INTELIGENTNI SISTEMI

as. ms Vladimir Jocović
as. ms Adrian Milaković

LINEARNA I LOGISTIČKA REGRESIJA

07

„Sometimes a straight line is all you need to make sense of the world. “
- Unknown author

LINEARNA REGRESIJA

Šta je linearna regresija i u kojim oblicima postoji?

Linearna regresija (*Linear regression*) je tip nadgledanih algoritama mašinskog učenja. Koristi se kod problema regresije.

Algoritam kreće od pretpostavke da zavisnost izlaznog podatka *y* od ulaznog podatka *x* ima oblik linearne kombinacije odlika i njihovih težina.

$$h(x) = w_0 + w_1 x_1 + w_2 x_2 + \ldots + w_n x_n$$

Cilj algoritma je pronaći težinske parametre modela w_0, w_1, ... tako da hipoteza h(x) bude bliska stvarnim izlaznim podacima *y* za sve ulazne parove (*x*, *y*).

U zavisnosti od broja ulaznih podataka razlikujemo jednostruku i višestruku linearnu regresiju. U zavisnosti od broja izlaznih podataka razlikujemo univarijantnu i multivarijantnu linearnu regresiju.

LINEARNA REGRESIJA

Na koji način izabrati parametre modela?

Parametre modela je potrebno izabrati tako da hipoteza što manje „odskače“ od stvarnih vrednosti *y*. U tu svrhu se definiše funkcija gubitka ***L(h(x), y)*** i funkcija greške ***J(w)*** koja predstavlja prosek vrednosti funkcije gubitka na svim podacima iz posmatranog skupa.

Obučavanje modela se svodi na izbor parametara takvih da je funkcija greške minimalna.

Kao funkcija greške najčešće se koristi prosek kvadrata odstupanja *h(x)* od *y* (*Mean Squared Error*):

$$L(h(x), y) = \frac{1}{2}\left(h\left(x^{(i)}\right) - y^{(i)}\right)^2$$

$$J(w) = \frac{1}{2m}\sum_{i=1}^{m}\left(h\left(x^{(i)}\right) - y^{(i)}\right)^2$$

LINEARNA REGRESIJA

Kako naći minimum funkcije greške?

Analitički pristup nalasku rešenja nalaženja minimuma funkcije greške može biti skupa operacija.

U praksi se najčešće koristi metoda gradijentnog spusta koji iterativno traži minimum funkcije računajući izvode za veći broj promenljivih (gradijente). Gradijent funkcije greške možemo definisati sa:

$$\nabla J(w) = (\frac{\partial J(w)}{\partial w_0}, \frac{\partial J(w)}{\partial w_1}, \dots, \frac{\partial J(w)}{\partial w_n})$$

Parametre u svakoj iteraciji treba menjati tako da funkcija greške opada:

$$w_i = w_i - \alpha \frac{\partial J(w)}{\partial w_i}$$

Gde je α pozitivan ceo broj koji predstavlja korak učenja i određuje koliko veliki skokovi će se praviti niz gradijent i mora biti pažljivo odabran.

LINEARNA REGRESIJA

Kako naći minimum funkcije greške?

Ukoliko je korak učenja premali, postoji preveliki broj iteracija i moguće je zaglaviti se u lokalnom minimumu funkcije.

Ukoliko je korak učenja preveliki, moguće je preskočiti globalni minimum funkcije.

Gradijentni spust konvergira ka minimumu smanjenjem vrednosti parcijalnih izvoda. Kako je vrednost izvoda u minimumu jednaka nuli, u nekom trenutku se parametri neće više menjati. Međutim, u praksi, često se gradijentni spust obustavlja kada promena vrednosti parametra u jednom koraku bude jako mala (približna nuli), odnosno manja od unapred definisane vrednosti.

JEDNOSTRUKA LINEARNA REGRESIJA

$$J(w) = \frac{1}{2m}\sum_{i=1}^{m}\left(h\left(x^{(i)}\right) - y^{(i)}\right)^2 = \frac{1}{2m}\sum_{i=1}^{m}\left(w_0 + w_1 x^{(i)} - y^{(i)}\right)^2$$

$$\frac{\partial J(w)}{\partial w_0} = \frac{1}{m}\sum_{i=1}^{m}\left(w_0 + w_1 x^{(i)} - y^{(i)}\right) = \frac{1}{m}\sum_{i=1}^{m}\left(h\left(x^{(i)}\right) - y^{(i)}\right)$$

$$\frac{\partial J(w)}{\partial w_1} = \frac{1}{m}\sum_{i=1}^{m}\left(w_0 + w_1 x^{(i)} - y^{(i)}\right)x^{(i)} = \frac{1}{m}\sum_{i=1}^{m}\left(h\left(x^{(i)}\right) - y^{(i)}\right)x^{(i)}$$

JEDNOSTRUKA LINEARNA REGRESIJA

Ukoliko je skup podataka mali, rešavanjem parcijalnih izvoda i izjednačavanjem sa nulom, možemo analitički pronaći parametre modela:

$$w_1 = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

$$w_0 = \frac{\sum y - w_1(\sum x)}{n}$$

Zadatak 6 - Zavisnost uspeha od učenja

★★★☆☆

Na osnovu posmatranog skupa podataka o zavisnosti uspeha na ispitu od broja sati provedenih u učenju izvesti zaključak o uspehu ukoliko je student učio 4.5 sata koristeći linearnu regresiju i izračunati grešku.

Nezavisni atribut	**Atribut odluke**
Sati učenja	Broj poena na ispitu
1	30
2	45
3	51
4	57
5	60
6	65
7	70
8	71

Zadatak 6 - Rešenje

Zadatak ćemo rešiti analitičkim putem. Potrebno je izračunati sve varijacije suma koje su pojavljuju u formulama za parametre modela.

	Sati učenja	Broj poena na ispitu	xy	x^2	y^2
	1	30	30	1	900
	2	45	90	4	2025
	3	51	153	9	2601
	4	57	228	16	3249
	5	60	300	25	3600
	6	65	390	36	4225
	7	70	490	49	4900
	8	71	568	64	5041
Σ	**36**	**449**	**2249**	**204**	**26541**

Zadatak 6 - Rešenje

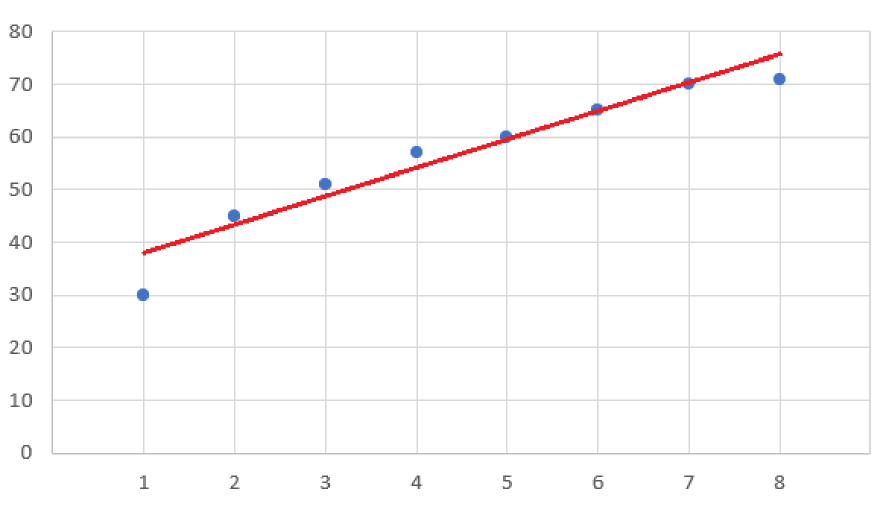
	Sati učenja	Broj poena na ispitu	xy	x^2	y^2
	1	30	30	1	900
	2	45	90	4	2025
	3	51	153	9	2601
	4	57	228	16	3249
	5	60	300	25	3600
	6	65	390	36	4225
	7	70	490	49	4900
	8	71	568	64	5041
Σ	36	449	2249	204	26541

$$w_1 = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} = \frac{8*2249 - 36*449}{8*204 - 36^2} = 5.4405$$

$$w_0 = \frac{\sum y - w_1(\sum x)}{n} = \frac{449 - 5.4405 * 36}{8} = 31.6429$$

Zadatak 6 - Rešenje

$$h(x) = 5.4405x + 31.6429$$



h(4.5) = 5.4405 * 4.5 + 31.6429 = 56.125

Zadatak 6 - Rešenje

$$h(x) = 5.4405x + 31.6429$$

	Sati učenja	Broj poena na ispitu	$\boldsymbol{h(x)}$	$\boldsymbol{h(x) - y}$	$\boldsymbol{(h(x) - y)^2}$
	1	30	37.08	7.08	50.17
	2	45	42.52	-2.48	6.13
	3	51	47.96	-3.04	9.21
	4	57	53.40	-3.60	12.92
	5	60	58.85	-1.15	1.33
	6	65	64.29	-0.71	0.51
	7	70	69.73	-0.27	0.07
	8	71	75.17	4.17	17.36
∑					**97.72**

$$J(w) = \frac{1}{2m}\sum_{i=1}^{m}\left(h\left(x^{(i)}\right) - y^{(i)}\right)^2 = \frac{1}{16} * 97.72 = 6.11$$

Zadatak 6 - Rešenje

Drugi način za proveru valjanosti modela jeste koeficijent korelacije. Ako skup podataka nije linearan, model neće dati dobre rezultate. Koeficijent korelacije meri jačinu linearne veze između dva atributa. Uzima vrednosti u opsegu [-1, 1], pri čemu granične vrednosti ukazuju na potpunu linearnost, dok vrednost 0 ukazuje na potpuno odsustvo linearnosti. Koeficijent korelacije je dat formulom:

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2} * \sqrt{n(\sum y^2) - (\sum y)^2}}$$

	Sati učenja	Broj poena na ispitu	xy	x^2	y^2
$\sum$	36	449	2249	204	26541

$$r = \frac{8(2249) - (36)(449)}{\sqrt{8(204) - (36)^2} * \sqrt{8(26541) - (449)^2}} = 0.9629$$

Vrednost koeficijenta je približna jedinici te možemo da kažemo da je korelacija jaka i da je model dobar.

Zadatak za samostalnu vežbu - Maratonci

★★★☆☆

Na osnovu posmatranog skupa podataka o zavisnosti rezultata na maratonu od starosti maratonca, proceniti rezultat osobe koja ima 50 godina koristeći model linearne regresije i izračunati grešku.

Nezavisni atribut	**Atribut odluke**
Starost (godina)	Rezultat (h:m)
23	2:56
25	3:01
35	3:23
47	3:27
52	3:17
60	3:57
62	4:13

Zadatak za samostalnu vežbu - Plata

★★★☆☆

Na osnovu posmatranog skupa podataka o zavisnosti plate na osnovu godina radnog iskustva izvesti zaključak o plati zaposlenog sa 2.5 godina radnog iskustva koristeći linearnu regresiju i izračunati grešku.

Nezavisni atribut	Atribut odluke
Sati učenja	Broj poena na ispitu
1.1	39434
1.3	46205
1.5	37731
2.0	43525
2.2	39891
2.9	56642
3.0	60150
3.2	54445
3.2	64445
3.7	57189
3.9	63218

LOGISTIČKA REGRESIJA

Šta je logistička regresija i u kojim oblicima postoji?

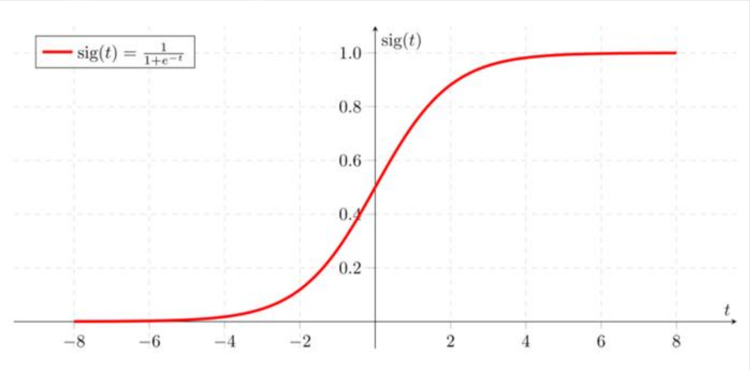
Logistička regresija (*Logistic regression*) je tip nadgledanih algoritama mašinskog učenja. Koristi se kod problema klasifikacije, kada je potrebno dati predikciju klase (kategorička vrednost) kojoj neki primerak pripada na osnovu jednog ili više njegovih atributa (kontinualnih ili kategoričkih).

Najprostiji oblik logističke regresije je **binarna logistička regresija**, kada je izlaz jedna od 2 moguće vrednosti (1 ili 0). U slučaju više od 2 moguće izlazne vrednosti radi se o multinomijalnoj logističkoj regresiji (koristi se softmax funkcija umesto sigmoid funkcije), a ukoliko je moguće uspostaviti redosled izlaznih vrednosti (npr. ocene 1-5) reč je o ordinalnoj logističkoj regresiji.

LOGISTIČKA REGRESIJA

Šta je logistička regresija i u kojim oblicima postoji?

Binarna logistička regresija koristi logističku (sigmoid) funkciju da modeluje zavisnu izlaznu promenljivu (verovatnoću pripadnosti klasi 1) na osnovu ulaznih podataka.



LOGISTIČKA REGRESIJA

Zašto se zove Logistička regresija, a ne Logistička klasifikacija?

Logistička regresija je generalizovani linearni model (utvrđuje odlike klasa, za razliku od diskriminativnog modela koji utvrđuje razlike između klasa) koji koristi formulu linearne zavisnosti izlaznog podatka od ulaznih, kao i linearna regresija

$$\boldsymbol{y = w_0 + w_1x_1 + w_2x_2 + \dots + w_nx_n}$$

koju upotrebljava kao argument sigmoid funkcije za predikciju izlaza:

$$\boldsymbol{h(x)} = \frac{1}{1+\ e^{-y}} = \frac{1}{1+\ \boldsymbol{e}^{-(\boldsymbol{w0}\,+\,\boldsymbol{w1x1}\,+\,\boldsymbol{w2x2}\,+\,\dots+\,\boldsymbol{wnxn})}}$$

Izlaz Linearne regresije može biti bilo koja kontinualna vrednost u opsegu [-∞, +∞], dok sigmoid funkcija dovodi te vrednosti u opseg [0, 1].

LOGISTIČKA REGRESIJA

Kako odrediti parametre modela?

Kao i kod Linearne regresije, parametri modela treba da budu takvi da predviđena vrednost izlaza bude što približnija stvarnoj vrednosti. Za razliku od Linearne regresije, kod Logističke regresije u opštem slučaju ne postoji analitički pristup određivanja parametara modela (nije moguće rešiti jednačinu gde je prvi izvod cost funkcije jednak nuli). Utvrđivanje parametara moguće je korišćenjem iterativnih algoritama optimizacije, kao što je **gradijentni spust uz minimizaciju funkcije greške**.

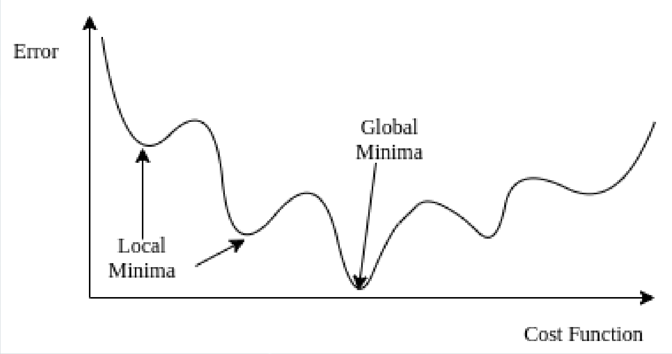
Da li se može koristiti MSE funkcija u metodi gradijentnog spusta?

Za razliku od Linearne regresije, gde je funkcija greške *Mean Squared Error* konveksna, kod Logističke regresije to nije slučaj, jer se linearna kombinacija ulaznih vrednosti i parametara modela koristi kao argument sigmoid funkcije (koja i sama nije konveksna).

LOGISTIČKA REGRESIJA

Da li se može koristiti MSE funkcija u metodi gradijentnog spusta?

Stoga, moguće je da metod gradijentnog spusta ne konvergira ka globalnom minimumu, već ka nekom od lokalnih minimuma, te ovakva funkcija greške ne odgovara modelu Logističke regresije. Zbog toga nije moguće korišćenje MSE funkcije kao cost funkcije.

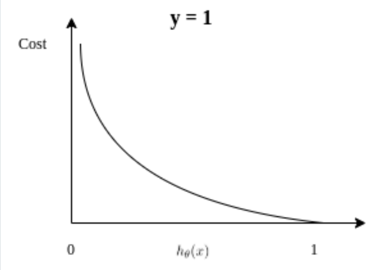


LOGISTIČKA REGRESIJA

<u>*Koju funkciju greške koristiti i zašto?*</u>

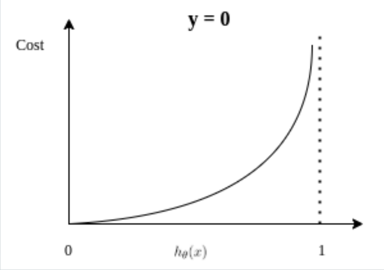
$\boldsymbol{y} = \mathbf{1} \mid Cost\,(h(x), y) = -\ln(h(x))$

- $h(x) = 0 \mid Cost = -\ln(0) = +\infty$
- $h(x) = 1 \mid Cost = -\ln(1) = 0$



$\boldsymbol{y} = \mathbf{0} \mid Cost\,(h(x), y) = -\ln\big(1 - h(x)\big)$

- $h(x) = 0 \mid Cost = -\ln(1) = 0$
- $h(x) = 1 \mid Cost = -\ln(0) = +\infty$



LOGISTIČKA REGRESIJA

Koju funkciju greške koristiti i zašto?

Objedinjena formula

$$\mathrm{Cost}\,(\mathrm{h(x)}, \mathrm{y}) = -\,\mathrm{y} \, * \ln\bigl(\mathrm{h(x)}\bigr) - (1 - \mathrm{y}) * \ln\bigl(1 - \mathrm{h(x)}\bigr)$$

S obzirom da je verovatnoća pripadanja primerka sa atributima x klasi 1

$$P\,(y = 1 \mid x) = h(x)$$

onda je u slučaju dve klase

$$P\,(y = 0 \mid x) = 1 - h(x)$$

odnosno $P\,(\mathrm{y} \mid \mathrm{x}) = h(x)^{y} \, * \, (1 - h(x))^{1-y}$

$$\mathrm{Cost}(\mathrm{h(x)}, y) = -\ln\left(h(x)^{y} \, * \, (1 - h(x))^{1-y}\right) = -\ln(P\,(y \mid x))$$

Veća verovatnoća pripadnosti primerka označenoj klasi daje manji cost!

Zadatak 7 - Ispit

★★★☆☆

Na osnovu posmatranog skupa podataka o zavisnosti polaganja ispita u odnosu na broj sati učenja, odrediti model Logističke regresije

$$h(x) = \frac{1}{1 + e^{-y}}$$

koji bolje reprezentuje podatke: y = 0.25 * x – 0.125 ili y = 0.125 * x + 0.25?

Nezavisni atribut	**Atribut odluke**
Sati učenja	Položio ispit
1	0
1.5	0
2	0
2.5	0
3	1
3.5	0
4	1
4.5	1
5	1
5.5	1

LOGISTIČKA REGRESIJA

Odnos ishoda i linearne kombinacije ulaznih parametara i težina

Inverzna funkcija sigmoid (logističke funkcije) je logit funkcija:

$$\mathrm{logit(p)} = \ sigmoid^{-1}(p)$$

$$= \ln\left(\frac{p}{1-p}\right) = \ln\left(\frac{h(x)}{1-h(x)}\right) = \ln\left(\frac{\frac{1}{1+e^{-y}}}{\frac{e^{-y}}{1+e^{-y}}}\right) = \ln(e^{y}) = y$$

odnosno $\left(\frac{p}{1-p}\right) = e^{y}$, tj. $\mathrm{p} = (1-p) * e^{y}$, tj. $p + p * e^{y} = e^{y}$. Odatle je:

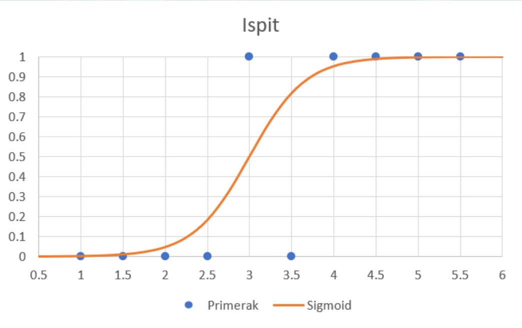
$$p = \frac{e^{y}}{e^{y}+1}$$

U slučaju da je p, odnosno h(x) = 0, onda je y = ln (0) = -∞.

Ukoliko je p, odnosno h(x) = 1, onda je y = ln (1 / 0) = +∞.

Zadatak 7 - Rešenje

Prikazom zavisnosti ishoda polaganja ispita (verovatnoća polaganja) od broja sati učenja možemo videti da podatke najbolje modeluje sigmoid funkcija. Raspoložive podatke možemo prikazati i u zavisnosti od logit (p) i pronaći hiperravan (u ovom slučaju jednog ulaznog parametra to je linija) koja je najbolje prilagođena zadatim podacima. Međutim, problem je što su vrednosti y (tj. logit (p)) za sve ulazne podatke +∞ ili −∞, te nije moguće metodom MSE oceniti kvalitet kandidat best-fit linije.



Zadatak 7 - Rešenje

Stoga, za svaki ulazni podatak izračunaće se best-fit vrednost y (logit (p)), a zatim će se na osnovu te vrednosti izračunati verovatnoća pripadanja klasi 1 podatka koji pripada best-fit liniji, a sve to u cilju ocene kvaliteta kandidat best-fit sigmoida.

Sada na osnovu labele svakog podatka (da li je ishod 1 ili 0) i njegove verovatnoće (pripadanja klasi 1) određujemo očekivanje da taj podatak **pripada labeliranoj klasi**.

likelihoods = [for i in len(data) p[i] if labels[i] == 1 else 1 – p[i]]

Nakon toga se sva očekivanja izmnože. Zbog mogućih malih vrednosti verovatnoća češće se izračunava **log-likelihood** (suma logaritama pojedinačnih očekivanja). Funkcija većeg očekivanja je bolji kandidat, odnosno ima manji cost.

Zadatak 7 - Rešenje

Za funkciju y = 0.25 * x – 0.125

Sum (Log (Likelihood)) = -6.216

Sati (x)	Ishod (y)	Logit (y)	Best-fit y	p	Likelihood	Log (Like.)
1	0	-inf	0.125	0.531	0.469	-0.758
1.5	0	-inf	0.250	0.562	0.438	-0.826
2	0	-inf	0.375	0.593	0.407	-0.898
2.5	0	-inf	0.5	0.622	0.378	-0.974
3	1	+inf	0.625	0.651	0.651	-0.429
3.5	0	-inf	0.750	0.679	0.321	-1.137
4	1	+inf	0.875	0.706	0.706	-0.348
4.5	1	+inf	1	0.731	0.731	-0.313
5	1	+inf	1.125	0.755	0.755	-0.281
5.5	1	+inf	1.250	0.777	0.777	-0.252

Zadatak 7 - Rešenje

Za funkciju y = 0.125 * x + 0.25

Sum (Log (Likelihood)) = -6.778, što znači da je prethodna funkcija bolja.

Sati (x)	Ishod (y)	Logit (y)	Best-fit y	p	Likelihood	Log (Like.)
1	0	-inf	0.375	0.593	0.407	-0.898
1.5	0	-inf	0.438	0.608	0.392	-0.936
2	0	-inf	0.500	0.622	0.378	-0.974
2.5	0	-inf	0.563	0.637	0.363	-1.013
3	1	+inf	0.625	0.651	0.651	-0.429
3.5	0	-inf	0.688	0.665	0.335	-1.095
4	1	+inf	0.750	0.679	0.679	-0.387
4.5	1	+inf	0.813	0.693	0.693	-0.367
5	1	+inf	0.875	0.706	0.706	-0.348
5.5	1	+inf	0.938	0.719	0.719	-0.330

PITANJA?

http://ri4es.etf.rs/