Lojistik Regresyon Katsayi Hesabi

1 Giris

Asagida verilen pseudo code'u kullanarak lojistik regresyon icin gerekli olan katsayilari bulabilirsiniz. Note edin: bu code *gradient descent* methoduna dayanmaktadir. Bu method iteratif (yinelemeli) bir methodtur. Ilk basta verilen baslangic katsayi degerleri her bir iterasyonda gelisir. Iterasyon tamamlandiginda umut edilirki vardigimiz katsayilar baslangic katsayilarina gore daha anlamlidir, yani lojistik regresyonda daha iyi sonuc verir. Burada iterasyon sayisi methodun bir paremetresidir. Biz bu paremetreyi 100 olarak sabit tutacagiz. Yani for loop 100 defa donecek. Bu methotun bir baska parametresi adim buyuklugudur (step size). Biz bu paremetreyi de sabitleyecegiz ve 0.01 olarak alacagiz.

Bu method uretecegimiz katsayilar icin bir baslangic degerine ihtiyac duyar. Biz bu aktivitelere denk gelecek katsayilarin tamaminin baslangic degerini 1 olarak alacagiz. Yani $\beta_1 = \beta_2 = \beta_3 = ... = \beta_{15} = 1$. Tum bu katsayilara ek olarak bir de β_0 katsayisina ihtiyacimiz var. Bu katsayinin baslangic degerini de 1 olarak alacagiz: $\beta_0 = 1$. Sonuc olarak katsayilar vektorunu $\beta = (\beta_0, \beta_1, \beta_2, ..., \beta_{15})$ olarak gosterecegiz ve bunlarin her birinin baslangic degerini 1 alacagiz.

2 Datanin Matematiksel Olarak Ifadesi

Lojistik regresyonu uygulayacagimiz datayi (egitim setini)

$$D = \{(x^i, y^i) | x^i \in \{0, 1, 2, ..., 9, 10\}^{15}, y^i \in \{0, 1\}, i \in \{1, 2, ..., N\}\}$$

seklinde verelim. Demekki bu datada N tane kisi var ve i. kisi (x^i, y^i) ikilisi ile gosteriliyor. Burada x^i, i. kisinin profilini olusturuyor. Yani x^i, 15 uzunlugunda bir vektortur, cunku hatirlarsaniz 15 tane aktivite vardi. Ve biz her bir aktiviteyi $0-10$ arasi oylamistik. Hepsi birlikte dusunuldugunde i. kisinin profiline karsilik gelen x^i, 15 uzunlugunda bir vektortur ve bu vektorun elemanlari $0 - 10$ arasi deger alir. Bu i. kisinin etiketi olan y^i, eger i. kisi sorgulanan kisinin arkadasi ise 1, degilse 0 degerini alir. Son olarak N, sorgulanan kisinin toplam arkadas sayisi ile bu kisinin arkadasi olmayan kisilerin toplam sayisinin yarisinin toplamindan olusur.

Yukarida bahsettigimiz gibi x^i, i. kisinin profilini gosterir ve bu kisinin 15 tane aktiviteye verdigi oylardan olusur. Ozel olarak i. kisinin j. aktivite icin verdigi oyu x_j^i ile gosterecegiz. Su halde i, $1-N$ arasi bir deger, j, $1-15$ arasi bir deger ve x_j^i, $1-10$ arasi bir deger alir.

3 Arkadaslik Olasiligin Hesaplanmasi

Elimizde $\beta = (\beta_0, \beta_1, ..., \beta_{15})$ katsayilar vektoru ve bir kisiye ait x ($x \in \{0, 1, 2, ..., 9, 10\}^{15}$) profili varken bu kisinin sorgulanan kisi ile arkadas olma olasiligini

$$P(y=1|x,\beta) = h_\beta(x) = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+...+\beta_{15}x_{15})}}$$

fonksiyonu ile hesaplayacagiz.

4 Lojistik Regresyon Katsayilarini Veren Algorithma

Algorithm 1 Gradient Descent ile Lojistik Regresyon Katsayilarinin Bulunmasi

Giris: Data=D, $maxIterSayisi = 100$, $stepSize = 0.01$,
baslangic katsayilari vektoru: $\beta = (1, 1, ..., 1)$.
Cikis: β katsayilar vektoru

for $t = 1 : maxIterSayisi$ **do**
$$\beta'_0 = \beta_0 - stepSize\left(\frac{1}{N}\sum_{i=1}^{N}\left(h_\beta(x^i) - y^i\right)\right)$$ // β_0'i ayri olarak hesapliyoruz.
for $j = 1 : 15$ **do**
$$\beta'_j = \beta_j - stepSize\left(\frac{1}{N}\sum_{i=1}^{N}\left(h_\beta(x^i) - y^i\right)x_j^i\right)$$
end for
$(\beta_0, \beta_1, \beta_2, ..., \beta_{15}) \leftarrow (\beta'_0, \beta'_1, \beta'_2, ..., \beta'_{15})$ // β lari ayni anda guncelliyoruz.
end for// iterasyon sonu
return $\beta = (\beta_0, \beta_1, ..., \beta_{15})$

Yukaridaki algorithmayi egitim seti uzerine uygulayarak β katsayilarini hesaplaybilirsiniz. Daha sonra egitim setine dahil etmedigimiz sorgulanan kisi ile arkadas olmayan kisilerin diger yarisindaki her bir kisi icin bu katsayilari ve yukarida verilen olasilik hesaplamasini kullanarak bu kisilerin arkadas olma olasiligini hesaplayabilirsiniz. En son olarak bu kisileri aldiklari olasiklarlara gore siralayip, arkadas olma olasilgi en yuksek ilk 10 kisiyi cikti olarak verin.