

# İLERİ DOĞAL DİL İŞLEME

BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ

BURSA TEKNİK ÜNİVERSİTESİ 2010

DR. ÖĞR. ÜYESİ HAYRİ VOLKAN AGUN

# Özet

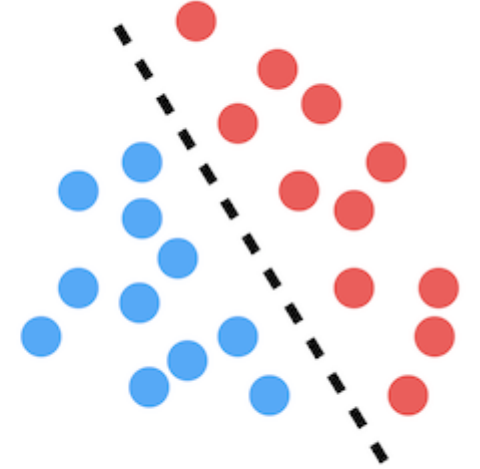
---

- ☐ Ayırt edici sınıflandırma (discrimantive)
- ☐ Logistic regression, gradient
- ☐ Multinomial logistic regression



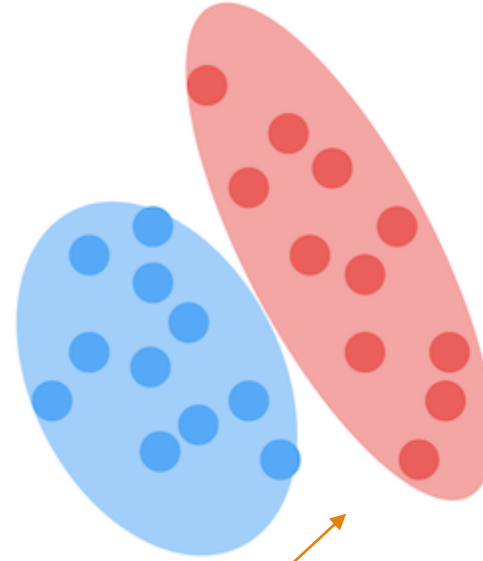
# Sınıflandırma Türleri – Ayırt Edici (Discriminative)

- Ayırt edici sınıflandırmada veri dağılımına bakılmaksızın bir sınır fonksiyonu elde edilmektedir. Kullanılan sınır fonksiyonu girilen bir girdi için doğru sınıfın bulunmasında kullanılır.

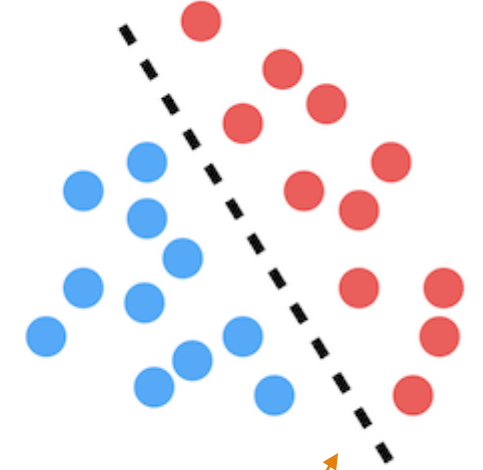


# Sınıflandırma Türleri – Ayırt Edici (Discriminative)

- ❑ Ayırt edici sınıflandırmada veri dağılımına bakılmaksızın bir sınır fonksiyonu elde edilmektedir. Kullanılan sınır fonksiyonu girilen bir girdi için doğru sınıfın bulunmasında kullanılır.
- ❑ Ayırt edici sınıflandırma ve üretici sınıflandırma modellerinde kullanılan matematik birbirine çok benzer olabilir ancak temel bu iki sınıflandırma yöntemi ya girdiyi yada ayırt edici modeli oluşturmada kullanılır.



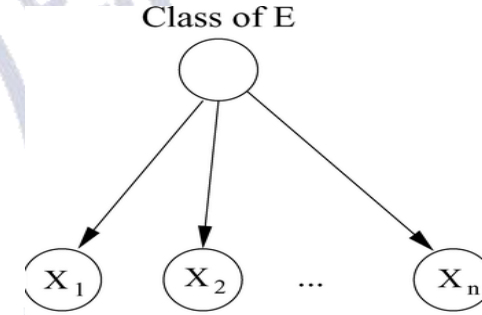
Girdi modellemesi



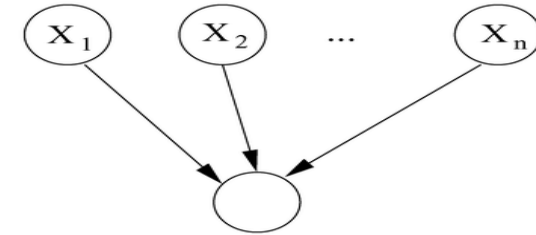
Sınır modellemesi

# Sınıflandırma Türleri – Ayırt Edici (Discriminative)

- Ayırt edici sınıflandırma türleri için ayırt edici sınır bir doğru, düzlem gibi iki farklı örnekleme ayırt etmekte kullanılan bir denklem şeklinde elde edilmelidir.
- Eğer bu denklem doğru yada düzlem denklemi ise o zaman sınıflandırma fonksiyonu doğrusal (lineer) olacaktır.
- En temel doğrusal sınıflandırma yöntemleri arasında Logistic Regression gösterilebilir.



Naive Bayes



Logistic regression

# Naive Bayes ve Logistic Regression Benzerlikler

- Önceki sunumlarda Naive Bayes sınıflandırma için kullandığımız modelde her bir kelime yada özelliğin bir birinden bağımsız olduğunu kabul etmiştik. Bu durumda bir dokümanda geçen kelimelerin o sınıfta geçme olasılığını çarparak dokümanın o sınıfa ait olma olasılığını bulabiliyoruz.
- Logistic Regression doğrusal bir sınıflandırma yapmak için bir sınıfa ait ağırlıklardan yararlanarak dokümanın doğrusal bir ayırt edici fonksiyonun neresinde kaldığına karar verir. Bu karar sonucunda sınıflandırılmak istenen doküman sınıfa ait etikete ya sahiptir yada değildir.

$$P(spam|win,buy) = P(spam|win)P(spam|buy)$$

$$\frac{P(win|spam)P(spam)}{P(win|spam)+P(win|notspam)} \frac{P(buy|spam)P(spam)}{P(buy|spam)+P(buy|notspam)}$$

$$P(Y=1|F) = \frac{1}{1 + \exp(\beta_0 + \sum_{i=1}^n \beta_i F_i)}$$

$$P(Y=0|F) = \frac{\exp(\beta_0 + \sum_{i=1}^n \beta_i F_i)}{1 + \exp(\beta_0 + \sum_{i=1}^n \beta_i F_i)}$$

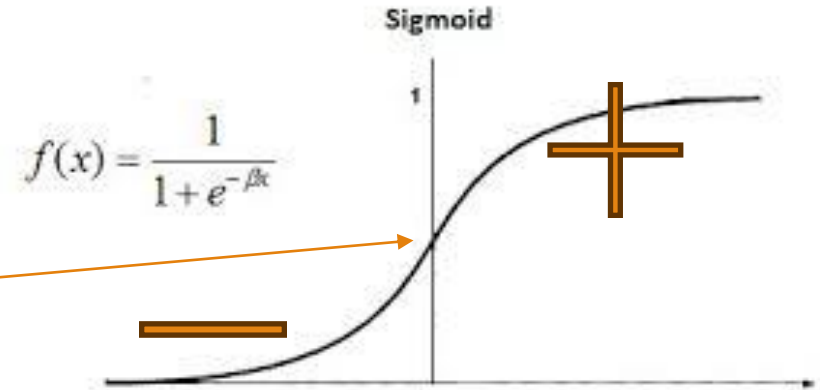
- Burada özellik değerleri ve ağırlık değerleri logistic (soft-max) fonksiyonda çarpılarak bir puan elde edilir. Bu puan sınıftan büyük ise bu doküman bu sınıfa aittir denir.

$$\beta_0 + \sum_{i=1}^n \beta_i F_i < 0$$



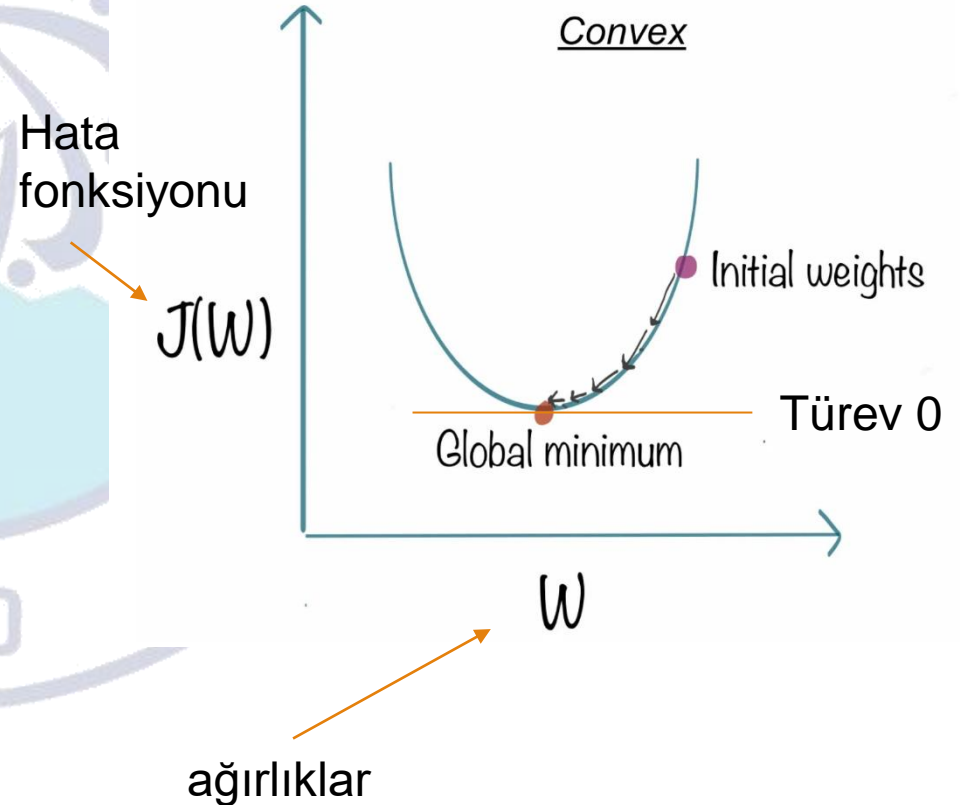
# Logistic Rgression

- Logistic Regression ağırlık parametrelerinin gradient metodları ile öğrenilmesi sonucu eğitilir.
- Logistic (soft-max) fonksiyonu tek bir x girdi değeri için yanda verilmiştir.
- Bu fonksiyonun değerinin 0.5'den büyük olduğu sayılar için girdiye ait sınıf pozitif olacaktır.



# Ağırlıkların bulunması

- Ağırlıkların bulunması için tüm girdiler için oluşan hata değerleri elde edilmeli ve bu değerleri minimum yapan ağırlıklar hesaplanmalıdır.
- Logistic Regression için hata fonksiyonu çok kolay türevlenebilir. Bu durumda türevin (gradient) 0 olduğu noktaya doğru azalan bir arama yaparak hata fonksiyonuna karşılık gelen ağırlıkları bulabiliriz.





# Hata fonksiyonu

$$Cost(h_{\theta}(x), y) = \begin{cases} -\log(h_{\theta}(x)) & \text{if } y = 1 \\ -\log(1 - h_{\theta}(x)) & \text{if } y = 0 \end{cases}$$

$$Cost(h_{\theta}(x), y) = -y \log(h_{\theta}(x)) - (1 - y) \log(1 - h_{\theta}(x))$$

Şimdiki parametreler  
ile elde edilen  
logistic fonksiyonu

$$J(\theta) = \frac{1}{m} \left[ \sum_{i=1}^m -y^{(i)} \log(h_{\theta}(x^{(i)})) + (1 - y^{(i)}) \log(1 - h_{\theta}(x^{(i)})) \right]$$

eğitim örneklerinin sayısı

eğitim örneklem vektörü (özellik vektörü)

# Hatanın Türevi ve Güncelleme

---

$$\frac{\partial J(\theta)}{\partial \theta} = \frac{1}{m} (h_{\theta}(x) - y) x$$

Iterate until convergence

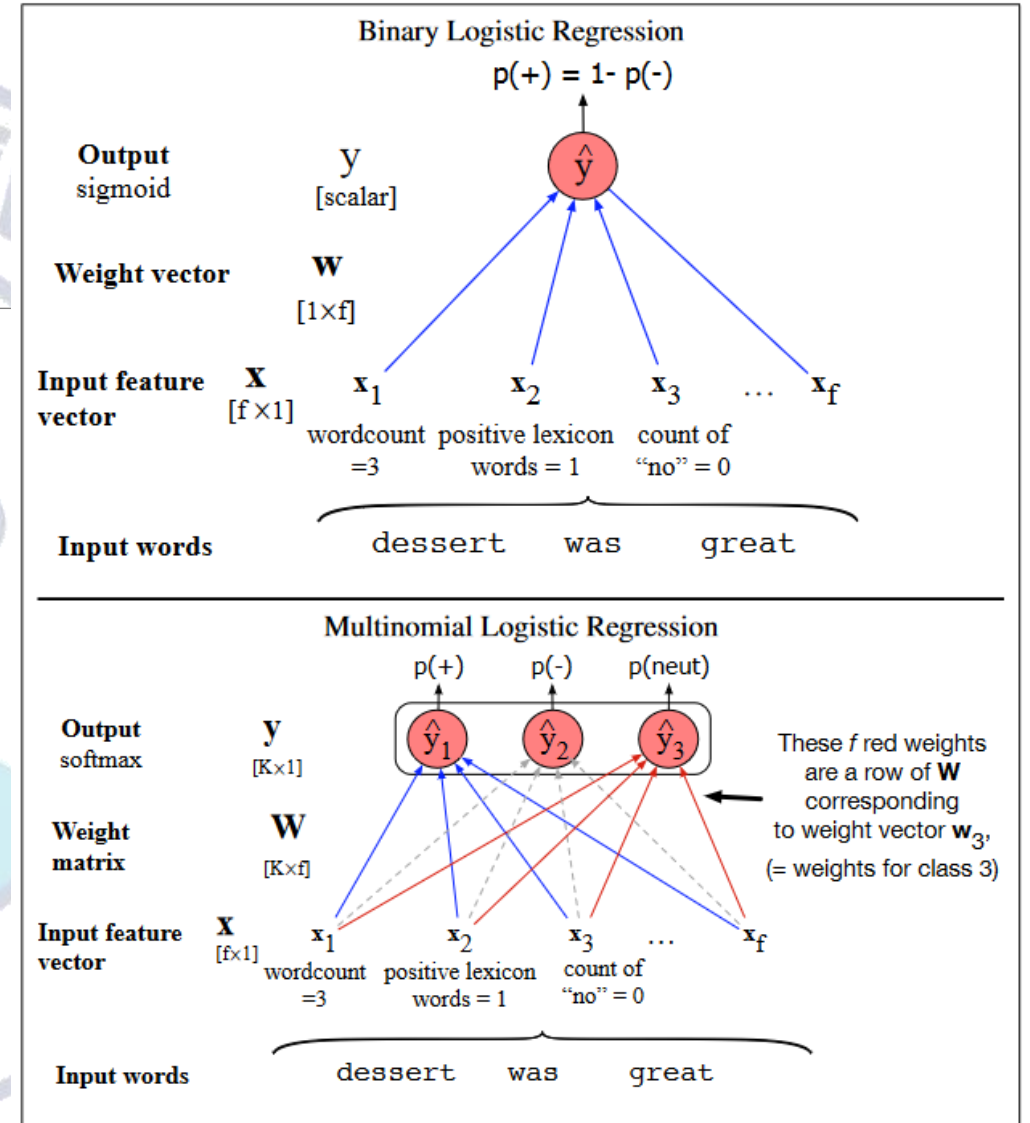
$$\left\{ \begin{array}{l} \theta_j = \theta_j - \alpha (h_{\theta}(x) - y) x \end{array} \right.$$

- $\alpha$  adım miktarını belirtir.
- $m$  eğitim örneklerinin sayısını belirtir. Bu sayı batch sayısı olarak seçilebilir.

# Birden fazla sınıf

- Örneğin bir cümlenin duygu durumu pozitif, negatif ve nötr olsun.
- Yandaki örnekte bu multinomial logistic regression ile açıklanmaktadır. Buna göre her bir sınıf için ayrı bir parametre vektörü vardır. Tüm sınıflar için bu bir matristir. Aynı anda tüm çıktılar için parametreler güncellenir.
- Bunun için örneğin aşağıdaki vektör K farklı sınıfın çıktısı için hata fonksiyonunun hesaplanmasında ve güncellemede kullanılır.

$$\text{softmax}(\mathbf{z}) = \left[ \frac{\exp(z_1)}{\sum_{i=1}^K \exp(z_i)}, \frac{\exp(z_2)}{\sum_{i=1}^K \exp(z_i)}, \dots, \frac{\exp(z_K)}{\sum_{i=1}^K \exp(z_i)} \right]$$



# Kaynaklar

---

1. <https://web.stanford.edu/~jurafsky/slp3/5.pdf>
2. <https://www.coursera.org/learn/machine-learning>
3. <https://towardsdatascience.com/optimization-loss-function-under-the-hood-part-ii-d20a239cde11>
4. <https://towardsdatascience.com/understanding-logistic-regression-step-by-step-704a78be7e0a>