Multinomial Regresyon

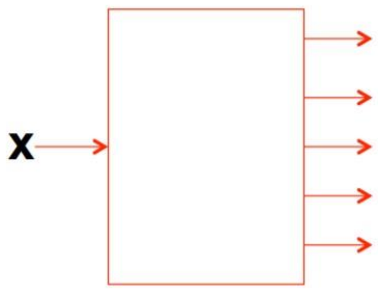
Bu gibi durumlarda, softmax regresyon olarak da adlandırılan multinomial lojistik regresyon kullanıyoruz. n (veya tarihsel olarak maxent sınıflandırıcı). Çok terimli lojistik regresyonda, hedef y ikiden fazla sınıfta değişen bir değişkendir; y'nin her potansiyel sınıf $c \in C$, p (y = c | x) içinde olma olasılığını bilmek istiyoruz.

Multinomial lojistik sınıflandırıcı, p (y = c | x) olasılığını hesaplamak için softmax fonksiyonu olarak adlandırılan sigmoidin bir genellemesini kullanır. Softmax işlevi, k keyfi değerden bir z = [z1, z2, ..., zk] vektörü alır ve bunları (0,1) aralığındaki her değer ve tüm değerlerin toplamı 1 olacak şekilde bir olasılık dağılımına eşler Sigmoid gibi, üstel bir fonksiyondur. K boyutsallığının bir z vektörü için softmax şu şekilde tanımlanır:

$$\text{softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{k} e^{z_j}} \quad 1 \leq i \leq k$$

Bu nedenle, z = [z1, z2, ..., zk] giriş vektörünün softmax'ı bir vektörün kendisidir:

$$\text{softmax}(z) = \left[\frac{e^{z_1}}{\sum_{i=1}^{k} e^{z_i}}, \frac{e^{z_2}}{\sum_{i=1}^{k} e^{z_i}}, \ldots, \frac{e^{z_k}}{\sum_{i=1}^{k} e^{z_i}} \right]$$



$$\mathrm{f}(\mathbf{x}) = \begin{bmatrix} P(y=1|\mathbf{x}) \\ P(y=2|\mathbf{x}) \\ \cdots \\ P(y=K|\mathbf{x}) \end{bmatrix} = \frac{1}{\sum_{j=1}^{K} \exp\left[\mathbf{w}_j^T \mathbf{x}\right]} \begin{bmatrix} \exp\left[\mathbf{w}_1^T \mathbf{x}\right] \\ \exp\left[\mathbf{w}_2^T \mathbf{x}\right] \\ .. \\ \exp\left[\mathbf{w}_k^T \mathbf{x}\right] \end{bmatrix}$$

$\sum_{i=1}^{k} e^{z_i}$ tüm değerleri olasılıklara normalleştirmek için kullanılır. Örneğin bir vektör

Payda
verildiğinde:
z = [0.6,1.1,−1.5,1.2,3.2,−1.1]

sonuç softmax (z):

[0.055,0.090,0.0067,0.10,0.74,0.010]

Yine sigmoid gibi, softmax'ın girdisi ağırlık vektörü w ile bir girdi vektörü x (artı bir sapma) arasındaki iç çarpım olacaktır. Ama şimdi her K sınıfı için ayrı ağırlık vektörlerine (ve sapmaya) ihtiyacımız olacak.

$$p(y=c|x) = \frac{e^{w_c \cdot x + b_c}}{\sum_{j=1}^{k} e^{w_j \cdot x + b_j}}$$

Sigmoid gibi softmax da değerleri 0 veya 1'e doğru bastırma özelliğine sahiptir. Dolayısıyla, eğer girdilerden biri diğerlerinden daha büyükse, olasılığını 1'e doğru itme ve daha küçük girdilerin olasılıklarını bastırma eğiliminde olacaktır.

Çok Terimli Logistic Regresyondaki Özellikler(Features in Multinomial Logistic Regression)

Çok sınıflı sınıflandırma için girdi özelliklerinin hem gözlem x'in hem de aday çıktı sınıfı c'nin bir işlevi olması gerekir. Bu nedenle, xi, fi veya fi (x) gösterimi yerine, özellikleri tartışırken fi (c, x) gösterimini kullanacağız, yani belirli bir x gözlemi için belirli bir c sınıfı için özellik i'dir. İkili sınıflandırmada, y = 1'e işaret eden bir özellik üzerindeki pozitif ağırlık ve y = 0'a doğru negatif bir ağırlık, ancak çok sınıflı sınıflandırmada bir özellik, tek bir sınıf lehine veya aleyhine kanıt olabilir. Hem x gözleminin hem de c sınıfının işlevleri olan özelliklerin belki de sezgisel olmayan kullanımını anlamaya yardımcı olması için birkaç NLP görevi için bazı örnek özelliklere bakalım. Metin sınıflandırması yaptığımızı varsayalım ve ikili sınıflandırma yerine görevimiz 3 sınıftan birini +, - veya 0 (nötr) bir belgeye atamaktır. Artık ünlem işaretleriyle ilgili bir özellik 0 belge için eksi ağırlığa ve + veya - belgeler için artı ağırlığa sahip olabilir:

Var	Definition	Wt
$f_1(0,x)$	$\begin{cases} 1 & \text{if "!"} \in \text{doc} \\ 0 & \text{otherwise} \end{cases}$	-4.5
$f_1(+,x)$	$\begin{cases} 1 & \text{if "!"} \in \text{doc} \\ 0 & \text{otherwise} \end{cases}$	2.6
$f_1(-,x)$	$\begin{cases} 1 & \text{if "!"} \in \text{doc} \\ 0 & \text{otherwise} \end{cases}$	1.3

Çok Terimli Lojistik Regresyonda Öğrenme(Learning in Multinomial Logistic Regression)

Çok terimli lojistik regresyon, ikili lojistik regresyondan biraz farklı bir kayıp fonksiyonuna sahiptir çünkü sigmoid sınıflandırıcı yerine softmax kullanır. Tek bir örnek x için kayıp işlevi, K çıktı sınıflarının günlüklerinin toplamıdır:

$$\begin{aligned} L_{CE}(\hat{y}, y) &= -\sum_{k=1}^{K} 1\{y=k\} \log p(y=k|x) \\ &= -\sum_{k=1}^{K} 1\{y=k\} \log \frac{e^{w_k \cdot x + b_k}}{\sum_{j=1}^{K} e^{w_j \cdot x + b_j}} \end{aligned}$$

Bu, parantez içindeki koşul doğruysa 1, aksi halde 0 olarak değerlendiren 1 {} işlevini kullanır. Burada türetmeyi göstermemekle birlikte, tek bir örneğin gradyanının lojistik regresyon gradyanına çok benzediği ortaya çıkmaktadır. Gerçek k sınıfı değeri (1'dir) ile sınıflandırıcının k sınıfı için çıkardığı olasılık arasındaki farktır ve xk girdisinin değerine göre ağırlıklandırılmıştır:

$$\begin{aligned} \frac{\partial L_{CE}}{\partial w_k} &= -(1\{y=k\} - p(y=k|x))x_k \\ &= -\left(1\{y=k\} - \frac{e^{w_k \cdot x + b_k}}{\sum_{j=1}^{K} e^{w_j \cdot x + b_j}}\right) x_k \end{aligned}$$

Simple

Logistic Regresyon nedir?

İsterseniz basitçe örneklendirerek açıklayalım. İki değerli (erkek / kadın, ölü / diri, vb.) Ve bir ölçüm değişkenine sahip bir nominal değişkeniniz olduğunda basit lojistik regresyon kullanın. Nominal değişken bağımlı değişkendir ve ölçüm değişkeni bağımsız değişkendir.

Ordinal Logistic Regresyon nedir?

Sıralı regresyon, bağımlı değişkeni "sıralı" çoklu kategoriler ve bağımsız değişkenlerle tahmin etmek için kullanılır. Başka bir deyişle, bağımlı değişkenlerin (birden fazla sıralı seviyeye sahip) bir veya daha fazla bağımsız değişkenle etkileşimini kolaylaştırmak için kullanılır.