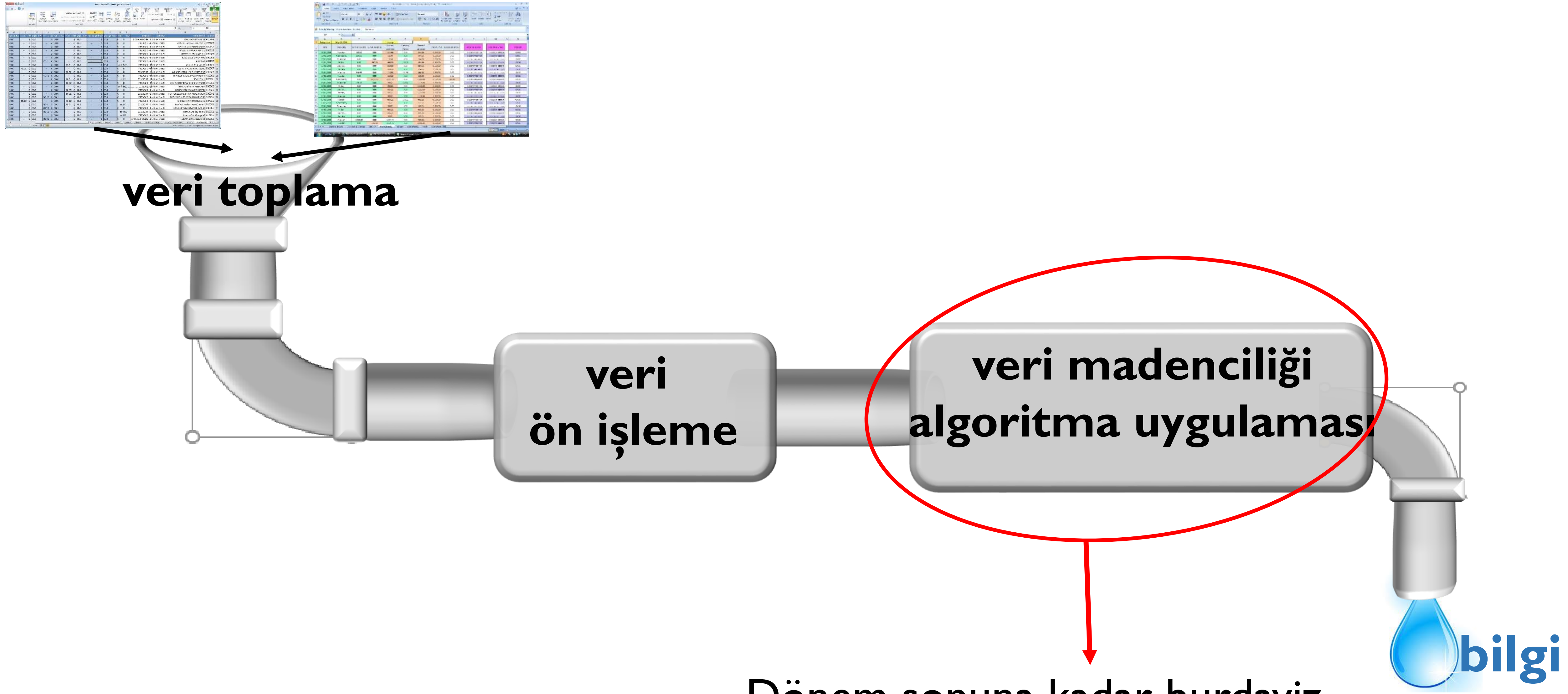
VERİ MADENCİLİĞİ

Fırat İsmailoğlu, PhD

Naïve Bayes ve Lojistik Regresyon



Dönem sonuna kadar burdayiz.

Bayes Teoremi

X ve Y iki rastgele degisken (random variable) olsun. Bu durumda X ve Y'nin bileşik olasiligi (joint probability) $P(X, Y)$ iki sekilde hesaplanabilir:

$$P(X, Y) = P(Y|X) \cdot P(X)$$

$$P(X, Y) = P(X|Y) \cdot P(Y)$$

Bu iki denklemi birbirine eşitlersek:

$$P(Y|X) \cdot P(X) = P(X|Y) \cdot P(Y)$$

olur. $P(Y|X)$ yalniz birakilirsa Bayes toerimi elde edilir :

$$\boxed{P(Y|X) = \frac{P(X|Y) \cdot P(Y)}{P(X)}}$$

$P(Y = y|X = x)$: X değişkeni x iken Y degiskeninin y olma olasiligi.

Ayrica bu ifadeye '*şartlı olasılık*' da denir. $X = x$ olmasi şartiyla Y'nin y olma olasılığı..

ör.Bir klinikte yapılan kanser testi gerçekte kanser olan hastaların %98'inde pozitif olarak sonuç veriyor. Bu test, gerçekte kanser olamayan hastaların %97'sinde negatif olarak sonuç veriyor. Ayrıca toplumdaki kişilerin 0.008'nin kanser olduğu biliniyor.

Buna gore kanser testi pozitif çıkmış bir kişinin gerçekte kanser olma olasılığı nedir?

Çözum:

Verilenler:

$P(kanser) = 0.008$ //kişi hakkında hiçbir bilgi yokken kişinin kanser olma olasiligi
$P(\neg kanser) = 0.992$

$P(test\ pozitif|kanser) = 0.98$ //kişi kanser iken testinin pozitif olma olasiligi
$P(test\ negatif|kanser) = 0.02$

$P(test\ negatif|\neg kanser) = 0.97$ //kişi kanser değilken testinin negatif çikma olasiligi
$P(test\ pozitif|\neg kanser) = 0.03$

İstenen:

$P(kanser|test\ pozitif) =?$

Bayes teoremi ile:

$$P(kanser|test\ pozitif) = \frac{P(test\ pozitif|kanser) \times P(kanser)}{P(test\ pozitif)}$$

Burada bilmediğimiz şey $P(test\ pozitif)$. Yani kişinin kanser olup olmadigina bagli olmaksizin testinin pozitif çıkma olasiligi.

Bir kişi ya kanserdir yada kanser değildir. O halde testin pozitif cikmasi kanserli kişilerde ve kanser olmayan kişilerde gorulur.

Bir kişinin kanser olmasi ve testinin pozitif cikmasi: $P(test\ pozitif, kanser)$.

Bir kişinin kanser olmamasi ve testinin pozitif cikmasi: $P(test\ pozitif, \neg kanser)$.

$$P(test\ pozitif) = P(test\ pozitif, kanser) + P(test\ pozitif, \neg kanser)$$
$$= P(test\ pozitif|kanser) \cdot P(kanser) + P(test\ pozitif|\neg kanser) \cdot P(\neg kanser)$$
$$= 0.98 \times 0.008 + 0.03 \times 0.992$$
$$= 0.0376$$

$$P(kanser|test\ pozitif) = \frac{0.98 \times 0.008}{0.0376} = 0.2$$

Su halde $P(\neg kanser|test\ pozitif) = 1 - 0.2 = 0.8$

Testi pozitif çikan kişinin kanser olmama olasiligi cok daha yüksektir. O halde bu test pek güvenilir değildir!

ör. Iki futbol takimi duşunun: T1 ve T2. Varsayalimki T1, T2 ile yaptigi maclarin %65'ni yenmiş olsun. T1'in T2'yi yendiği maclarin %70'i kendi sahasinda gerçekleşmiş olsun. Öte yandan T2'nin T1'i yendiği maclarin %75'i kendi sahasinda gerçeklemiş olsun. Eğer gelecek maç T1'in sahasinda olacaksa bu maçı hangi takimin kazanmasi daha olasidir?

Çözüm:

$P(T1'in\ kazanmasi) = 0.65$

$P(T2'nin\ kazanmasi) = 0.35$

$P(T1'in\ sahasi|T1'in\ kazanmasi) = 0.7$

$P(T2'nin\ sahasi|T2'nin\ kazanmasi) = 0.75$

$P(T1'in\ sahasi|T2'nin\ kazanmasi) = 0.25$

$P(T1'nin\ kazanmasi|T1'nin\ sahasi)$ Vs $P(T2'nin\ kazanmasi|T1'nin\ sahasi)$

$$P(T1'nin\ kazanmasi|T1'nin\ sahasi) = \frac{P(T1'nin\ sahasi|T1'nin\ kazanmasi) \times P(T1'in\ kazanmasi)}{P(T1'in\ sahasi)}$$

$$P(T2'nin\ kazanmasi|T1'nin\ sahasi) = \frac{P(T1'nin\ sahasi|T2'nin\ kazanmasi) \times P(T2'in\ kazanmasi)}{P(T1'in\ sahasi)}$$

Yukaridaki her iki olasilik hesaplamasinda da paydalar aynidir. Amacimiz yalnızca bu olasilikları sıralamak olduğundan paydayi hesaplamaya gerek yoktur.

(Yinede hesaplanmak istenirse:

$P(T1'in\ sahasi)$= $P(T1'nin\ sahasi|T1'nin\ kazanmasi) \times P(T1'in\ kazanmasi) +$
$P(T1'nin\ sahasi|T2'nin\ kazanmasi) \times P(T2'in\ kazanmasi)$)

$$P(T1'nin\ kazanmasi|T1'nin\ sahasi) = \frac{0.7 \times 0.65}{P(T1'in\ sahasi)} = \frac{0.455}{P(T1'in\ sahasi)}$$

$$P(T2'nin\ kazanmasi|T1'nin\ sahasi) = \frac{0.25 \times 0.35}{P(T1'in\ sahasi)} = \frac{0.08}{P(T1'in\ sahasi)}$$

Ilk olasillikta pay daha büyük olduğundan T1'in sahasinda oynanacak maçta T1'in kazanma olasiligi daha yüksektir.

$$P(T1'nin\ kazanmasi|T1'nin\ sahasi) > P(T2'nin\ kazanmasi|T1'nin\ sahasi)$$

Elimizde bir sonuç (gözlem) varken biz bu sonucun ortaya çikmasini sağlayabilecek birçok nedenden en olası olanını bulmak isteriz. Yani olabilecek her neden için $P(neden|sonuç)$ olasiliğini hesaplariz.

Örnegin gözlemimiz 'öksürük' olsun ve biz 'grip'ten şüphelenelim. Bu durumda $P(grip|öksürük)$ olasiligini hesaplamamiz gerekir.

Öte yandan genellikle $P(sonuç|neden)$'i hesaplamak daha kolaydır. Yani elimizdeki sonuca yol açabilecek nedenin bu sonuca yol açma olasiligini hesaplamak daha kolaydır.

Örnek olarak $P(öksürük|grip)$ olasılığı.

Bayes teoremi bize $P(neden|sonuç)$'tan $P(sonuç\ |neden)$'e gitmemizi sağlar:

$$\boxed{P(neden|sonuç) = \frac{P(sonuç|neden) \times P(neden)}{P(sonuç)}}$$

Öksürük şikayeti (bulgusu) olan hastada bu şikayete yol açabilecek nedenler grip ve zatürre olsun. Amacımız $P(grip\ |öksürük)$ ve $P(zatürre\ |öksürük)$ şartlı olasılıklarını karşılaştırmaktır. Bayes teoremi ile:

$$P(grip|öksürük) = \frac{P(öksürük|grip) \times P(grip)}{P(öksürük)}$$

$$P(zatürre|öksürük) = \frac{P(öksürük|zatürre) \times P(zatürre)}{P(öksürük)}$$

Her iki olasılık hesabında da payda aynıdır. O halde bu iki olasiligi karşilaştirirken yalnız payları göz önüne alabiliriz.

Genellersek $P(sonuç)$ tüm nedenler için aynı olduğundan, $P(neden|sonuç)$; $P(sonuç|neden) \times P(neden)$ ile orantilidir:

$$\boxed{P(neden|sonuç) \propto P(sonuç|neden) \times P(neden)}$$

Şartlı Bağımsızlık (Conditional Independence)

Ahmet ve Mehmet iki kardeş olsun. Bu kardeşlerde belirli bir genetik hastaligin gorulme olasılıklarına bakalım.

Bu genetik bir hastalik oldugundan Ahmet'in ve Mehmet'in hasta olma olasiliklari birbirinden ayrı düşünülemez; birbirinden bagimsiz degildir. O halde

$$P(Ahmet\ hasta, Mehmet\ hasta) \neq P(Ahmet\ hasta) \times P(Mehmet\ hasta).$$

Fakat Mehmet'in evlatlık olduğu bilinirse; Ahmet'in ve Mehmet'in bu genetik hastaligi tasima olasliklari birbirinden bagimsiz olur. O halde

$$P(Ahmet\ hasta, Mehmet\ hasta | Mehmet\ evlatlik) \\ = P(Ahmet\ hasta | Mehmet\ evlatlik) \times P(Mehmet\ hasta | Mehmet\ evlatlik).$$

Mehmet'in evlatlik olmasi sartiyla Ahmet in ve Mehmetin bu hastaligi tasima olasiliklari birbirinden bağımsızdır.

Şartlı Bağımsızlık (Conditional Independence)

Genel formül:

$X_1, X_2, \ldots, X_n$ rastgele değişkenler olsun. Eğer

$$P(X_1, X_2, \ldots, X_n|Z) = P(X_1|Z) \times P(X_2|Z) \times \cdots \times P(X_n|Z)$$

oluyorsa Z'nin varliginda; $X_1, X_2, \ldots, X_n$ rastgele değişkenlerinin gerçekleşme olasiliklari biribirinden bagimsizdir.

Bayes Teoreminin Sınıflandırmada Kullanılması

Bir test örneği $x = (x_1, x_2, \ldots, x_n)$ seklinde verilsin. Bu örnegi $y_1, \ldots, y_k$ sınıflarından en olası olanına eşlemek istiyoruz.

Burada her y_i ($i \in \{1, \ldots, k\}$) sınıfı için $P(y_i|x)$ olasılığını hesaplarız. Bu olasılık elimizde x varken bunun y_i sınıfına ait olma olaslığıdır.

Test örnegini en yüksek $P(y_i|x)$ olasliliğina sahip sınıfa eşleriz.Formal olarak:

$$\boxed{\hat{y} = \underset{i \in \{1, \ldots, k\}}{argmax}\, P(y_i|x)}$$

$P(y_i|x)$ olasılığını Bayes teoremi kullanarak aşağıdaki gibi hesaplayabiliriz:

$$P(y_i|x) = \frac{P(x|y_i) \times P(y_i)}{P(x)}$$

Amacımız $P(y_i|x)$ olasılıklarını kıyaslarak en yüksek olasılığa sahip sınıfı seçmekti. $P(y_i|x)$ hesaplamasında bütün sınıflar için (her $i \in \{1, \dots, k\}$ için) payda $(P(x))$ aynı olduğundan sınıflandırma yaparken $P(x)$'i hesaplamiyoruz.

Bu durumda $P(y_i|x)$ hesaplamasinda yalnızca $P(x|y_i) \times P(y_i)$ çarpımını dikkate alacağiz:

$$P(y_i|x) \propto P(x|y_i) \times P(y_i)$$

$P(y_i)$: eğitim setinde y_i sınıfının görülme olasılığı. Örnegin eğitim setindeki örneklerin %20'sinin sınıfı grip olsun. Bu durumda $P(grip) = 0.2$ olur.

$P(x|y_i)$: y_i sınıfında x örneğinin görülme olasılığı.

Örneğin $x = (bulantı, yüksek\ ateş, burun\ tıkanıklığı), y_i = grip$ olsun.

$P(x|y_i)$: hasta grip iken bulantı, yüksek ateş ve burun tıkanıklığının görülme olasılığı.

Bulantı, yüksek ateş ve burun tıkanıklığının gripin varliginda gerçekleşme olasiliklarinin birbirinden bagimsiz olduğunu varsayalim. Yani bu belirtileri şartli bagimsiz varsayalım.

Bu durumda:

$$P(x|y_i) = P(bulantı, yüksek\ ateş, burun\ tıkanıklığı|grip)$$
$$= P(bulantı|grip) \times P(yüksek\ ateş|grip) \times P(burun\ tıkanıklığı|grip).$$

Bu varsayımı formulize edersek:

$$P(x|y_i) = P(x_1|y_i) \times P(x_2|y_i) \times \cdots \times P(x_n|y_i)$$
$$= \prod_{j=1}^{n} P(x_j|y_i).$$

Amacımız $P(y_i|x)$ olasılıklarini karşılaştırmaktı. Gördükki bu olasilik $P(x|y_i) \times P(y_i)$ ile orantili. Ayrica sınıfın (y_i)'nin varlliginda $x_1, x_2, \ldots, x_n$ özelliklerinin biribirinden bagimsiz olduğunu varsaydık. Tüm bunları birleştirirsek:

$$\hat{y} = \underset{i \in \{1,\ldots,k\}}{argmax}\, P(y_i|x)$$

$$\hat{y} = \underset{i \in \{1,\ldots,k\}}{argmax}\, P(y_i) \times P(x|y_i)$$

$$\boxed{\hat{y} = \underset{i \in \{1,\ldots,k\}}{argmax}\, P(y_i) \times \prod_{j=1}^{n} P(x_j|y_i)}$$

Yukarıdaki kural ile test örneklerini sınıflandıran sınıflandırıcıya 'Naive Bayes Sınıflandırıcı' denir.

ör.

Deri	Renk	Büyüklük	Et	Sınıf
Tüylü	Kahverengi	Büyük	Sert	Güvenli
Tüylü	Yeşil	Büyük	Sert	Güvenli
Tüysüz	Kırmızı	Büyük	Yumuşak	Tehlikeli
Tüylü	Yeşil	Büyük	Yumuşak	Güvenli
Tüylü	Kırmızı	Küçük	Sert	Güvenli
Tüysüz	Kırmızı	Küçük	Sert	Güvenli
Tüysüz	Kahverengi	Küçük	Sert	Güvenli
Tüylü	Yeşil	Küçük	Yumuşak	Tehlikeli
Tüysüz	Yeşil	Küçük	Sert	Tehlikeli
Tüylü	Kırmızı	Büyük	Sert	Güvenli
Tüysüz	Kahverengi	Büyük	Yumuşak	Güvenli
Tüysüz	Yeşil	Küçük	Yumuşak	Tehlikeli
Tüylü	Kırmızı	Küçük	Yumuşak	Güvenli
Tüysüz	Kırmızı	Büyük	Sert	Tehlikeli
Tüysüz	Kırmızı	Küçük	Sert	Güvenli
Tüylü	Yeşil	Küçük	Sert	Tehlikeli

Eğitim Seti

Elimizde bu eğitim seti varken '*Tüylü, Kırmızı,Büyük,Yumuşak*' özelliklere sahip bir hayvanı yemek güvenli midir yoksa tehlikeli midir?

$P(güvenli|tüylü, kırmızı, büyük, yumuşak)$ Vs $P(tehlikeli|tüylü, kırmızı, büyük, yumuşak)$

$$P(güvenli|tüylü, kırmızı, büyük, yumuşak) \propto$$
$$P(tüylü, kırmızı, büyük, yumuşak|güvenli) \times P(güvenli)$$

Eğitim setindeki 16 hayvandan 10'u güvenli sınıfındadır. O halde $P(güvenli) = \frac{10}{16} = 0{,}62.$

Şartlı bağımsız varsayımı sayesinde:

$$P(tüylü, kırmızı, büyük, yumuşak|güvenli) =$$
$$= P(tüylü|güvenli) \times P(kırmızı|güvenli) \times P(büyük|güvenli) \times P(yumuşak|güvenli)$$

Deri	Renk	Büyüklük	Et	Sınıf
Tüylü	Kahverengi	Büyük	Sert	Güvenli
Tüylü	Yeşil	Büyük	Sert	Güvenli
Tüysüz	Kırmızı	Büyük	Yumuşak	Tehlikeli
Tüylü	Yeşil	Büyük	Yumuşak	Güvenli
Tüylü	Kırmızı	Küçük	Sert	Güvenli
Tüysüz	Kırmızı	Küçük	Sert	Güvenli
Tüysüz	Kahverengi	Küçük	Sert	Güvenli
Tüylü	Yeşil	Küçük	Yumuşak	Tehlikeli
Tüysüz	Yeşil	Küçük	Sert	Tehlikeli
Tüylü	Kırmızı	Büyük	Sert	Güvenli
Tüysüz	Kahverengi	Büyük	Yumuşak	Güvenli
Tüysüz	Yeşil	Küçük	Yumuşak	Tehlikeli
Tüylü	Kırmızı	Küçük	Yumuşak	Güvenli
Tüysüz	Kırmızı	Büyük	Sert	Tehlikeli
Tüysüz	Kırmızı	Küçük	Sert	Güvenli
Tüylü	Yeşil	Küçük	Sert	Tehlikeli

Güvenli sınıfındaki 10 hayvandan 6'sı tüylüdür. $P(tüylü|güvenli) = \frac{6}{10} = 0,6.$

Güvenli sınıfındaki 10 hayvandan 5'i kırmızıdır. $P(kırmızı|güvenli) = \frac{5}{10} = 0,5.$

Güvenli sınıfındaki 10 hayvandan 5'i büyüktür. $P(büyük|güvenli) = \frac{5}{10} = 0,5.$

Güvenli sınıfındaki 10 hayvandan 3'ü yumuşaktır. $P(yumuşak|güvenli) = \frac{3}{10} = 0,3.$

$$P(güvenli|tüylü, kırmızı, büyük, yumuşak) \propto$$
$$P(tüylü|güvenli) \times P(kırmızı|güvenli) \times P(büyük|güvenli) \times P(yumuşak|güvenli)) \times P(güvenli)$$
$$= 0{,}6 \times 0{,}5 \times 0{,}5 \times 0{,}3 \times 0{,}62 = 0{,}0279.$$

Benzer şekilde $P(tehlikeli|tüylü, kırmızı, büyük, yumuşak)$ yi hesaplayalim.

$$P(tehlikeli|tüylü, kırmızı, büyük, yumuşak) \propto$$
$$P(tüylü|tehlikeli) \times P(kırmızı|tehlikeli) \times P(büyük|tehlikeli) \times P(yumuşak|tehlikeli) \times P(tehlikeli)$$
$$= 0{,}33 \times 0{,}33 \times 0{,}33 \times 0{,}5 \times 0{,}38 = 0{,}006.$$

$$P(güvenli|tüylü, kırmızı, büyük, yumuşak) > P(tehlikeli|tüylü, kırmızı, büyük, yumuşak)$$

olduğundan test örnegini güvenli olarak siniflandiririz. Yani '*Tüylü, Kırmızı,Büyük,Yumuşak*' özelliklere sahip bir hayvanı yemek güvenlidir.

Naive Bayes Sınıflandırıcıda Dikkat Edilecek Hususlar

1. Bir değer bir sınıfta hiç görülmemiş olabilir.

Örnegin kahverengi bir hayvan Tehlikeli sınıfında hiç görülmemiştir (verilen eğitim setinde böyle bir örnek yoktur). Bu durumda $P(kahverengi|tehlikeli) = 0$ olur. Yani kahverengi bir hayvanin tehlikeli olma olasiligi direkt 0 olur. Diğer hiçbir özellik dikkate alinmaz. Bu ise çoğu kez hatta yapmamiza neden olur.

Basit bir çözüm olarak böyle bir durumda 0 yerine çok küçük bir olaslılık verebiliriz. Örneğin 0.0000001 gibi.

2. Genel olarak olasılıkları çarpmak iyi bir fikir değildir.

Bir veri setinde 10000 tane özellik oldugunu düşünün: $X_1, X_2, \dots, X_{10000}$.

$P(X_1|y_i) \times P(X_2|y_i) \times \cdots \times P(X_{10000}|y_i)$ çarpımı çok çok küçük bir sayı olabilir, örnegin 10^{-20000} gibi. Bu durumda kullandigimiz program bunu 0'a çevirebilir. Böylece test örneğini y_i sınıfına eşleştirme olasılığımız 0 olur.

Çözüm olarak çarpımın logunu alırız:

$$\boxed{\hat{y} = argmax\ log\left(P(y_i) \times \prod_{j=1}^{n} P(x_j|y_i)\right)}$$

Sayısal Özellik Varken Naive Bayes Sınıflandırıcı

Özellikler kategorik iken Naive Bayes sınıflandırıcı inşa etmek oldukça kolaydir. Fakat sayısal özellikler varken $P(x|y)$ olasılığini hesap etmek pek kolay degildir.

Bu durumda genel yaklasim *sayisal ozelliklerdeki değerlerin her bir sınıf için normal dagilima sahip olduğunu varsaymaktır.*

ör.

Ev Sahibi	Yıllık Gelir	Krediyi ödemiş mi?
Evet	125K	Evet
Hayır	100K	Hayır
Hayır	70K	Hayır
Evet	120K	Hayır
Hayır	105K	Evet
Hayır	60K	Hayır
Evet	220K	Evet

Eğitim seti yukaridaki gibi verilmisken ev *sahibi olmayan ve yıllık geliri* $100K$ olan bir test örneginin alacagi krediyi ödeme olasiligini ve odememe olasiligini karsilastiralim.

$$P(Evet|Ev\ sahibi\ degil, 100K) \text{ Vs } P(Hayır|Ev\ sahibi\ degil, 100K)$$

$$P(Evet|Ev\ sahibi\ degil, 100K) \propto P(Ev\ sahibi\ degil, 100K|Evet) \times P(Evet)$$

$$P(Ev\ sahibi\ degil, 100K|Evet) = P(Ev\ sahibi\ degil|Evet) \times P(100K|Evet) \times P(Evet)$$

Aldigi krediyi odeyen 3 kisi vardir. Bu uc kisinin birinin evi yoktur; ikisinin evi vardir.

$$P(Ev\ sahibi\ degil|Evet) = \frac{1}{3} = 0,33$$

$P(100K|Evet)$ olasiligini hesaplarken; krediyi ödeyen kisilerin gelirlerinin normal dagilima sahip oldugunu varsayacagiz. Yani krediyi odeyen kisilerin gelirleri:

$$125, 105,220$$

normal dagilima sahip olsun.

Bu dagilimin ortalamasi $\mu = 150$, standart sapması $\sigma = 61{,}4$

Bir x değerinin ortalamasi μ , standart sapmasi σ olan bir normal dagilim tarafindan üretilme olasiliği $P(x|\mu, \sigma)$ şu formülle hesaplanir:

$$P(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{(x-\mu)^2}{2\sigma^2}\right)}$$

Bu formulde $x = 100, \mu = 150$ ve $\sigma = 61{,}4$ degerleri yerlerine konursa:

$$P(100K|Evet) = P(100|150{,}61.4) = 0.0047$$

$$P(Ev\ sahibi\ degil, 100K|Evet) \propto 0.33 \times 0.0047 \times 0.42 = 0.00651$$

$$P(Ev\ sahibi\ degil, 100K|Hayır) \propto 0.75 \times 0.013 \times 0.57 = 0.056$$

O halde test örnegindeki kisinin alacagi krediyi odememe olasailigi daha yuksektir.

Lojistik Regresyon

Elimizde bir *ikili sınıflandırma* (yalnızca iki tane sınıfın olduğu sınıflandırma) eğitim seti olsun. Eğitim setindeki sınıflardan birini 0, diger sınıfı 1 olarak etiketleyelim. Böylece bağlı değişken sayısal (nümerik) değerler almış olur. Bu bize sınıflandırma problemini regresyon problemine dönüştürmemizi sağlar.

ör.

Tümör Büyüklüğü	Kanser?
0.3	Kanser değil
4.8	Kanser
5	Kanser
1.5	Kanser degil
0.7	Kanser degil
1	Kanser degil
5.2	Kanser

Sınıflandırma Problemi

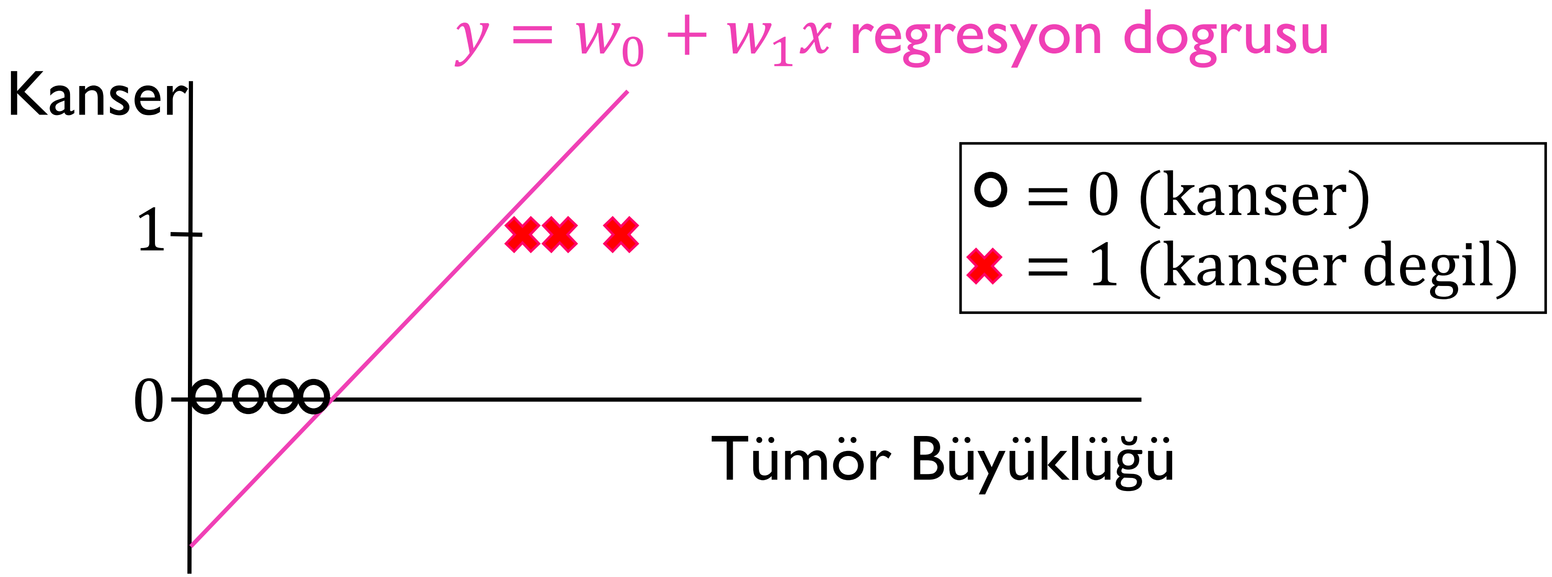
0/1 dönüştürmesi

Tümör Büyüklüğü	Kanser?
0.3	0
4.8	1
5	1
1.5	0
0.7	0
1	0
5.2	1

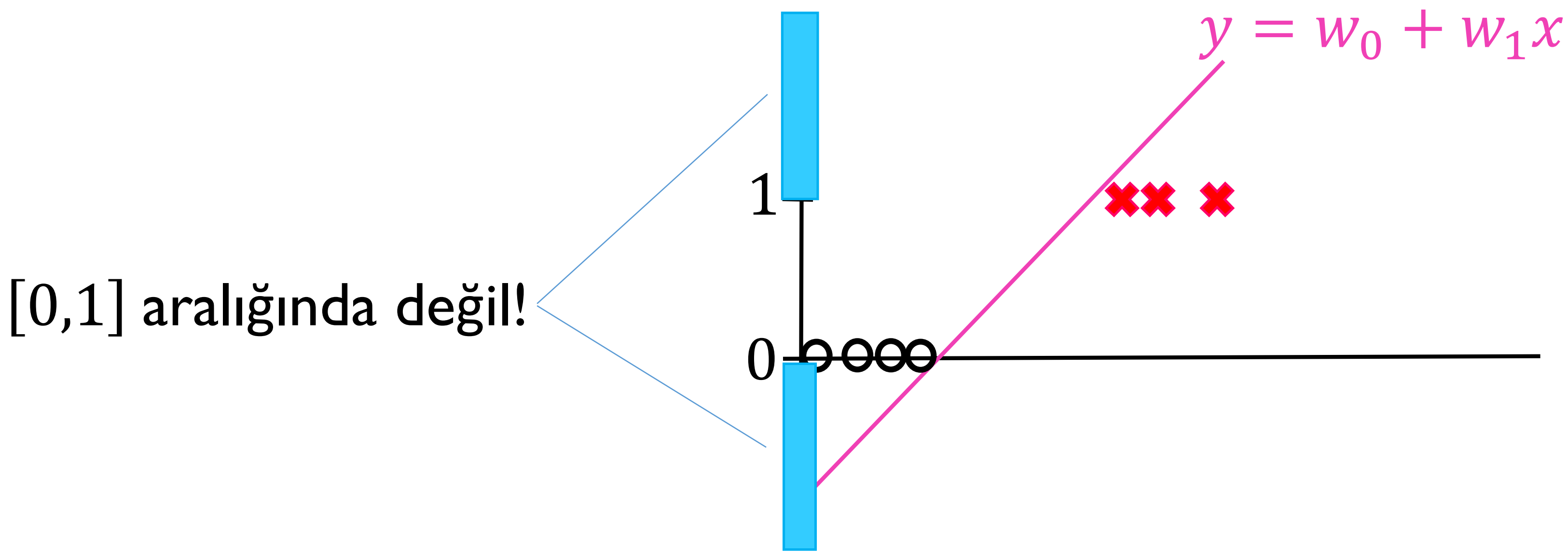
Regresyon Problemi

Lojistik Regresyon

Tümör Büyüklüğü	Kanser?
0.3	0
4.8	1
5	1
1.5	0
0.7	0
1	0
5.2	1



Verilen ikili sınıflandırma problemini regresyon problemine dönüştürsek ve klasik anlamda bir regresyon doğrusu oluşturursak ($y = w_0 + w_1 x$) elde edecegimiz y değerlerinin $[0,1]$ arasinda olmasini garanti edemeyiz.

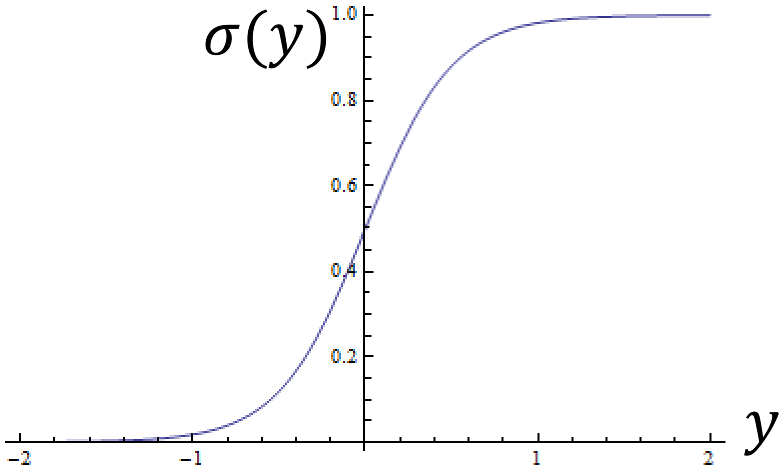


$y = w_0 + w_1 x$ değerlerini 0-1 aralığına sığdırmak için bu değerleri lojistik (sigmoid) fonksiyonuna sokarız. Lojistik fonksiyon:

$$\sigma : \mathbb{R} \rightarrow [0,1]$$

$$\sigma(y) = \frac{e^y}{1 + e^y}$$

şeklindedir; aldığı her değeri $[0,1]$ araliğinda bir yere eşler.



Her zaman $[0,1]$ aralığında!!

Lojistik fonksiyon ile regresyonu birleşitirerek lojistik regresyon sınıflandırıcısı elde ederiz. Bu sınıflandırıcı $x = (x_1, \dots, x_n)$ örneğini

$$f(x) = \sigma(w_0 + w_1x_1 + \cdots + w_nx_n) = \frac{e^{w_0+w_1x_1+\cdots+w_nx_n}}{1 + e^{w_0+w_1x_1+\cdots+w_nx_n}}$$

fonksiyonu ile $[0,1]$ arası bir değere eşler. Bu değer x örneğinin pozitif sınıfa $(Y = 1)$ ait olma olasiligidir.

$$\boxed{P(Y = 1|x) = \frac{e^{w_0+w_1x_1+\cdots+w_nx_n}}{1 + e^{w_0+w_1x_1+\cdots+w_nx_n}}}$$

Doğal olarak, örnegin negatif sınıfa ait olma olasiligi

$$P(Y = 0|x) = 1 - P(Y = 1|x)$$

şeklinde hesaplanir.

Eğer olasılıklarla değil, direkt pozitif-negatif (kanser- kanser degil) sınıflarıyla ilgileniyorsak $P(Y = 1|x) \geq 0.5$ durumunda x örneğini pozitif; $P(Y = 1|x) < 0.5$ ise x örneğini negatif olarak tahmin ederiz.

Lojistik Regresyon Maliyet Fonksiyonu

Diyelimki eğitim setimizde m tane örnek olsun. Her bir örnek

$$\left(x^i, y^i\right) \text{ öyleki } x^i = (x_1^i, x_2^i, \dots, x_n^i) \text{ ve } y^i \in \{0,1\}$$

$i \in \{1, \dots, m\}$ formundadır.

x^i, nin sınıfı $w_0, w_1, \dots, w_n$ katsayıları varken lojistik regresyonla

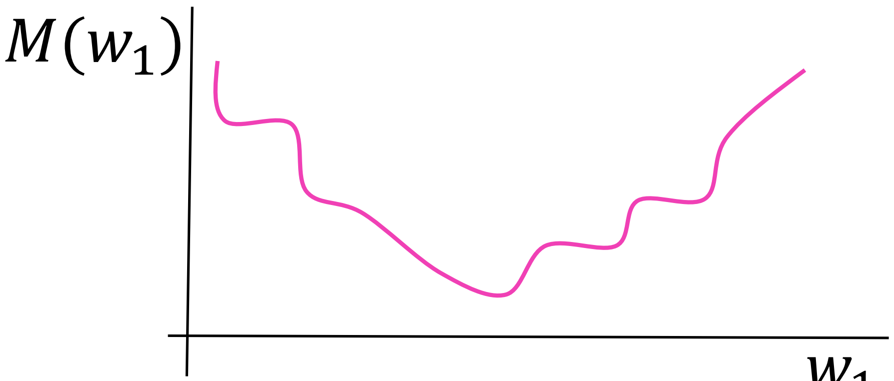
$$\sigma\left(w_0 + w_1 x_1^i + \cdots + w_n x_n^i\right)$$

olarak tahmin edilir.

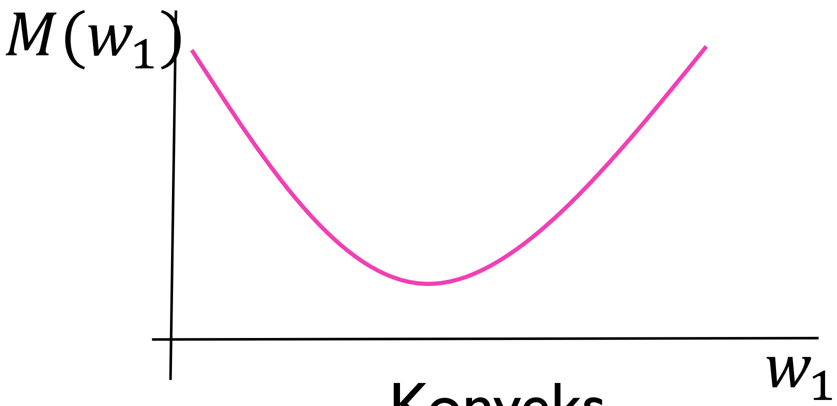
Lineer regresyonda oldugu gibi her (x^i, y^i) örnegi icin tahmin edilen deger $\sigma(w_0 + w_1 x_1^i + \cdots + w_n x_n^i)$ ile gerçek değer y^i arasındaki farkın karesini alarak oluşturacagimiz maliyet fonksiyonu şöyle olur:

$$M(w_0, w_1, \ldots, w_n) = \frac{1}{2m} \sum_{i=1}^{m} \left(\sigma\left(w_0 + w_1 x_1^i + \cdots + w_n x_n^i\right) - y^i\right)^2$$

Fakat bu fonksiyon konveks degildir; bir çok lokal minimum içerir. Bu durumda bu fonksiyonu minimize eden $w_0, w_1, \ldots, w_n$ değerlerini hesaplamak için gradient descent algoritmasindan faydalanamayiz.



Konveks değil



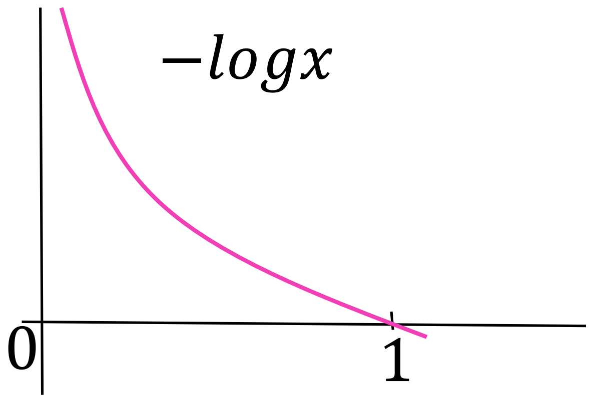
Konveks

Gerçek sınıfı 1 olan bir x^i örneginin

i. sınıfını 0 olarak tahmin edersek maliyetin cok buyuk olmasi gerekir;

ii. sınıfını 1 olarak tahmin edersek maliyetin cok küçük olmasi gerekir.

$[0,1]$ aralığında; 0'da maksimum; 1'de minimum degeriğini alan konveks fonksiyon $-logx$ fonksiyonudur.



Lojistik regresyonda x^i örnegi için sinif tahminimiz $\sigma(w_0 + w_1 x_1^i + \cdots + w_n x_n^i)$ idi. Bu tahmine göre maliyet değisir. O yuzden maliyeti veren $-log$ fonksiyonu $\sigma(w_0 + w_1 x_1^i + \cdots + w_n x_n^i)$ değerinin bir fonksiyonudur.

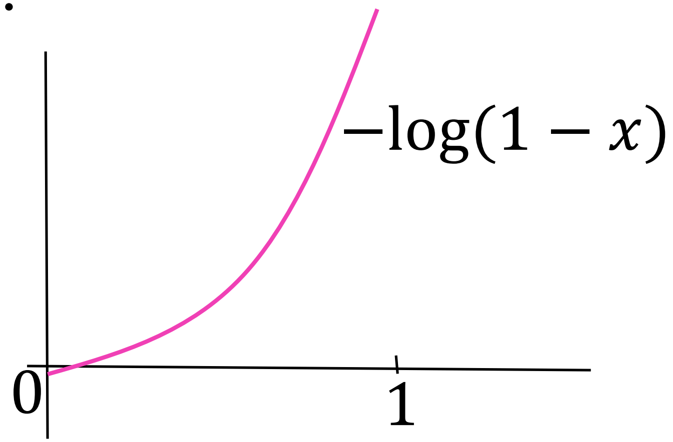
Şu halde $y^i = 1$ iken maliyet: $-log\left(\sigma\left(w_0 + w_1 x_1^i + \cdots + w_n x_n^i\right)\right)$ olur. (1)

Gerçek sınıfı 0 olan bir x^i örneginin

i. sınıfını 1 olarak tahmin edersek maliyetin cok buyuk olmasi gerekir;

ii. sınıfını 0 olarak tahmin edersek maliyetin cok küçük olmasi gerekir.

$[0,1]$ aralığında; 0'da minimum; 1'de maksimum degeriğini alan konveks fonksiyon $-\log(1-x)$ fonksiyonudur.



$y^i = 0$ iken maliyet: $-log\left(1 - \sigma\left(w_0 + w_1 x_1^i + \cdots + w_n x_n^i\right)\right)$ olur. (2)

(1) ve (2) deki maliyetleri birleştirersek:

$$-y^i \cdot log\left(\sigma\left(w_0 + w_1 x_1^i + \cdots + w_n x_n^i\right)\right) - (1 - y^i) \cdot log\left(1 - \sigma\left(w_0 + w_1 x_1^i + \cdots + w_n x_n^i\right)\right)$$

olur. Şu halde ortalama maliyet:

$$\boxed{M(w_0, w_1 \ldots, w_n) = \frac{1}{2m} \sum_{i=1}^{m} -y^i \cdot log\left(\sigma\left(w_0 + w_1 x_1^i + \cdots + w_n x_n^i\right)\right) - (1 - y^i) \cdot log\left(1 - \sigma\left(w_0 + w_1 x_1^i + \cdots + w_n x_n^i\right)\right)}$$

Maliyet fonksiyonun w_0'a göre türevi

$$\frac{\partial M}{\partial w_0} = \frac{1}{m} \sum_{i=1}^{m} \sigma(w_0 + w_1 x_1^i + w_2 x_2^i + \cdots + w_n x_n^i) - y^i$$

Maliyet fonksiyonun w_j'ye göre türevi

$$\frac{\partial M}{\partial w_j} = \frac{1}{m} \sum_{i=1}^{m} (\sigma(w_0 + w_1 x_1^i + w_2 x_2^i + \cdots + w_n x_n^i) - y^i) x_j^i$$

Buldugumuz bu türevler gradient descent algoritmasinda w_j guncellemesinde

$$w_j' := w_j - \alpha \cdot \frac{\partial M}{\partial w_j}$$

yerine konur.

Lojistik Regresyon Icin Gradient Descent Algoritması

Giris: başlangıc vektoru, $M = M(w_0, w_1, \ldots, w_n)$ fonksiyonu

α öğrenme oranı, maxIter maksimum iterasyon sayısı

Cikis: Optimize edilmiş w vektoru

for i=1:maxIter

$$w_0' := w_0 - \alpha \cdot \frac{1}{m} \sum_{i=1}^{m} \sigma(w_0 + w_1 x_1^i + w_2 x_2^i + \cdots + w_n x_n^i) - y^i$$

for j=1:n

$$w_j' := w_j - \alpha \cdot \frac{1}{m} \sum_{i=1}^{m} (\sigma(w_0 + w_1 x_1^i + w_2 x_2^i + \cdots + w_n x_n^i) - y^i) x_j^i$$

end for

$(w_0, w_1, \ldots, w_n) \leftarrow (w_0', w_1', \ldots, w_n')$ //w'ları aynı anda güncelle

end for