Master studije
GEOINFORMATIKA

Mašinsko učenje - uopšteni linearni modeli, logistička regresija, Njutnova metoda.

Tatjana Jakšić Krüger

tatjana@turing.mi.sanu.ac.rs

23. novembar 2021.

Građevinski fakultet Univerziteta u Beogradu

Podsetnik

- Predviđanja se mogu vršiti linearnim modelom regresije.
- Algoritam učenja je optimizaciona metoda (metoda najstrmijeg spusta).
- Pod kojim uslovima možemo i ne možemo direktno da odredimo koeficijenate linearnog modela $\boldsymbol{\theta}$.
- Preprilagođavanje. Generalizacija. Regularizacija.

Cilj za danas

- Uopšteni linearni modeli.
- Logistička regresija.
- Njutnova metoda.

Kada koristiti linearnu regresiju

- Kada želimo da primenimo linearni model regresije?
  - ↪ Kada vršimo predviđanja.
- Da li bilo koje podatke možemo da opišemo linearnim modelom?
  - ↪ Ne.
- Koje uslove moraju podaci da zadovolje?
  - ↪ Zavisna promenljiva y je realna vrednost.
  - ↪ Postoji linearna zavisnost između nezavisne x i zavisne y promenljive.
  - ↪ Nema značajnih ekstremnih vrednosti (*eng.* outliers).
  - ↪ Reziduali su nezavisni i $\epsilon_i \sim \mathcal{N}(0, 1)$.
- Razlozi za ne-normalnu raspodelu rezidua linearnog modela regresije:
  1. Nezavisna ili zavisna promenljiva su isuviše ne-normalne.
  2. Postoje ekstremene vrednosti koje kvare prosek.
     Primena Box-Cox transformacije nad ulaznim podacima kako bismo dobili $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$.

Uopšteno linearno modelovanje

- *eng*. Generalized linear model (GLM).
- Kada koristimo uopšteno linearno modelovanje?
  1. Zavisna promenljiva y se opisuje raspodelom iz familije eksponencijalnih funkcija:
     - ↪ normalna, log-normalna, eksonencijalna, gama, hi-kvadratna, beta, Bernulijeava, Poasonova, itd.
     - ↪ nisu uključene Studentova t-raspodela i mešane raspodele.
  2. Transformacija srednje vrednosti zavisne promenljive linearno je povezana sa nezavisnim promenljivama.

Uopšteno linearno modelovanje

- Modelovana promenljiva $\hat{y}$ je linearna funkcija od promenljivih $x_j^{(i)}$:

$$g(\hat{y}^{(i)}) = \theta_0 + \theta_1 x_1^{(i)} + \theta_2 x_2^{(i)} + \dots$$

$\hookrightarrow$ Transformacija g se zove funkcija veze, link funkcija (*eng.* link function).

- Postoji tipično uparivanje funkcije veze i raspodele zavisne promenljive y.

Raspodela	Upotreba	Naziv	Funkcija veze
Normalna	Linearni odziv	Identička	$g(y) = y$
Eksponencijalna ili gama	Eksponencijalni odziv; parametar skaliranja	Inverz	$g(y) = \frac{1}{y}$
Poasonova	broj pojavljivanja za fiksirano vreme/prostor	Logaritam	$g(y) = \ln(y)$
Bernulijeva	prebrojavanje "da"/"ne"	Logit	$g(y) = \ln(\frac{y}{1-y})$

GLM kada je y binarna

- Pogledajmo primer za GLM kada je y binarna promenljiva.
- Primer binarne promenljive:
  - bračno stanje je "venčan" ili "nevenčan"..
  - proizvodni deo je "ispravan" ili "neispravan".
  - proizvod je "kupljen" ili "nije kupljen"...
- Tipično je kodiranje sa 1 i 0.
- Prebrojavanje kada je $y = 1$ daje Bernulijevu raspodelu verovatnoća za y.

Bernulijeva raspodela verovatnoće

- **Bernulijev događaj** - događaj čija je verovatnoća da se ostvari/realizuje/desi jednaka p.
  - ↪ Verovatnoća da se neće desiti je $1 - p$.
- **Bernulijev eksperiment** - realizacija Bernulijevog događaja
  - ↪ Ujedno je i najjednostavniji eksperiment u oblasti verovatnoće.
- Primer Bernulijevog eksperimenta:
  1. Bacanje novčića. Ako je "uspeh" da će pasti glava, a "neuspeh" da će pasti pismo, bacanje novčića je Bernulijev eksperiment.
  - ↪ Više bacanja novčića čini **Bernulijev proces** tokom kojeg prebrojavamo "uspeh".
  2. Fudbal. Ako je postizanje gola "uspeh" onda nas interesuje da prebrojimo koliko puta je postignut go.
- Bernulijeva funkcija gustine verovatnoće $f(x)$:

$$f(x) = \begin{cases} p^x(1-p)^{1-x}, & x = \{0, 1\} \\ 0, & \text{u suprotnom} \end{cases}$$

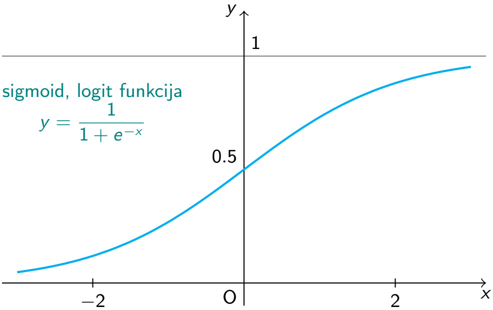
  - Matematičko očekivanje: $E[X] = \sum_{i=1}^{2} x_i p_i = p$
  - Varijansa: $Var[X] = E[X^2] - (E[X])^2 = p(1-p)$

Logit funkcija

- Verovatnoća uspeha se može proceniti putem formule:

$$\hat{y}^{(i)} = \theta_0 + \theta_1 x_1^{(i)} + \theta_2 x_2(i) + \dots$$

- Problem je što ovakvo y nije ograničeno na interval $[0, 1]$!
- Potrebna nam je logit funkcija.



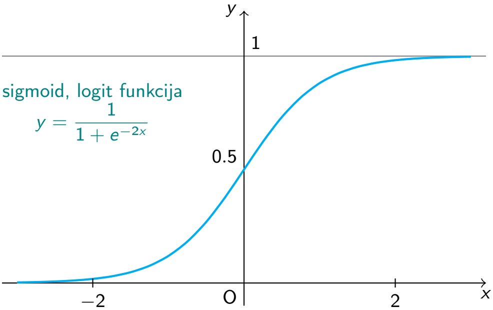
- Preslikava $[-\infty, \infty]$ u interval $[0, 1]$.
- $\lim_{x \to -\infty} \frac{1}{1+e^{-x}} = 0$
- $\lim_{x \to +\infty} \frac{1}{1+e^{-x}} = 1$

Logit funkcija

- Verovatnoća uspeha se može proceniti putem formule:

$$\hat{y}^{(i)} = \theta_0 + \theta_1 x_1^{(i)} + \theta_2 x_2(i) + \dots$$

- Problem je što ovakvo y nije ograničeno na interval $[0, 1]$!
- Potrebna nam je logit funkcija.

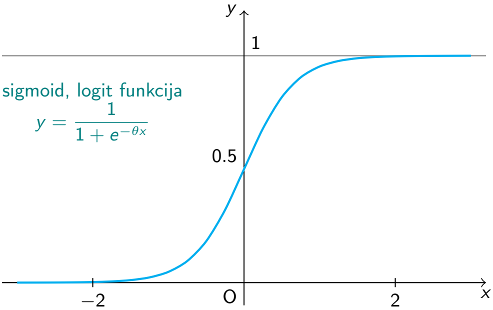


- Preslikava $[-\infty, \infty]$ u interval $[0, 1]$.
- $\lim_{x \to -\infty} \frac{1}{1+e^{-2x}} = 0$
- $\lim_{x \to +\infty} \frac{1}{1+e^{-2x}} = 1$

Logit funkcija

- Verovatnoća uspeha se sada može proceniti putem formule:

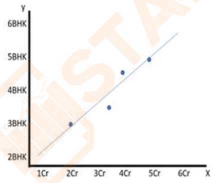
$$\mathrm{logit}(p) = \ln(\frac{p}{1-p}) = \theta_0 + \theta_1 x_1^{(i)} + \theta_2 x_2(i) + \ldots$$

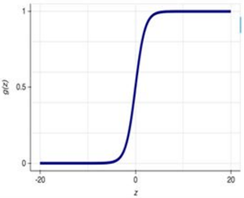


- Preslikava $[-\infty, \infty]$ u interval $[0, 1]$.
- $\lim_{x \to -\infty} \frac{1}{1+e^{-\theta x}} = 0$
- $\lim_{x \to +\infty} \frac{1}{1+e^{-\theta x}} = 1$

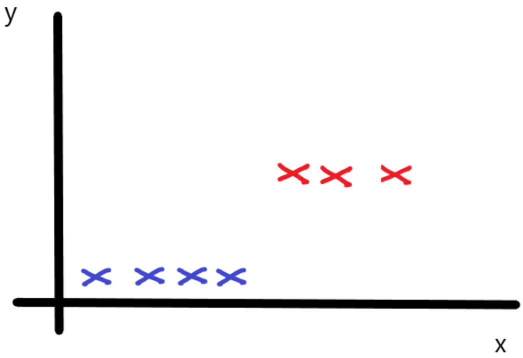
Poredjenje

Linear Regression	Logistic Regression
Target is an interval variable	Target is discrete (binary or ordinal) variable
Predicted values are the mean of the target variable at the given values of the input variable	Predicted values are the probability of the particular levels of the given values of the input variable
Solve regression problems	Solve classification problems
Example : What is the Temperature?	Example : Will it rain or not?
Graph is straight line	Graph is S-curve

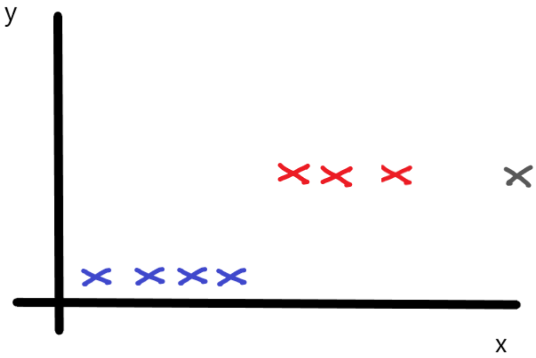




- Problem binarne klasifikacije - kada je izlazna vrednost 0 ili 1 tj. $y = \{0, 1\}$.



- Problem binarne klasifikacije - kada je izlazna vrednost 0 ili 1 tj. $y = \{0, 1\}$.



Hipoteza problema binarne klasifikacije

- Oblik hipoteze (modela) je:

$$h(\theta) = g(\theta^T x) = \frac{1}{1 + e^{-\theta^T x}}$$

- Na osnovu linearnog modela imamo:

$$h_\theta(x^{(i)}) = \theta^T x^{(i)}.$$

- Pretpostavimo da možemo opisati podatke sledećom raspodelom:

$$P(y = 1|x; \theta) = h_\theta(x).$$

- Posledica osobine verovatnoca:

$$P(y = 0|x; \theta) = 1 - h_\theta(x).$$

- Jedna jednacina:

$$P(y|x; \theta) = h_\theta(x)^y (1 - h_\theta(x))^{1-y}$$

Funkcija verodostojnosti

$$\mathcal{L}(\theta) = P(y|x;\theta) = \prod_{i=1}^{m} p(y^{(i)}|x^{(i)};\theta)$$

Funkcija verodostojnosti

$$\ell(\theta) = \log \mathcal{L}(\theta)$$

$$= \sum_{i=1}^{m} \left(y^{(i)} \log h_\theta(x^{(i)}) + (1-y^{(i)}) \log(1 - h_\theta(x^{(i)}) \right)$$

Gradijentni metod:

$$\theta_k := \theta_k + \lambda \frac{\partial}{\partial \theta_k}$$

Funkcija verodostojnosti

$$\mathcal{L}(\theta) = \prod_{i=1}^{m} p(y^{(i)} | x^{(i)}; \theta)$$

$$= \prod_{i=1}^{m} h_\theta(x^{(i)})^{y^{(i)}} \cdot \left(1 - h_\theta(x^{(i)})\right)^{1-y^{(i)}}$$

$$P(y|x;\theta) = h_\theta(x)^{y} \left(1 - h_\theta(x)\right)^{1-y}$$

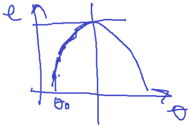
$$h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$$

$$\log(a \cdot b) = \log a + \log b$$

$$\log a^2 = 2 \log a$$

$$\ell(\theta) = \log \mathcal{L}(\theta)$$

$$= \sum_{i=1}^{m} \left(y^{(i)} \log h_\theta(x^{(i)}) + (1 - y^{(i)}) \log\left(1 - h_\theta(x^{(i)})\right) \right)$$



$$\max_\theta \ell(\theta)$$

GRADIJENTNA METODA

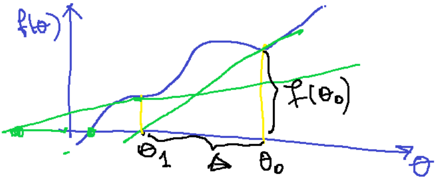
$$\theta_{k+1} = \theta_k + \lambda \frac{\partial \ell(\theta)}{\partial \theta_k}$$

$$\theta_{k+1} = \theta_k + \lambda \sum_{i=1}^{m} \left(y^{(i)} - h_\theta(x^{(i)}) \right) \cdot x_k^{(i)}$$

$\max_{\theta} \ell(\theta)$, $\theta_k \dots\dots$ $|\theta_{k+1} - \theta_k| < \varepsilon$ $\lambda = 0.01$

NJUTNOVA METODA

$f(\theta) = 0$



$$\ell'(\theta_0) = \frac{f(\theta_0)}{\Delta}$$

$$H_{ij} = \frac{\partial^2 \ell}{\partial \theta_i \partial \theta_j}$$

$$\theta_1 = \theta_0 - \frac{f(\theta_0)}{\Delta}$$

$$\theta_{k+1} = \theta_k - \frac{f(\theta_k)}{f'(\theta_k)}$$

$$\ell'(\theta) = f(\theta)$$

$$\Rightarrow \theta_{k+1} = \theta_k - \frac{\ell'(\theta_k)}{\ell''(\theta_k)}$$

, $\theta = [\theta_0 \ \theta_1 \ \dots \theta_n]^T$

$$\theta_{k+1} = \theta_k + H^{-1} \nabla_{\theta} \ell$$

H – Hesian matrica

Imamo 31 poruku. Klasifikujemo e-poruke {"spam", "nisu spam"}

1 0

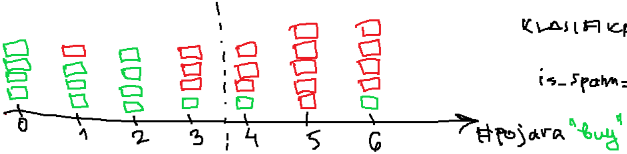
ATRIBUTI

1 "buy"

2 spelling speling ↵

nisu spam

spam



KLASIFIKATOR ("buy")

$$\text{is\_spam} = \begin{cases} T, & \#(\text{"buy"}) \geq 4 \\ F, & < 4 \end{cases}$$

□. #("buy") < 4 ⇒ □

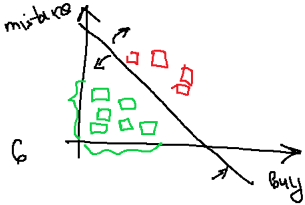
("buy") ≥ 4 ⇒ □

Funkcija verodostojnosti

Poruka	BUY	MISTAKE	BUY + MISTAKE
1	4	3	7
2	1	2	3

KLASIFIKATOR >4 >2

$$\underbrace{y = is\_spam}_{} = \begin{cases} T, & BUY + MISTAKE \geq 6 \\ F, & \text{u suprotnom.} \end{cases}$$



Hvala na pažnji.

Sada idemo na prikaz komandi.

Da li imate pitanja?