J.T. 2024

Maximum Likelihood -menetelmä MLE

(suom.: suurimman uskottavuuden menetelmä)

Maximum Likelihood Estimate (MLE)is **a method of estimating the parameters of an assumed probability distribution, given some observed data**. This is achieved by maximizing a likelihood function so that, under the assumed statistical model, the observed data is most probable.

Maximum Likelihood Estimate (MLE) on menetelmä oletetun todennäköisyysjakauman parametrien arvioimiseksi käytettävissä olevan havaintodatan perusteella. Tämä saavutetaan maksimoimalla ns. uskottavuusfunktio L (Likelihood function) siten, että oletetun tilastollisen mallin mukaan havaittu data on todennäköisin.

Merkinnät:

- Olkoon x satunnaismuuttuja ja f(x,a,b) sen tiheysfunktio, missä a ja b ovat jakaumaparametrit (esim. a=keskiarvo ja b=keskihajonta (useissa jakaumissa jakaumaparametrit ovat muut kuin keskiarvo ja keskihajonta)).

- Olkoot x_1, x_2, …., x_n havaitut muuttujan x arvot.

Todennäköisyys sille, että havainto x_i on peräisin oletetusta jakaumasta on $f(x_i,a,b)$. (Tämä on itse asiassa tiheysfunktion f määritelmä)

Tuloperiaatteen nojalla todennäköisyys sille, että koko havaintosarja x_1, x_2, …., x_n on peräisin jakaumasta f(x,a,b) = Likelihood funktio L (suomeksi uskottavuusfunktio)

$$L = \prod_{i=1}^{n} f(x_i, a, b)$$

L on siis funktio, jonka muuttujina ovat jakaumaparametrit a ja b. MLE menetelmässä etsitään uskottavuusfunktion L maksimikohta. Tuloksena saadaan jakaumaparametreille a ja b arvot.

Monissa tapauksissa uskottavuusfunktio L sisältää potenssilausekkeiden tuloja ja osamääriä, sekä mahdollisesti eksponenttifunktioita. Tällöin uskottavuusfunktion *L* sijasta haetaan maksimia sen logaritmille ln(L).

$$LL = \ln(L)$$

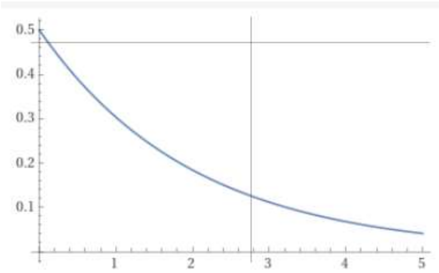
Perustelu: Jos g(x) on aidosti kasvava funktio, niin funktio g(f(x)) saavuttaa maksimiarvon kohdassa, jossa f(x):llä on maksimi. Funktion ln(x) on aidosti kasvava, joten ln(f(x)) saavuttaa maksimiarvon kohdassa, jossa f(x):llä on maksimi

----

Alla on neljä esimerkkiä MLE menetelmän käytöstä. Esimerkit 1 - 3 voidaan laskea helpommin muilla menetelmillä kuin MLE:llä. Esimerkki 4 kuvaa koneoppimiseen liittyvää Logistisen Regression -algoritmia, joka perustuu MLE menetelmään.

J.T. 2024

Esim1: Eksponenttijakauma on yksiparametrinen, yksinkertainen todennäköisyysjakauma, jota käytetään esimerkiksi hehkulamppujen vikaantumisaikojen mallinnuksessa. Eksponenttijakauman tiheysfunktio $f(x,\lambda) = \lambda * e^{-\lambda x}$, missä $x \geq 0$.



Testissä oli 6 hehkulamppua, jotka rikkoontuivat ajoissa (1.1, 1.35, 1.5, 1.75, 1.8 ja 1.97) * 1000 h. Määritä tiheysfunktio $f(x,\lambda)$ eli parametrin λ arvo MLE menetelmällä.

Maksimoitava uskottavuusfunktio $L = \prod \lambda * e^{-\lambda xi} = \lambda^6 e^{-\lambda(x1+x2+...+x)} = \lambda^6 e^{-\lambda * 9.47}$

Funktio on sellaista muoto, että on helpompi maksimoida sen logaritmi

$$LL = 6 \ln(\lambda) - 9.47 \lambda$$ (*käytetyt säännöt liite1:ssä)

Maksimi on derivaatan nollakohdassa: $6/\lambda - 9.47 = 0 \Rightarrow \lambda = 6/9.47 = 0.6336$

Vastaus: Kysytty tiheysfunktio $f(x) = 0.6336 * e^{-0.6336\,x}$

! Edellinen tehtävä voidaan ratkaista helpommin käyttäen tietoa, että eksponenttijakauman $f(x, \lambda)$ keskiarvo on $1/\lambda$. Otoskeskiarvo luvuista (1.1, 1.35, 1.5, 1.75, 1.8 ja 1.97) on 1.5783 ,
joten parametri $\lambda = 1/1.5783 = 0.6336$.

Esim2: binomijakauma. Suomalaisen ampumahiihtäjän pystyammuntapaikalla pudotettujen taulujen lukumäärät kevään MC kiertueella olivat 3,4,4,5,4,5,3,3,4,4. Oletetaan, lukumäärät noudattavat binomijakaumaa, jossa yksittäiseen tauluun osumisen todennäköisyys p on vakio. Määritä p:n arvo perustuen dataan. Käytä MLE menetelmää.

Maksimoitava uskottavuusfunktio on seuraava:

$$L = \prod_{i=1}^{10} \binom{5}{x_i} p^{x_i} (1-p)^{5-x_i}$$

Funktio on muodoltaan sellainen, että sen ääriarvokohdan määritys on helpompi suorittaa käyttämällä logaritmia ln(L), joka saavuttaa maksimiarvon samassa kohdassa.

$$LL = \sum_{i=1}^{10} (\binom{5}{x_i} + x_i ln(p) + (5 - x_i) ln(1-p))$$

Sen derivaatta $\partial LL/\partial p = \sum(0 + x_i/p - (5-x_i)/(1-p))$ jonka nollakohta saadaan yhtälöstä
$\sum x_i/p = \sum(5-x_i)/(1-p)$, josta ristiinkertomalla $(1-p)\sum x_i = p\sum(5-x_i)$, josta edelleen
$\sum x_i = p(\sum x_i + \sum(5-x_i))$ josta saadaan $p = \sum x_i/((\sum x_i + \sum(5-x_i)) = 39/(39+11) = 0.78$.

! Tehtävän ratkaiseminen MLE:llä on turhan monimutkainen tapa, koska tulos on itsestään selvä muutoinkin. Urheilija pudotti taulun 39 laukauksella 50 yrityksestä, joten P(osuu tauluun) = 39/50 = 0.78.

Esim3. Gaussin jakauma , jonka tiheysfunktio on

$$f(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Määritä MLE menetelmällä jakaumaparametrit μ ja σ perustuen seuraaviin satunnaismuuttujan x havaintoarvoihin: 175, 178, 184, 183, 181, 176, 185.

Likelihood-funktio on tässä tapauksessa seuraava:

$$L = \prod_{i=1}^{7} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

Tässäkin tapauksessa on helpompaa käyttää logaritmista ln(L) – funktiota

$$LL = \sum_{i=1}^{n} (ln \frac{1}{\sqrt{2\pi\sigma^2}} + -\frac{(x_i-\mu)^2}{2\sigma^2})$$
$$= \sum_{i=1}^{n} (ln \frac{1}{\sqrt{2\pi}} - ln\sigma - \frac{(x_i-\mu)^2}{2\sigma^2})$$

Minimi löytyy kohdasta, jossa osittaisderivaatat ovat nollia:
$\partial LL/\partial\mu = 2/2\sigma^2 \sum(x_i-\mu)=0 \Rightarrow \sum x_i = n\,\mu \Rightarrow \mu = \sum x_i/n$ (= otoskeskiarvo, kuten voi olettaa)

$\partial LL/\partial\sigma = \sum(-1/\sigma + 2(x_i-\mu)^2/2\,\sigma^3) = 0 \Rightarrow \sum(-\sigma^2 + (x_i-\mu)^2) = 0 \Rightarrow -7\sigma^2 + \sum(x_i-\mu)^2 = 0$
=> varianssi $\sigma^2 = \sum(x_i-\mu)^2/n$, jonka neliöjuuri on keskihajonta σ

Esimerkissä n = 7 ja $\sum x_i = 1262$ => keskiarvo μ = 1262/7 = 180.3. Keskihajonnaksi tulee vastaavasti 4.0.

Tässäkään esimerkissä MLE metodista ei ole varsinaista hyötyä, tulokset olisi saatu muutenkin helpommalla tavalla:
”Parhaat estimaatit populaatiokeskiarvolle, ja – keskihajonnalle ovat otoskeskiarvo ja otoskeskihajonta”

J.T. 2024

Esim4. Logistinen regressio

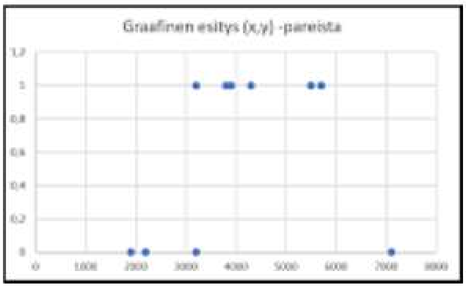
- yksi koneoppimisen perusmenetelmiä.

Riippumaton muuttuja on välimatka-asteikollinen ja riippuva muuttuja on 0,1 -muuttuja. .
Logistisen regression menetelmässä muodostetaan datan perusteella funktion p(x), jonka arvojoukko on [0, ∞[.   Funktion p(x) arvo = todennäköisyys, että selitettävä muuttuja saa arvon 1 kyseisellä x:n arvolla.

Esim4. Selittävä muuttuja x = henkilön kk-tulot, selitettävä muuttuja y on 0,1 -muuttuja ”harkitsee sähköauton ostoa”. Alla on dataa 10 henkilön mitatuista (x,P) -arvoista.

KK-tulot = X	Harkitsen sähköautoa = P
2200	0
3200	1
1900	0
5500	1
4300	1
3800	0
7100	1
5700	1
3200	0
3900	1

Datan graafinen esitys



Havaintoaineisto pistepareina:
Kk -palkka muuttujan x arvot muunnettu yksikköön k€

```
data = [[2.2, 0], [3.2, 1], [1.9, 0], [5.5, 1], [4.3, 1], [3.8, 0], [7.1, 1],
[5.7, 1], [3.2, 0], [3.9, 1]];
```

Logistisen regressioanalyysin tuloksena on sigmoidifunktio, joka esittää todennäköisyyttä sähköauton ostamiselle kk-tulojen funktiona.

Ratkaisun vaiheet.

1. Muuttujan vaihdos

Logistinen regression on sukua lineaariselle regressiolle. Erona on se, että selitettävällä muuttujalla P on vain kaksi mahdollista arvoa: 0 ja 1, kun tavallisessa lineaarisessa regressiossa arvoalue on kaikki reaaliluvut.

J.T. 2024

Logistisessa regressiossa muuttujalle P tehdään muunnos

$$y = ln(odds(p)) = ln(\frac{p}{1-p})$$

Funktiota odds(p) = p/(1-p) sanotaan vedonlyöntikerroinfunktioksi eli riskifunktioksi. Esim. jos hevosen voiton todennäköisyys p on 0.75, odds funktio antaa kertoimeksi 0.75/(1-0.25) = 3 eli vedonlyöntikielellä 3:1.

Logistisessa regressiossa todennäköisyysmuuttujan p arvot muunnetaan riskifunktion odds luonnolliseksi logaritmiksi.

Muunnoksessa arvo p = 0 muuttuu y – arvoksi ln(0/1) = -∞
ja p = 1 muuttuu y-arvoksi ln(1/0) = ∞.

Datataulukko muuntuu muotoon

KK-tulot(k€) = X	y
2.2	-∞
3.2	∞
1.9	-∞
5.5	∞
4.3	∞
3.8	-∞
7.1	∞
5.7	∞
3.2	-∞
3.9	∞

Muunnettuun dataan ei muunnoksen jälkeenkään voida soveltaa perinteistä lineaarista regressiomallia y = a x + b, jossa minimioidaan neliösumma $\sum(axi + b - yi)^2$. Syynä on se, että arvot y_i ovat joko -∞ tai ∞ .

2. Tehdään muuttujalle y käänteismuunnos y - > p ja palataan alkuperäiseen todennäköisyysfunktioon p(x):

$y = ln(p/(1-p)) \Rightarrow e^y = p/(1-p) \Rightarrow e^y - p\, e^y = p \Rightarrow e^y = (1+e^y)p$
$\Rightarrow p = ey/(1+ey) \Rightarrow$ $p = 1/(1+e^{-y})$

Sijoittamalla lineaarinen malli y = a x + b, saadaan lopputuloksena kaava

$$p(x) = \frac{1}{1 - e^{-ax-b}}$$

3. Mallin parametrit määritetään käyttämällä suurimman uskottavuuden menetelmää MLE ja alkuperäistä dataa.

Funktion p(x) tulkinnasta seuraa, että todennäköisyys sille, että selitettävä muuttuja p saa arvon 1 on p(x) ja todennäköisyys sille, että selitettävä muuttuja p saa arvon 0 on 1 – p(x).

Likelihood funktio on datataulukon perusteella seuraavanlainen.

$$L = (1 - \frac{1}{1 - e^{-2.2a-b}}) * \frac{1}{1 - e^{-3.2a-b}} * \ldots * \frac{1}{1 - e^{-3.9a-b}}$$

L on muotoa, jossa kannattaa siirtyä käyttämään logaritmista funktiota LL.

$$LL = ln(1 - \frac{1}{1 - e^{-2.2a-b}}) + ln\frac{1}{1 - e^{-3.2a-b}} + \ldots + ln\frac{1}{1 - e^{-3.9a-b}})$$

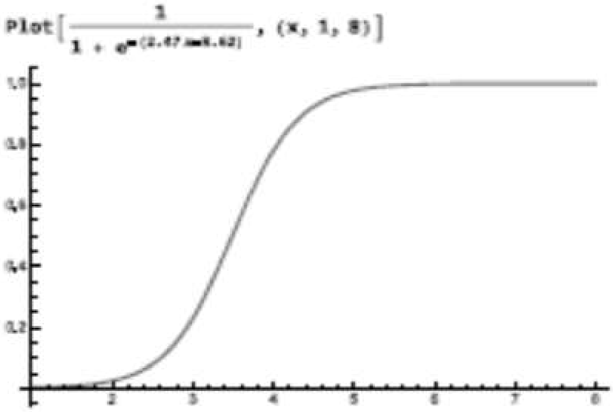
Minimi löytyy kohdasta, jossa funktion LL osittaisderivaatat sekä parametrin a , että parametrin b suhteen saavat arvon 0. Käsin minimiarvoa on vaikea, jopa mahdotonta määrittää. On käytettävä iteratiivista menetelmää (Gradient Descent) menetelmää tai hyvän laskinohjelmiston maximize toimintoa.

Mathematica – ohjelman komento

$$Maximize[\, ln(1 - \frac{1}{1 - e^{-2.2a-b}}) + ln\frac{1}{1 - e^{-3.2a-b}} + \ldots + ln\frac{1}{1 - e^{-3.9a-b}})]$$

antaa maximikohdan parameterille arvot a = 2.47 ja b = - 8.62.

Todennäköisyyttä kuvaava sigmoidi p(x) näyttää seuraavalta



Kuvan perusteella sähköauton ostamisen todennäköisyys on yli 50%, kun tulot ylittävät 3.5 k€/kk

Alla on sama tehtävä suoritettu Mathematica- ohjelmiston valmisfunktiolla **LogitModelFit**. Tulos on sama kuin yllä kuvatussa manuaaliratkaisussa.

```
malli = LogitModelFit[data, x, x]
```

FittedModel[$\frac{1}{1+e^{8.62413-2.47077x}}$]

Malli funktiomuodossa

```
Normal[malli]
```

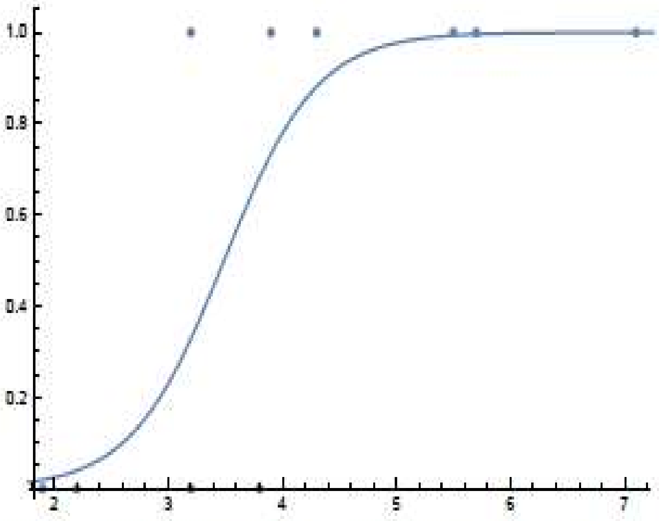
$$\frac{1}{1+e^{8.62413-2.47077x}}$$

Lasketaan mallia käyttäen todennäköisyys kohdassa x =3.7

```
malli[3.7]
0.626618
```

Kuva alkuperäisestä datasta ja siihen sovitetusta sigmoidista.

```
Show[ListPlot[data], Plot[malli[x], {x, 1.0, 7.5}]]
```



Lasketaan x:n arvo, jossa todennäköisyys ylittää 0.5.

```
Solve[malli[x] == 0.5, x]
{{x → 3.49046}}
```

Tulkinta: Jos kk-tulot ylittävät 3500 Euroa, on todennäköistä, että henkilö harkitsee sähköauton

LIITE1: Esimerkeissä käytettyjä laskusääntöjä:

$e^a e^b = e^{a+b}$

$\ln(x*y) = \ln(x) + \ln(y)$
$\ln(x/y) = \ln(x) – \ln(y)$
$\ln(x^r) = r \ln(x)$

Derivointikaavoja:

$Dx^n = n x^{n-1}$
$D(1/x^n) = -n/x^{n+1}$
$De^x = e^x$
$De^{ax} = a e^{ax}$
$D\ln(x) = 1/x$