę estymacji największej wiarygodności (ang. maximum likelihood estimation)
- Dla dyskretnego rozkładu y funkcja wiarygodności jest iloczynem prawdopodobieństw pojawienia się poszczególnych obserwacji z próby przy danych parametrach modelu

$$P(Y = y_i \mid x = x_i) = \left(p(x_i)\right)^{y_i} \cdot \left(1 - p(x_i)\right)^{1-y_i}$$

Estymacja paramterów

Zakładając, że przykłady uczące (obserwacje) są niezależne, łączna funkcja wiarygodności jest

$$L(\beta) = \prod_{i=1}^{n} \left(p(x_i) \right)^{y_i} \cdot \left(1 - p(x_i) \right)^{1-y_i}$$

Należy wybrać β maksymalizujące $L(\beta)$

Dogodniejsza obl. jest maksymalizacja logarytmu $L(\beta)$, tj.

$$= \sum_{i=1}^{n} y_i \log p(x_i) + (1 - y_i) \log \left(1 - p(x_i)\right).$$

Który można przekształcić $$= \sum_{i=1}^{n} y_i (x_i^T \beta) - \log \left(1 + \exp(x_i^T \beta)\right)$$

Rozwiązuje się numeryczne algorytmem Newtona-Raphsona

Wnioskowanie w modelu logistycznym

- Można wykorzystać prawdopodobieństwa jako przynależności do klas
- Oszacować granicę decyzyjną
- Dla logistycznej funkcji prawdopodobieństwa p(y|x)
- Progowanie decyzji: $$\hat{f}(x) = \begin{cases} 0 & \hat{p}(x) \leq 0.5 \\ 1 & \hat{p}(x) > 0.5 \end{cases}$$
- Lub korzystając z przekształcenia logit $$\operatorname{logit} \hat{p}(x) = \hat{\beta}^T x,$$
- Granica decyzyjna

$$\hat{f}(x) = \begin{cases} 0 & \hat{\beta}^T x \leq 0 \\ 1 & \hat{\beta}^T x > 0 \end{cases}$$

Przykład ilustracyjny

- Powróćmy do przykładu oceny prawdopodobieństwa choroby w zależności od wieku
- Estymatory największej wiarygodności β
  b0=-4.372 oraz b1=0.06696

Rozważmy dane pacjent w wieku 50 lat

Prawdopodobieństwo choroby y=1 równa sie

$$\frac{e^{-4.372+0.06696(age)}}{1+e^{-4.372+0.06696(age)}}$$

Z przekształcenia logit $-4.372 + 0.06696(50) = -1.024$

Prawdopodobieństwo, że pacjent choruje $: \frac{e^{-1.024}}{1+e^{-1.024}} = 0.26$
i że nie ma choroby 1-0.36 = 0.74

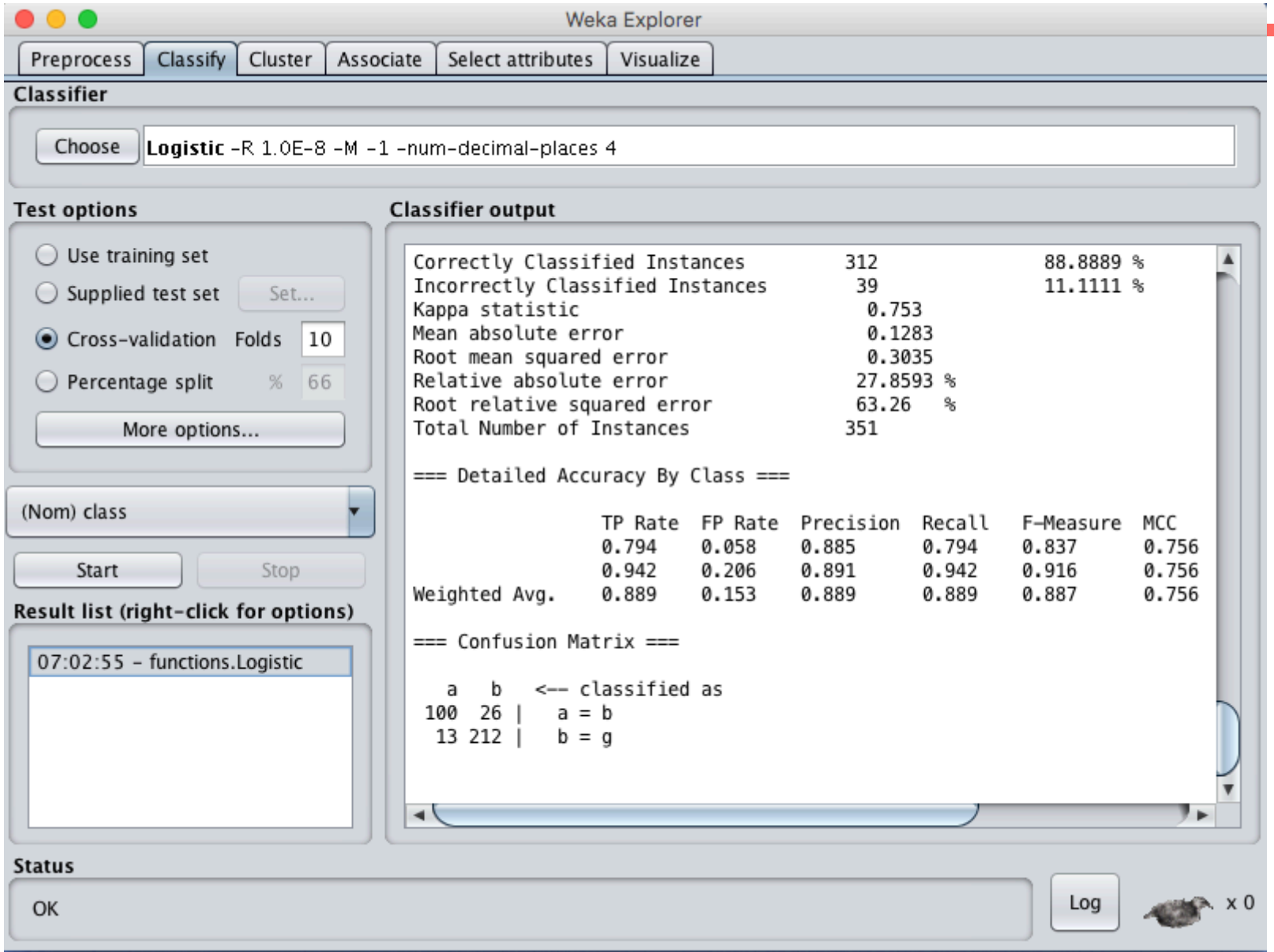
Analogiczne obl. dla wieku 72 / prawd choroby 0.61, nie ma 0.39

Przykład cd

Interpretacja szans (odds)

- Szansa na zachorowanir dla72 letniego pacjenta odds = 0.61/0.39 = 1.56
- Szansa choroby dla 50 letniego pacjenta 0.26/0.74=0.35

Weka - prościej



Literatura

- D.Larose: Metody i modele eksploracji danych. (polskie tłumaczenie PWN 2008)
- T.Górecki, M. Krzyśko, M. Skorzybut, W. Wołyński: Systemy uczące się. WNT 2009
- Agresti, A.: Foundations of Linear and Generalized Linear Models. Wiley Probability and Statistics Series (2015).
- J.Horbert: Logistic regression − wykład ML
- Inne wykłady online

Podsumowanie

Zalety:

- Prostota i łatwość użycia
- Interpretowalność (zwłaszcza wpływu zmiennych)
- Efektywność obliczeniowa
- Możliwe uogólnienia (wiele klas; różne typy zmiennych)
- Dobre podstawy prob. i statystyczne
- Odporna na przeuczenie

Ograniczenia:

- Liniowa granica decyzyjna (w przekształconej przestrzeni)

Matematyka za obl. p_i

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 X$$

$$\frac{p_i}{1-p_i} = \exp(\beta_0 + \beta_1 X)$$

$$p_i = (1-p_i)\exp(\beta_0 + \beta_1 X)$$

$$p_i = \exp(\beta_0 + \beta_1 X) - p_i \exp(\beta_0 + \beta_1 X)$$

$$p_i + p_i \exp(\beta_0 + \beta_1 X) = \exp(\beta_0 + \beta_1 X)$$

$$p_i(1 + \exp(\beta_0 + \beta_1 X)) = \exp(\beta_0 + \beta_1 X)$$

$$p_i = \frac{\exp(\beta_0 + \beta_1 X)}{1 + \exp(\beta_0 + \beta_1 X)}$$

Koniec wykładu – dziękuję za uwagę

Pytania lub komentarze?

Zapraszam na „konsultacje”

Czytaj książki, artykuły, …

Kontakt:
Jerzy.Stefanowski@cs.put.poznan.pl
http://www.cs.put.poznan.pl/jstefanowski