Probabilistički modeli

Mašinsko učenje 2020/21

Matematički fakultet
Univerzitet u Beogradu

Probabilistički modeli

- Dizajn algoritama probabilističkih modela mašinskog učenja se zasniva na definisanju raspodele verovatnoće koju je potrebno oceniti iz podataka.

Probabilistički modeli

- Dizajn algoritama probabilističkih modela mašinskog učenja se zasniva na definisanju raspodele verovatnoće koju je potrebno oceniti iz podataka.
- Kako bi ocena, a kasnije i zaključivanje bili računski izvodljivi, obično se uvode neke pretpostavke vezane za:

Probabilistički modeli

- Dizajn algoritama probabilističkih modela mašinskog učenja se zasniva na definisanju raspodele verovatnoće koju je potrebno oceniti iz podataka.
- Kako bi ocena, a kasnije i zaključivanje bili računski izvodljivi, obično se uvode neke pretpostavke vezane za:
  - kako promenljive zavise jedne od drugih (da li su vrednosti ciljnih promenjlivih na različitim instancama nezavisne ili postoji neka zavisnost)

Probabilistički modeli

- Dizajn algoritama probabilističkih modela mašinskog učenja se zasniva na definisanju raspodele verovatnoće koju je potrebno oceniti iz podataka.
- Kako bi ocena, a kasnije i zaključivanje bili računski izvodljivi, obično se uvode neke pretpostavke vezane za:
  - kako promenljive zavise jedne od drugih (da li su vrednosti ciljnih promenjlivih na različitim instancama nezavisne ili postoji neka zavisnost)
  - kakva je forma raspodele (npr. normalna)

Probabilistički modeli

- Dizajn algoritama probabilističkih modela mašinskog učenja se zasniva na definisanju raspodele verovatnoće koju je potrebno oceniti iz podataka.
- Kako bi ocena, a kasnije i zaključivanje bili računski izvodljivi, obično se uvode neke pretpostavke vezane za:
  - kako promenljive zavise jedne od drugih (da li su vrednosti ciljnih promenjlivih na različitim instancama nezavisne ili postoji neka zavisnost)
  - kakva je forma raspodele (npr. normalna)
- U nastavku pretpostavljaćemo međusobnu nezavisnost vrednosti ciljnih promenljivih

Pregled

Linearna regresija

Logistička regresija

Multinomijalna logistička regresija

Linearna regresija

- Osnovna pretpostavka je pretpostavka *normalne raspodele* ciljne promenljive y pri datim vrednostima atributa x

$$p(y|x) = \mathcal{N}(f(x), \sigma^2)$$

Linearna regresija

- Osnovna pretpostavka je pretpostavka *normalne raspodele* ciljne promenljive y pri datim vrednostima atributa x

$$p(y|x) = \mathcal{N}(f(x), \sigma^2)$$

- f je funkcija koja uspostavlja vezu između atributa i očekivanja ciljne promenljive i kod linearne regresije je linearna

$$f_w(x) = w \cdot x$$

Linearna regresija

- Osnovna pretpostavka je pretpostavka *normalne raspodele* ciljne promenljive y pri datim vrednostima atributa x

$$p(y|x) = \mathcal{N}(f(x), \sigma^2)$$

Linearna regresija

- Osnovna pretpostavka je pretpostavka *normalne raspodele* ciljne promenljive y pri datim vrednostima atributa x

$$p(y|x) = \mathcal{N}(f(x), \sigma^2)$$

- f je funkcija koja uspostavlja vezu između atributa i očekivanja ciljne promenljive i kod linearne regresije je linearna

$$f_w(x) = w \cdot x$$

Linearna regresija

- Osnovna pretpostavka je pretpostavka *normalne raspodele* ciljne promenljive y pri datim vrednostima atributa x

$$p(y|x) = \mathcal{N}(f(x), \sigma^2)$$

- f je funkcija koja uspostavlja vezu između atributa i očekivanja ciljne promenljive i kod linearne regresije je linearna

$$f_w(x) = w \cdot x$$

- najčešće se podrazumeva da je skup atributa proširen atributom koji je konstantne vrednosti 1 kako bi model sadržao slobodni koeficijent

Linearna regresija

- Osnovna pretpostavka je pretpostavka *normalne raspodele* ciljne promenljive y pri datim vrednostima atributa x

$$p(y|x) = \mathcal{N}(f(x), \sigma^2)$$

- f je funkcija koja uspostavlja vezu između atributa i očekivanja ciljne promenljive i kod linearne regresije je linearna

$$f_w(x) = w \cdot x$$

- najčešće se podrazumeva da je skup atributa proširen atributom koji je konstantne vrednosti 1 kako bi model sadržao slobodni koeficijent
- Ukupni model:

$$p(y|x) = \mathcal{N}(w \cdot x, \sigma^2)$$

Linearna regresija

- Da li je model generativni ili diskriminativni?

$$p(y|x) = \mathcal{N}(w \cdot x, \sigma^2)$$

Linearna regresija

- Da li je model generativni ili diskriminativni?

$$p(y|x) = \mathcal{N}(w \cdot x, \sigma^2)$$

- $$y^* = \text{argmax}_y p(y|x)$$

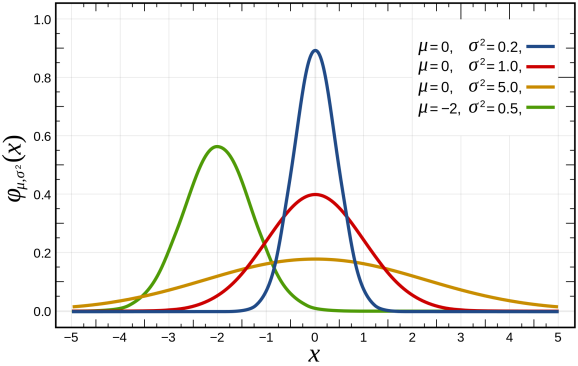
Linearna regresija

- Da li je model generativni ili diskriminativni?

$$p(y|x) = \mathcal{N}(w \cdot x, \sigma^2)$$

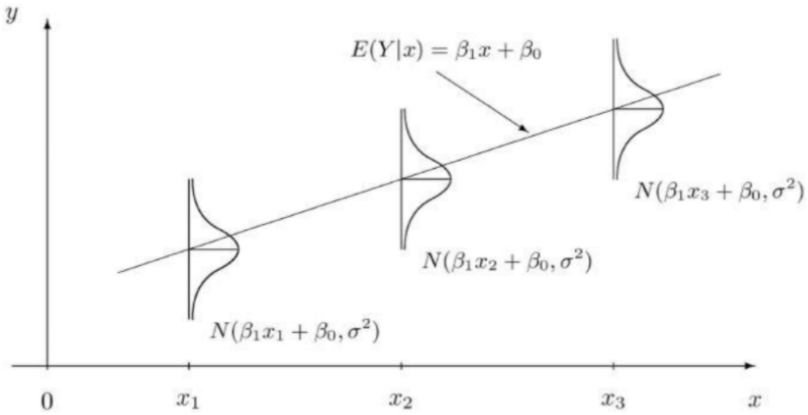
-

$$y^* = \text{argmax}_y p(y|x)$$



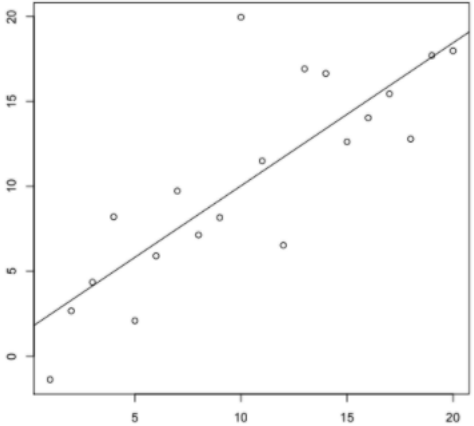
Linearna regresija

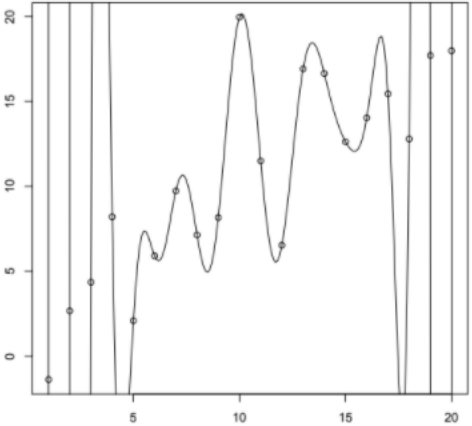
- μ se menja u zavisnosti od x a σ je konstantno



Linearna regresija

- linearnost po parametrima





Ocene parametara metodom maksimalne verodostojnosti

- funkcija verodostojnosti (eng. *likelihood*)

$$\mathcal{L}(w) = p_w(y_1, \ldots, y_n | x_1, \ldots, x_n)$$

Ocene parametara metodom maksimalne verodostojnosti

- funkcija verodostojnosti (eng. *likelihood*)

$$\mathcal{L}(w) = p_w(y_1, \ldots, y_n | x_1, \ldots, x_n)$$

- ako pretpostavimo nezavisnost instanci

$$\mathcal{L}(w) = \prod_{i=1}^{N} p_w(y_i | x_i)$$

Ocene parametara metodom maksimalne verodostojnosti

- funkcija verodostojnosti (eng. *likelihood*)

$$\mathcal{L}(w) = p_w(y_1, \ldots, y_n | x_1, \ldots, x_n)$$

- ako pretpostavimo nezavisnost instanci

$$\mathcal{L}(w) = \prod_{i=1}^{N} p_w(y_i | x_i)$$

- potrebno je rešiti problem

$$\max_w \mathcal{L}(w)$$

Ocene parametara metodom maksimalne verodostojnosti

- funkcija verodostojnosti (eng. *likelihood*)

$$\mathcal{L}(w) = p_w(y_1, \ldots, y_n | x_1, \ldots, x_n)$$

- ako pretpostavimo nezavisnost instanci

$$\mathcal{L}(w) = \prod_{i=1}^{N} p_w(y_i | x_i)$$

- potrebno je rešiti problem

$$\max_w \mathcal{L}(w)$$

- *negative log likelihood*

$$\min_w -\log \mathcal{L}(w)$$

Ocene parametara metodom maksimalne verodostojnosti

- $$\min_{w} -\log \mathcal{L}(w)$$

Ocene parametara metodom maksimalne verodostojnosti

- $$\min_{w} -\log \mathcal{L}(w)$$
- $$-\log \mathcal{L}(w) = -\sum_{i=1}^{N} \log p_w(y_i|x_i)$$

Ocene parametara metodom maksimalne verodostojnosti

- $$\min_w -\log \mathcal{L}(w)$$
- $$-\log \mathcal{L}(w) = -\sum_{i=1}^{N} \log p_w(y_i|x_i)$$
- $-\log p_w(y_i|x_i)$ se može uzeti kao funkcija greške

Ocene parametara metodom maksimalne verodostojnosti

- $$\min_w -\log \mathcal{L}(w)$$
- $$-\log \mathcal{L}(w) = -\sum_{i=1}^{N} \log p_w(y_i|x_i)$$
- $-\log p_w(y_i|x_i)$ se može uzeti kao funkcija greške
- $$p_w(y_i|x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - f_w(x_i))^2}{2\sigma^2}\right)$$

Linearna regresija

- Forma modela:

$$f_w(x) = w_0 + \sum_{i=1}^{n} w_i x_i$$

- Minimizacioni problem

$$\min_w \sum_{i=1}^{N} (y_i - f_w(x_i))^2$$

- Matrična formulacija

$$\min_w \|y - Xw\|$$

- Rešenje

$$w = (X^T X)^{-1} X^T y$$

$$\sigma = \frac{1}{N-n} \sum_{i=1}^{N} (y_i - f(x_i))^2$$

Grebena (eng. ridge) linearna regresija

- Šta ako matrica $X^T X$ nije invertibilna ili je loše uslovljena?

Grebena (eng. ridge) linearna regresija

- Šta ako matrica $X^T X$ nije invertibilna ili je loše uslovljena?
- Zato se razmatra regularizovani problem:

$$\min_w \sum_{i=1}^{N} (y_i - f_w(x_i))^2 + \frac{\lambda}{2} \|w\|_2^2$$

Grebena (eng. ridge) linearna regresija

- Šta ako matrica $X^T X$ nije invertibilna ili je loše uslovljena?
- Zato se razmatra regularizovani problem:

$$\min_w \sum_{i=1}^{N} (y_i - f_w(x_i))^2 + \frac{\lambda}{2} \|w\|_2^2$$

- Rešenje

$$w = (X^T X + \lambda I)^{-1} X^T Y$$

Grebena (eng. ridge) linearna regresija

- Šta ako matrica $X^T X$ nije invertibilna ili je loše uslovljena?
- Zato se razmatra regularizovani problem:

$$\min_w \sum_{i=1}^{N} (y_i - f_w(x_i))^2 + \frac{\lambda}{2} \|w\|_2^2$$

- Rešenje

$$w = (X^T X + \lambda I)^{-1} X^T Y$$

- Regularizacija čini da matrica koja se invertuje ima pun rang

Izazovi i nedostaci

- ako je matrica $X^T X$ prevelika, od izvedenih rešenja se odustaje u korist gradijentnih tehnika optimizacije

Izazovi i nedostaci

- ako je matrica $X^T X$ prevelika, od izvedenih rešenja se odustaje u korist gradijentnih tehnika optimizacije
- osetljivost na odudarajuće podatke

Izazovi i nedostaci

- ako je matrica $X^T X$ prevelika, od izvedenih rešenja se odustaje u korist gradijentnih tehnika optimizacije
- osetljivost na odudarajuće podatke
- nezavisnost dejstva atributa na ciljnu promenljivu

Prednosti

- interpretabilnost

Prednosti

- interpretabilnost
- moguće je analizirati i interpretirati dobijeni model ali tek nakon pretprocesiranja podataka (npr. standardizacijom)

Prednosti

- interpretabilnost
- moguće je analizirati i interpretirati dobijeni model ali tek nakon pretprocesiranja podataka (npr. standardizacijom)
- standardizacija ili neka slična transformacija se uvek vrši zbog boljih računskih svojstava kao što je brža konvergencija metoda

Prednosti

- interpretabilnost
- moguće je analizirati i interpretirati dobijeni model ali tek nakon pretprocesiranja podataka (npr. standardizacijom)
- standardizacija ili neka slična transformacija se uvek vrši zbog boljih računskih svojstava kao što je brža konvergencija metoda
- korišćenje kategoričkih atributa

Primene

- Ustanovljavanje veze između pušenja i rizika od bolesti
- Prvi algoritam koji vredi isprobati u svakom regresionom problemu

Pregled

Linearna regresija

Logistička regresija

Multinomijalna logistička regresija

Logistička regresija

- Neka važi $y \in \{0, 1\}$

Logistička regresija

- Neka važi $y \in \{0, 1\}$
- Želimo da modelujemo uslovnu verovatnoću $p(y|x)$

Logistička regresija

- Neka važi $y \in \{0, 1\}$
- Želimo da modelujemo uslovnu verovatnoću $p(y|x)$
- Pretpostavimo da ciljna promenljiva y ima Bernulijevu raspodelu pri datim vrednostima atributa x

Logistička regresija

- Neka važi $y \in \{0, 1\}$
- Želimo da modelujemo uslovnu verovatnoću $p(y|x)$
- Pretpostavimo da ciljna promenljiva y ima Bernulijevu raspodelu pri datim vrednostima atributa x
- To znači da za date vrednosti atributa x postoji parametar $\mu \in [0, 1]$ tako da važi

$$p(y|x) = \begin{cases} \mu, & y = 1 \\ 1 - \mu, & y = 0 \end{cases} \tag{1}$$

Logistička regresija

- Neka važi $y \in \{0, 1\}$
- Želimo da modelujemo uslovnu verovatnoću $p(y|x)$
- Pretpostavimo da ciljna promenljiva y ima Bernulijevu raspodelu pri datim vrednostima atributa x
- To znači da za date vrednosti atributa x postoji parametar $\mu \in [0, 1]$ tako da važi

$$p(y|x) = \begin{cases} \mu, & y = 1 \\ 1 - \mu, & y = 0 \end{cases} \tag{1}$$

- Ovako zadat model nije kompletan jer nije ustanovljena zavisnost parametra μ od atributa x

Logistička regresija

- Razmotrimo linearnu formu modela:

$$f_w(x) = w_0 + \sum_{i=1}^{n} w_i x_i$$

Logistička regresija

- Razmotrimo linearnu formu modela:

$$f_w(x) = w_0 + \sum_{i=1}^{n} w_i x_i$$

- Da li uslovnu verovatnoću možemo modelovati ovom linearnom formom? $p(y|x) = f_w(x)$?

Logistička regresija

- Razmotrimo linearnu formu modela:

$$f_w(x) = w_0 + \sum_{i=1}^{n} w_i x_i$$

- Da li uslovnu verovatnoću možemo modelovati ovom linearnom formom? $p(y|x) = f_w(x)$?
- Ne, jer $f_w(x)$ nije u intervalu $[0, 1]$ već u $[-\infty, +\infty]$

Logistička regresija

- Razmotrimo linearnu formu modela:

$$f_w(x) = w_0 + \sum_{i=1}^{n} w_i x_i$$

- Da li uslovnu verovatnoću možemo modelovati ovom linearnom formom? $p(y|x) = f_w(x)$?
- Ne, jer $f_w(x)$ nije u intervalu $[0, 1]$ već u $[-\infty, +\infty]$
- Treba nam neka monotona (i po mogućstvu neprekidna i diferencijabilna) funkcija koja će preslikati interval $[-\infty, +\infty]$ u interval $[0, 1]$

Logistička regresija

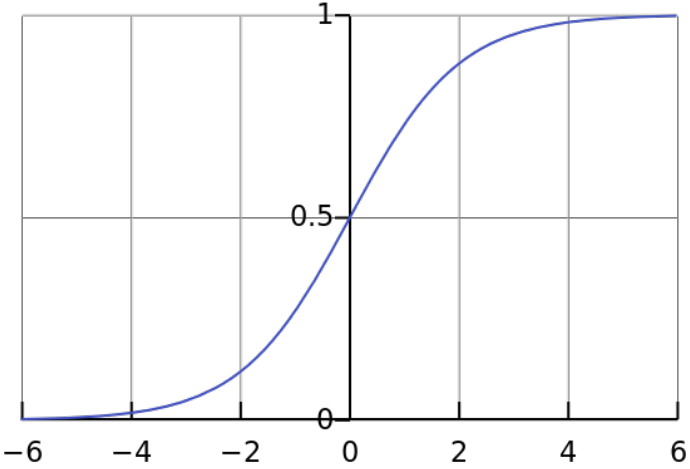
- Razmotrimo linearnu formu modela:

$$f_w(x) = w_0 + \sum_{i=1}^{n} w_i x_i$$

- Da li uslovnu verovatnoću možemo modelovati ovom linearnom formom? $p(y|x) = f_w(x)$?
- Ne, jer $f_w(x)$ nije u intervalu $[0, 1]$ već u $[-\infty, +\infty]$
- Treba nam neka monotona (i po mogućstvu neprekidna i diferencijabilna) funkcija koja će preslikati interval $[-\infty, +\infty]$ u interval $[0, 1]$
- U ovakvu svrhu se često koristi sigmoidna funkcija:

$$\sigma(x) = \frac{1}{1 + \exp(-x)}$$

Sigmoidna funkcija



Slika: https://en.wikipedia.org/wiki/Sigmoid_function

Logistička regresija

- Forma modela:

$$f_w(x) = \sigma(w \cdot x)$$

Logistička regresija

- Forma modela:

$$f_w(x) = \sigma(w \cdot x)$$

- $p_w(y = 1|x) = \sigma(w \cdot x)$

Logistička regresija

- Forma modela:
$$f_w(x) = \sigma(w \cdot x)$$
- $p_w(y = 1|x) = \sigma(w \cdot x)$
- Željenu verovatnoću modelujemo kao $p_w(y|x) = \sigma(w \cdot x)^y (1 - \sigma(w \cdot x))^{1-y}$

Logistička regresija

- Forma modela:
$$f_w(x) = \sigma(w \cdot x)$$
- $p_w(y = 1|x) = \sigma(w \cdot x)$
- Željenu verovatnoću modelujemo kao $p_w(y|x) = \sigma(w \cdot x)^y (1 - \sigma(w \cdot x))^{1-y}$
- Da li je model generativni ili diskriminativni?

Logistička regresija

- Forma modela:
$$f_w(x) = \sigma(w \cdot x)$$
- $p_w(y = 1|x) = \sigma(w \cdot x)$
- Željenu verovatnoću modelujemo kao $p_w(y|x) = \sigma(w \cdot x)^y (1 - \sigma(w \cdot x))^{1-y}$
- Da li je model generativni ili diskriminativni?
- Kako izabrati parametre w?

Logistička regresija

- Uslovna verovatnoća opažanja iz skupa podataka je

$$\prod_{i=1}^{N} p_w(y_i|x_i)$$

Logistička regresija

- Uslovna verovatnoća opažanja iz skupa podataka je

$$\prod_{i=1}^{N} p_w(y_i|x_i)$$

- Ova vrednost se naziva verodostojnošću parametara

Logistička regresija

- Uslovna verovatnoća opažanja iz skupa podataka je

$$\prod_{i=1}^{N} p_w(y_i|x_i)$$

- Ova vrednost se naziva verodostojnošću parametara
- Potrebno je naći vrednosti parametara w čija je verodostojnost najveća

Logistička regresija

- Uslovna verovatnoća opažanja iz skupa podataka je

$$\prod_{i=1}^{N} p_w(y_i|x_i)$$

- Ova vrednost se naziva verodostojnošću parametara
- Potrebno je naći vrednosti parametara w čija je verodostojnost najveća
- Kako su sume numerički i analitički pogodnije od proizvoda, razmatra se logaritam ovog proizvoda

Logistička regresija

- Uslovna verovatnoća opažanja iz skupa podataka je

$$\prod_{i=1}^{N} p_w(y_i|x_i)$$

- Ova vrednost se naziva verodostojnošću parametara
- Potrebno je naći vrednosti parametara w čija je verodostojnost najveća
- Kako su sume numerički i analitički pogodnije od proizvoda, razmatra se logaritam ovog proizvoda
- Obično se umesto maksimizacije koristi minimizacija, pa se razmatra negativna vrednost logaritma

Logistička regresija

- Uslovna verovatnoća opažanja iz skupa podataka je

$$\prod_{i=1}^{N} p_w(y_i|x_i)$$

- Ova vrednost se naziva verodostojnošću parametara
- Potrebno je naći vrednosti parametara w čija je verodostojnost najveća
- Kako su sume numerički i analitički pogodnije od proizvoda, razmatra se logaritam ovog proizvoda
- Obično se umesto maksimizacije koristi minimizacija, pa se razmatra negativna vrednost logaritma
- Negativna vrednost logaritma verodostojnosti:

$$\mathcal{L}(w) = -\sum_{i=1}^{N} \log p_w(y_i|x_i) = -\sum_{i=1}^{N} [y_i \log \sigma(w \cdot x_i) + (1 - y_i) \log(1 - \sigma(w \cdot x_i)))]$$

Logistička regresija

- Optimizacioni problem

$$\min_{w} -\sum_{i=1}^{N}[y_i \log \sigma(w \cdot x_i) + (1 - y_i) \log(1 - \sigma(w \cdot x_i))]$$

- Ova funkcija je konveksna, pa ima jedinstven globalni minimum
- Za optimizaciju se obično koristi Njutnova metoda ili bilo koja gradijentna metoda

Unakrsna entropija

- Optimizacioni problem

$$\min_{w} -\sum_{i=1}^{N}[y_i \log \sigma(w \cdot x_i) + (1 - y_i)\log(1 - \sigma(w \cdot x_i))]$$

Unakrsna entropija

- Optimizacioni problem

$$\min_{w} -\sum_{i=1}^{N}[y_i \log \sigma(w \cdot x_i) + (1 - y_i)\log(1 - \sigma(w \cdot x_i))]$$

- $L(u, v) = -u \log v - (1 - u)\log(1 - v)$

Unakrsna entropija

- Optimizacioni problem

$$\min_{w} -\sum_{i=1}^{N}[y_i \log \sigma(w \cdot x_i) + (1 - y_i)\log(1 - \sigma(w \cdot x_i))]$$

- $L(u, v) = -u \log v - (1 - u)\log(1 - v)$
- u je tačna vrednost a v predviđanje

Unakrsna entropija

- Optimizacioni problem

$$\min_{w} -\sum_{i=1}^{N}[y_i \log \sigma(w \cdot x_i) + (1 - y_i)\log(1 - \sigma(w \cdot x_i))]$$

- $L(u, v) = -u \log v - (1 - u)\log(1 - v)$
- u je tačna vrednost a v predviđanje
- L je funkcija greške (kad nema greške, vrednost funkcije je 0 a kad ima, onda je pozitivna) koja se naziva *unakrsna entropija*

Zašto se zove regresija kad je klasifikacioni model?

- Logistička regresija modeluje logaritam količnika verovatnoća dve različite klase linearnim modelom

$$\log \frac{p(y=1|x)}{p(y=0|x)} = \log \frac{\frac{1}{1+\exp(-w\cdot x)}}{\frac{\exp(-w\cdot x)}{1+exp(-w\cdot x)}} = w \cdot x$$

Primena – ...svuda

- Predviđanje 30-dnevnog mortaliteta od infarkta na osnovu zdravstvenih podataka (pol, visina, težina, pušenje, krvni pritisak, itd.)
- Posebna popularnost u medicniskim primenama
- Standardni model za poređenje prilikom klasifikacije

Pregled

Linearna regresija

Logistička regresija

Multinomijalna logistička regresija

Multinomijalna logistička regresija

- Metod višeklasne klasifikacije

Multinomijalna logistička regresija

- Metod višeklasne klasifikacije
- Predstavlja ocenu *kategoričke* raspodele:

$$p(y|x) = Cat(p_1, \ldots, p_C) = \begin{cases} p_1, & y = 1 \\ p_2, & y = 2 \\ \vdots & \vdots \\ p_C, & y = C \end{cases}$$

Multinomijalna logistička regresija

- Metod višeklasne klasifikacije
- Predstavlja ocenu *kategoričke* raspodele:

$$p(y|x) = Cat(p_1, \ldots, p_C) = \begin{cases} p_1, & y = 1 \\ p_2, & y = 2 \\ \vdots & \vdots \\ p_C, & y = C \end{cases}$$

- C je broj klasa

Multinomijalna logistička regresija

- Metod višeklasne klasifikacije
- Predstavlja ocenu *kategoričke* raspodele:

$$p(y|x) = Cat(p_1, \ldots, p_C) = \begin{cases} p_1, & y = 1 \\ p_2, & y = 2 \\ \vdots & \vdots \\ p_C, & y = C \end{cases}$$

- C je broj klasa
- Mora da važi $p_1 + \ldots + p_C = 1$ i $p_i \geq 0$ za svako $i = 1, \ldots, C$

Multinomijalna logistička regresija

- Metod višeklasne klasifikacije
- Predstavlja ocenu *kategoričke* raspodele:

$$p(y|x) = Cat(p_1, \ldots, p_C) = \begin{cases} p_1, & y = 1 \\ p_2, & y = 2 \\ \vdots & \vdots \\ p_C, & y = C \end{cases}$$

- C je broj klasa
- Mora da važi $p_1 + \ldots + p_C = 1$ i $p_i \geq 0$ za svako $i = 1, \ldots, C$
- svako p_i zavisi od vektora atributa x

Multinomijalna logistička regresija

- Na koji način p_i zavisi od x?

Multinomijalna logistička regresija

- Na koji način p_i zavisi od x?
- Ovu zavisnost ćemo modelovati analogno logističkoj regresiji za binarni slučaj

Multinomijalna logistička regresija

- Na koji način p_i zavisi od x?
- Ovu zavisnost ćemo modelovati analogno logističkoj regresiji za binarni slučaj
- Kod logističke regresije važi

$$\log \frac{p(y=1|x)}{p(y=0|x)} = \log \frac{\frac{1}{1+\exp(-w\cdot x)}}{\frac{\exp(-w\cdot x)}{1+exp(-w\cdot x)}} = w \cdot x$$

Multinomijalna logistička regresija

- Na koji način p_i zavisi od x?
- Ovu zavisnost ćemo modelovati analogno logističkoj regresiji za binarni slučaj
- Kod logističke regresije važi

$$\log \frac{p(y=1|x)}{p(y=0|x)} = \log \frac{\frac{1}{1+\exp(-w\cdot x)}}{\frac{\exp(-w\cdot x)}{1+exp(-w\cdot x)}} = w \cdot x$$

- Analogno, modelovaćemo linearnim modelom logaritam odnosa verovatnoća *svih klasa* u odnosu na *jednu* privilegovanu klasu, na primer poslednju klasu C

$$\log \frac{p(y=i|x)}{p(y=C|x)} = w_i \cdot x \qquad i = 1, \ldots, C \tag{2}$$

Multinomijalna logistička regresija

- Želimo da odredimo vektor parametara w tako da za svako i važi:

$$\log \frac{p(y = i|x)}{p(y = C|x)} = w_i \cdot x \qquad i = 1, \ldots, C \tag{3}$$

Multinomijalna logistička regresija

- Želimo da odredimo vektor parametara w tako da za svako i važi:

$$\log \frac{p(y = i|x)}{p(y = C|x)} = w_i \cdot x \qquad i = 1, \ldots, C \tag{3}$$

- Obratimo pažnju da su ovde pojedinačni koeficijenti w_i zapravo *vektori* a w predstavlja *matricu* koeficijenata:

$$w = \begin{pmatrix} w_{11} & \ldots & w_{1n} \\ \vdots & & \vdots \\ w_{C1} & \ldots & w_{Cn} \end{pmatrix}, \qquad w_i = [w_{i1}, \ldots, w_{in}], \qquad i = 1, \ldots, C$$

Multinomijalna logistička regresija

- Primenimo eksponencijalnu funkciju:

$$\log \frac{p(y = i|x)}{p(y = C|x)} = w_i \cdot x \qquad i = 1, \ldots, C \tag{4}$$

Multinomijalna logistička regresija

- Primenimo eksponencijalnu funkciju:

$$\log \frac{p(y = i|x)}{p(y = C|x)} = w_i \cdot x \qquad i = 1, \ldots, C \tag{4}$$

- Dobijamo

$$p(y = i|x) = p(y = C|x) \exp(w_i \cdot x), \qquad i = 1, \ldots, C$$

Multinomijalna logistička regresija

- Primenimo eksponencijalnu funkciju:

$$\log \frac{p(y=i|x)}{p(y=C|x)} = w_i \cdot x \qquad i = 1, \ldots, C \tag{4}$$

- Dobijamo

$$p(y=i|x) = p(y=C|x)\exp(w_i \cdot x), \qquad i = 1, \ldots, C$$

- Sumiramo za $i = 1, \ldots, C-1$

$$1 - p(y=C|x) = p(y=C|x)\sum_{i=1}^{C-1}\exp(w_i \cdot x)$$

Multinomijalna logistička regresija

- Sumiramo za $i = 1, \ldots, C-1$

$$1 - p(y = C|x) = p(y = C|x) \sum_{i=1}^{C-1} \exp(w_i \cdot x)$$

Multinomijalna logistička regresija

- Sumiramo za $i = 1, \ldots, C-1$

$$1 - p(y = C|x) = p(y = C|x) \sum_{i=1}^{C-1} \exp(w_i \cdot x)$$

- Odavde:

$$1 = p(y = C|x)(1 + \sum_{i=1}^{C-1} \exp(w_i \cdot x))$$

Odnosno:

$$p(y = C|x) = \frac{1}{1 + \sum_{i=1}^{C-1} \exp(w_i \cdot x)}$$

Multinomijalna logistička regresija

- Digresija:

Multinomijalna logistička regresija

- Digresija:
- Setimo se da za svako $i = 1, \ldots, C$ važi

$$p(y = i|x) = p(y = C|x) \exp(w_i \cdot x), \qquad i = 1, \ldots, C$$

Multinomijalna logistička regresija

- Digresija:
- Setimo se da za svako $i = 1, \ldots, C$ važi

$$p(y = i|x) = p(y = C|x) \exp(w_i \cdot x), \qquad i = 1, \ldots, C$$

- Specijalno, za $i = C$:

$$p(y = C|x) = p(y = C|x) \exp(w_C \cdot x)$$

Multinomijalna logistička regresija

- Digresija:
- Setimo se da za svako $i = 1, \ldots, C$ važi

$$p(y = i|x) = p(y = C|x) \exp(w_i \cdot x), \qquad i = 1, \ldots, C$$

- Specijalno, za $i = C$:

$$p(y = C|x) = p(y = C|x) \exp(w_C \cdot x)$$

- Ova jednakost može važiti samo za $\exp(w_C \cdot x) = 1$ odnosno $w_C = 0$

Multinomijalna logistička regresija

- Vratimo se na formulu:

$$p(y = C|x) = \frac{1}{1 + \sum_{i=1}^{C-1} \exp(w_i \cdot x)}$$

Multinomijalna logistička regresija

- Vratimo se na formulu:

$$p(y = C|x) = \frac{1}{1 + \sum_{i=1}^{C-1} \exp(w_i \cdot x)}$$

- Jedinicu u imeniocu možemo zameniti sa $\exp(w_C \cdot x) = 1$ i tada dobijamo:

$$p(y = C|x) = \frac{1}{\exp(w_C \cdot x) + \sum_{i=1}^{C-1} \exp(w_i \cdot x)} = \frac{1}{\sum_{i=1}^{C} \exp(w_i \cdot x)}$$

Multinomijalna logistička regresija

- Vratimo se na formulu:

$$p(y = C|x) = \frac{1}{1 + \sum_{i=1}^{C-1} \exp(w_i \cdot x)}$$

- Jedinicu u imeniocu možemo zameniti sa $\exp(w_C \cdot x) = 1$ i tada dobijamo:

$$p(y = C|x) = \frac{1}{\exp(w_C \cdot x) + \sum_{i=1}^{C-1} \exp(w_i \cdot x)} = \frac{1}{\sum_{i=1}^{C} \exp(w_i \cdot x)}$$

- Kada ovaj izraz uvrstimo u formulu za $p(y = i|x)$, dobijamo

$$p(y = i|x) = \frac{\exp(w_i \cdot x)}{\sum_{i=1}^{C} \exp(w_i \cdot x)}, \qquad i = 1, \ldots, C$$

Multinomijalna logistička regresija

- Vratimo se na formulu:

$$p(y = C|x) = \frac{1}{1 + \sum_{i=1}^{C-1} \exp(w_i \cdot x)}$$

- Jedinicu u imeniocu možemo zameniti sa $\exp(w_C \cdot x) = 1$ i tada dobijamo:

$$p(y = C|x) = \frac{1}{\exp(w_C \cdot x) + \sum_{i=1}^{C-1} \exp(w_i \cdot x)} = \frac{1}{\sum_{i=1}^{C} \exp(w_i \cdot x)}$$

- Kada ovaj izraz uvrstimo u formulu za $p(y = i|x)$, dobijamo

$$p(y = i|x) = \frac{\exp(w_i \cdot x)}{\sum_{i=1}^{C} \exp(w_i \cdot x)}, \qquad i = 1, \ldots, C$$

- Za $C = 2$, dobija se baš logistička regresija

Multinomijalna logistička regresija

- Posmatrajmo formulu:

$$p(y = i|x) = \frac{\exp(w_i \cdot x)}{\sum_{i=1}^{C} \exp(w_i \cdot x)}, \qquad i = 1, \ldots, C$$

Multinomijalna logistička regresija

- Posmatrajmo formulu:

$$p(y = i|x) = \frac{\exp(w_i \cdot x)}{\sum_{i=1}^{C} \exp(w_i \cdot x)}, \qquad i = 1, \ldots, C$$

- Imamo ukupno C linearnih modela $w_i \cdot x$ i svaki određuje verovatnoću da data instanca pripada klasi i

Multinomijalna logistička regresija

- Posmatrajmo formulu:

$$p(y = i|x) = \frac{\exp(w_i \cdot x)}{\sum_{i=1}^{C} \exp(w_i \cdot x)}, \qquad i = 1, \ldots, C$$

- Imamo ukupno C linearnih modela $w_i \cdot x$ i svaki određuje verovatnoću da data instanca pripada klasi i
- Što je $w_i \cdot x$ veće, to će i verovatnoća da instanca pripada klasi i biti veća

Multinomijalna logistička regresija

- Posmatrajmo formulu:

$$p(y = i|x) = \frac{\exp(w_i \cdot x)}{\sum_{i=1}^{C} \exp(w_i \cdot x)}, \qquad i = 1, \ldots, C$$

- Imamo ukupno C linearnih modela $w_i \cdot x$ i svaki određuje verovatnoću da data instanca pripada klasi i
- Što je $w_i \cdot x$ veće, to će i verovatnoća da instanca pripada klasi i biti veća
- Ovi modeli se međusobno nadmeću i predviđena klasa će biti ona sa najvećom verovatnoćom

Multinomijalna logistička regresija

- Posmatrajmo formulu:

$$p(y = i|x) = \frac{\exp(w_i \cdot x)}{\sum_{i=1}^{C} \exp(w_i \cdot x)}, \qquad i = 1, \ldots, C$$

- Imamo ukupno C linearnih modela $w_i \cdot x$ i svaki određuje verovatnoću da data instanca pripada klasi i
- Što je $w_i \cdot x$ veće, to će i verovatnoća da instanca pripada klasi i biti veća
- Ovi modeli se međusobno nadmeću i predviđena klasa će biti ona sa najvećom verovatnoćom
- Vrednost $w_i \cdot x$ može biti i pozitivna i negativna

Multinomijalna logistička regresija

- Posmatrajmo formulu:

$$p(y = i|x) = \frac{\exp(w_i \cdot x)}{\sum_{i=1}^{C} \exp(w_i \cdot x)}, \qquad i = 1, \ldots, C$$

- Imamo ukupno C linearnih modela $w_i \cdot x$ i svaki određuje verovatnoću da data instanca pripada klasi i
- Što je $w_i \cdot x$ veće, to će i verovatnoća da instanca pripada klasi i biti veća
- Ovi modeli se međusobno nadmeću i predviđena klasa će biti ona sa najvećom verovatnoćom
- Vrednost $w_i \cdot x$ može biti i pozitivna i negativna
- Kada primenimo eksponencijalnu funkciju, postaje pozitivna

Multinomijalna logistička regresija

- Posmatrajmo formulu:

$$p(y = i|x) = \frac{\exp(w_i \cdot x)}{\sum_{i=1}^{C} \exp(w_i \cdot x)}, \qquad i = 1, \ldots, C$$

- Imamo ukupno C linearnih modela $w_i \cdot x$ i svaki određuje verovatnoću da data instanca pripada klasi i
- Što je $w_i \cdot x$ veće, to će i verovatnoća da instanca pripada klasi i biti veća
- Ovi modeli se međusobno nadmeću i predviđena klasa će biti ona sa najvećom verovatnoćom
- Vrednost $w_i \cdot x$ može biti i pozitivna i negativna
- Kada primenimo eksponencijalnu funkciju, postaje pozitivna
- Kada takvu vrednost podelimo sa sumom, biće između 0 i 1, što znači da je ovako zadata verovatnoća dobro definisana

Ocena parametara metodom maksimalne verodostojnosti

- Model:

$$p(y = i|x) = \frac{\exp(w_i \cdot x)}{\sum_{i=1}^{C} \exp(w_i \cdot x)}, \qquad i = 1, \ldots, C$$

Ocena parametara metodom maksimalne verodostojnosti

- Model:

$$p(y = i|x) = \frac{\exp(w_i \cdot x)}{\sum_{i=1}^{C} \exp(w_i \cdot x)}, \qquad i = 1, \ldots, C$$

- Potrebno je oceniti parametre tako da model što bolje odgovara datim podacima

Ocena parametara metodom maksimalne verodostojnosti

- Model:

$$p(y = i|x) = \frac{\exp(w_i \cdot x)}{\sum_{i=1}^{C} \exp(w_i \cdot x)}, \qquad i = 1, \ldots, C$$

- Potrebno je oceniti parametre tako da model što bolje odgovara datim podacima
- Izvođenje je analogno izvođenju kod logističke regresije uz jedan dogovor: vrednost ciljne promenljive $y = i$ predstavlja se vektorom dimenzije C koji ima vrednost 1 na poziciji i i vrednost 0 na ostalim pozicijama: $(0, \ldots, 0, 1, 0, \ldots, 0)$

Ocena parametara metodom maksimalne verodostojnosti

- Model:

$$p(y = i|x) = \frac{\exp(w_i \cdot x)}{\sum_{i=1}^{C} \exp(w_i \cdot x)}, \qquad i = 1, \ldots, C$$

- Potrebno je oceniti parametre tako da model što bolje odgovara datim podacima
- Izvođenje je analogno izvođenju kod logističke regresije uz jedan dogovor: vrednost ciljne promenljive $y = i$ predstavlja se vektorom dimenzije C koji ima vrednost 1 na poziciji i i vrednost 0 na ostalim pozicijama: $(0, \ldots, 0, 1, 0, \ldots, 0)$
- Funkcija verodostojnosti je data izrazom:

$$\mathcal{L}(w) = p(y_1 \ldots y_N | x_1 \ldots x_N) = \prod_{i=1}^{N} p(y_i|x_i) = \prod_{i=1}^{N} \prod_{j=1}^{C} \left(\frac{\exp(w_j \cdot x_i)}{\sum_{k=1}^{C} \exp(w_k \cdot x_i)} \right)^{y_{ij}}$$

Ocena parametara metodom maksimalne verodostojnosti

- Prelaskom na negativnu vrednost logaritama funkcije verodostojnosti, nakon izvođenja se dobija

$$\min_{w_1,\ldots,w_{C-1}} -\sum_{i=1}^{N}\left[\sum_{j=1}^{C} y_{ij} w_j \cdot x_i - \log\left(\sum_{k=1}^{C} \exp(w_k \cdot x_i)\right)\right]$$

Ocena parametara metodom maksimalne verodostojnosti

- Prelaskom na negativnu vrednost logaritama funkcije verodostojnosti, nakon izvođenja se dobija

$$\min_{w_1,\ldots,w_{C-1}} -\sum_{i=1}^{N}\left[\sum_{j=1}^{C} y_{ij} w_j \cdot x_i - \log\left(\sum_{k=1}^{C}\exp(w_k \cdot x_i)\right)\right]$$

- Vektor koeficijenata w_C je nula vektor pa se po njemu ne minimizuje

Ocena parametara metodom maksimalne verodostojnosti

- Prelaskom na negativnu vrednost logaritama funkcije verodostojnosti, nakon izvođenja se dobija

$$\min_{w_1,\ldots,w_{C-1}} -\sum_{i=1}^{N}\left[\sum_{j=1}^{C} y_{ij} w_j \cdot x_i - \log\left(\sum_{k=1}^{C}\exp(w_k \cdot x_i)\right)\right]$$

- Vektor koeficijenata w_C je nula vektor pa se po njemu ne minimizuje
- Logaritam sume eksponencijalnih funkcija je konveksna funkcija, pa je i ceo problem konveksan i pogodan za gradijentne metode optimizacije