Makine Öğrenmesi
Destek Vektör Makinalar

Doç.Dr. İlhan AYDIN

SVM sınıflandırma

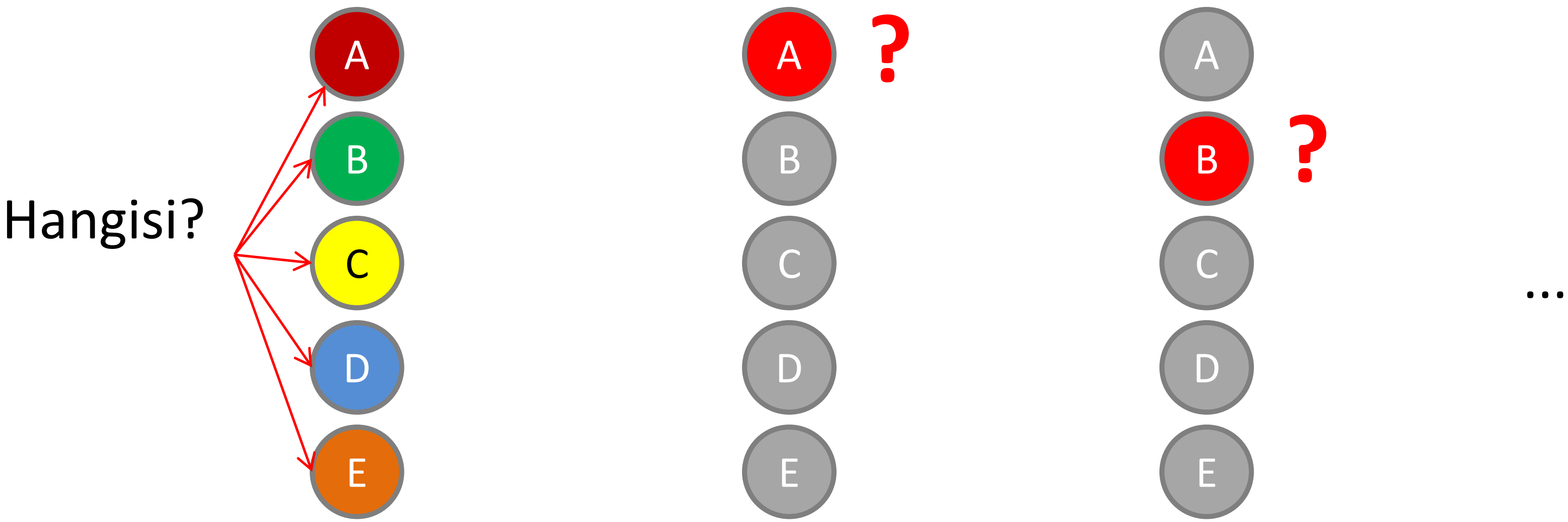
- SVM, lineer sınıflandırma makine öğrenme metodudur
- Son derecede basit; aynı zamanda gerçek uygulamalarda Yapay Sınır Ağlarına yakın performans gösteren güçlü bir yöntemdir

SVM sınıflandırma

- Spam mesajları yakalamak
- Erken olarak hastaları bulmak
- Aday işine uygunluğu belirlemek
- Konuşmadaki sözün var olduğunu belirtmek
- Resimdeki belirli türden nesne bulmak
- Haber mesajının belirli insan için ilginç olabilmesi belirlemek

SVM sınıflandırma

- “Tüm karşı bir” birkaç sınıflı sınıflandırma;



Birkaç-olasılıklı soru

“A mı?” iki-olasılıklı soru

“B mı?” iki-olasılıklı soru

SVM sınıflandırma

- Notasyon
  - x’ler – özellikler (bir özellik vektörü veya dizisi)
  - y’ler – sonuçlar (sınıf etiketleri, evet/hair, 1/0, vb)
  - n – özelliklerin sayısı (x’in boyutu)
  - x_j – j. özellik
  - (x^i,y^i) – bir örnek, önceden var olan özellikler için sonuç
  - m –örneklerin sayısı
  - bütün (x^i,y^i) –“eğitim kümesi”

SVM sınıflandırma

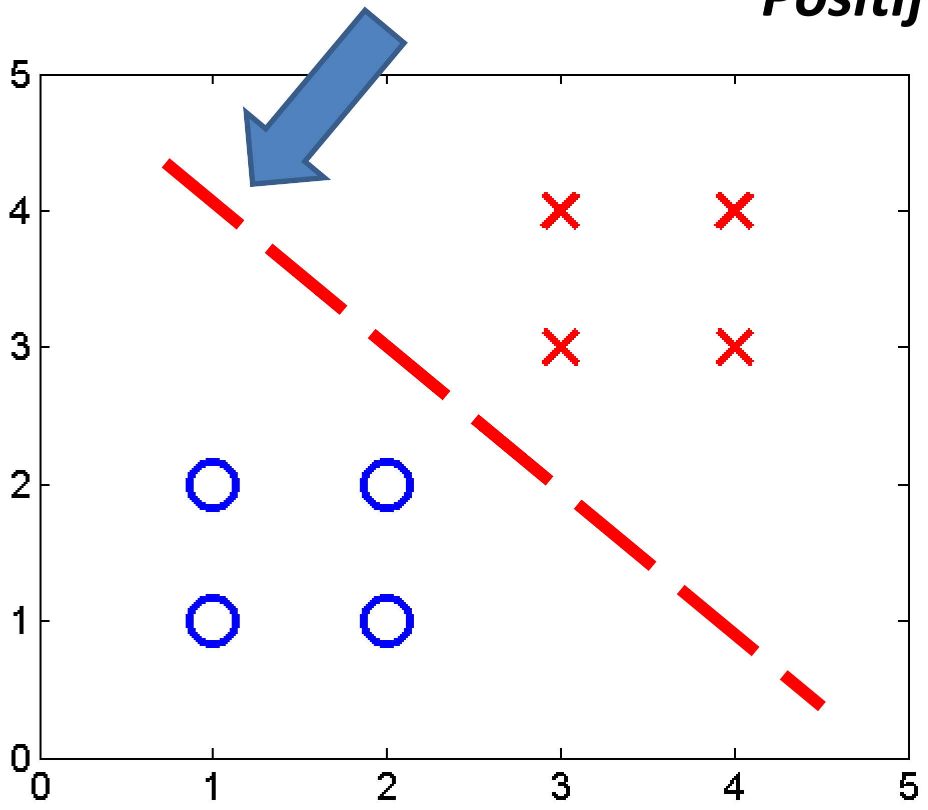
SVM – “Support Vector Machine” yada “Destek Vektör Makinesi”, önce görülen lojistik regresyona benzer bir lineer sınıflandırma metodudur;

- “pozitif” (y=+1) ve “negatif” (y=-1) örnekler için, bu örnekleri doğru şekilde bölen lineer karar sınırı buluyor

SVM sınıflandırma

Karar sınırı:

Positif örnekler: y=1



Negatif örnekler: y=-1

SVM sınıflandırma

Lojistik regresyon böyle sınırı bulmak için, özel maliyet fonksiyonunu azaltılma sorunu formülleştirdi:

$$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}(-y^i \log h_\theta(x^i) - (1-y^i)\log(1-h_\theta(x^i)))$$

$$h_\theta(x) = \frac{1}{1+e^{-\theta^T \cdot x}}$$

model

$$y^i = \{0,1\}$$

gereken çıktılar

SVM sınıflandırma

SVM, bu sınırı farklı açıdan buluyor ...

SVM sınıflandırma

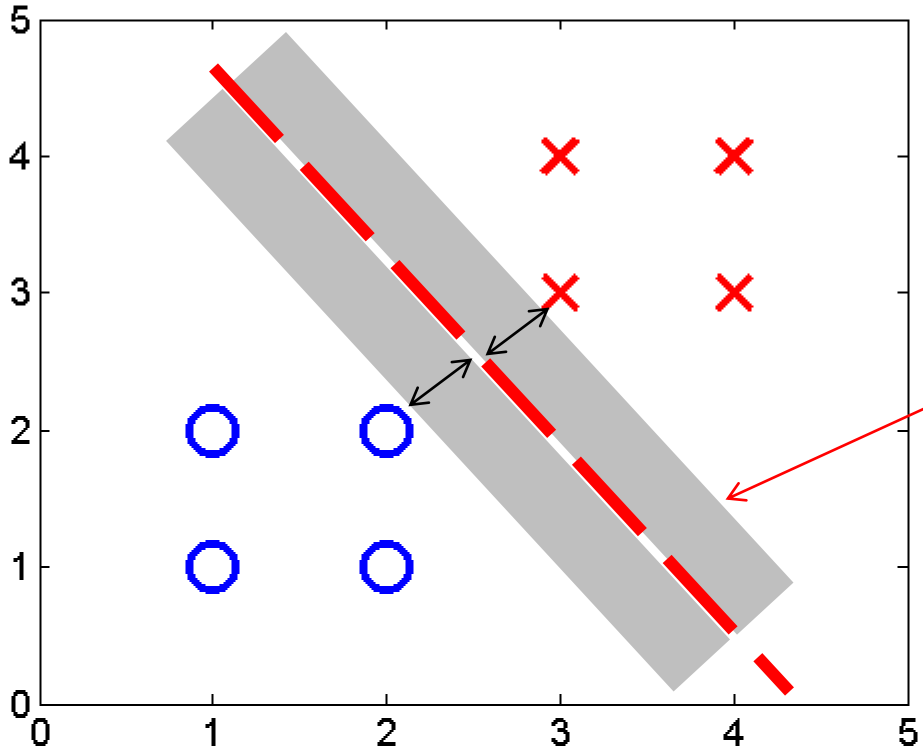
“Büyük Kenar Prensibi”

- Büyük Kenar Prensibine göre, karar sınırı pozitif ve negatif örneklerin arasında en büyük kenarı alır. Diğer bir deyişle iki sınıf arasındaki marjini maksimum yapar.

SVM sınıflandırma

Karar kenarı:

Positif örnekler: y=1



Kenar

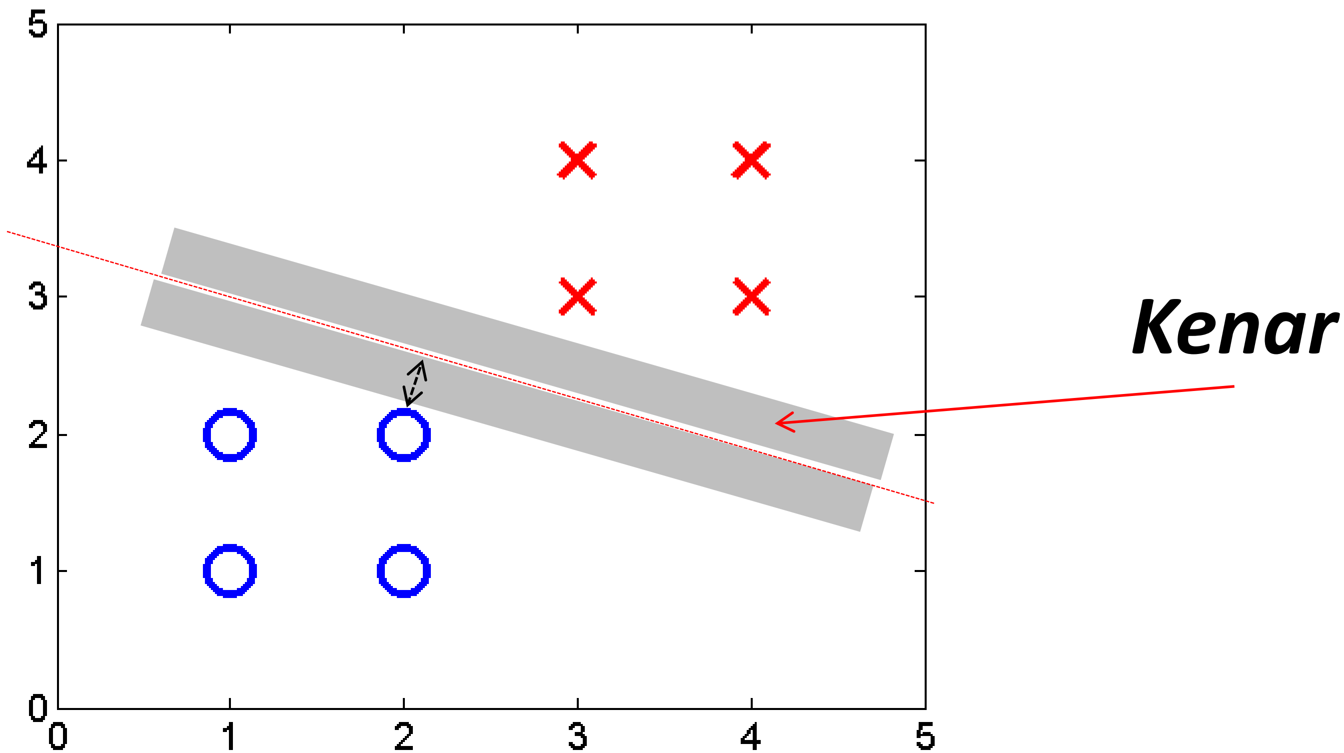
Pozitif ve negatif örneklerin arasındaki mesafe

Negatif örnekler: y=-1

SVM sınıflandırma

En büyük karar kenarı?

Positif örnekler: y=1

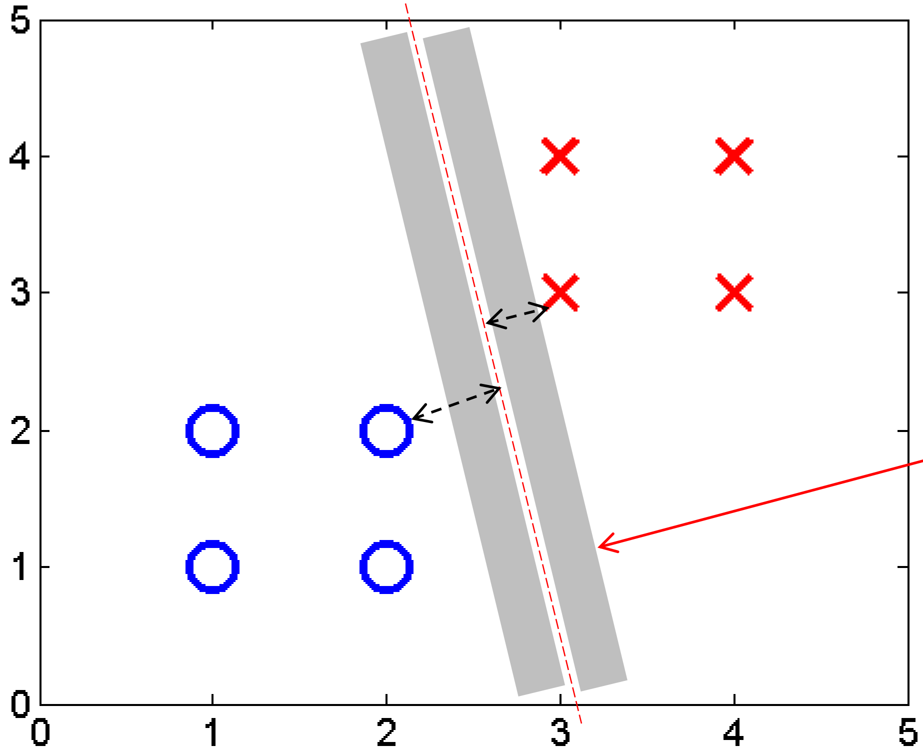


Negatif örnekler: y=-1

SVM sınıflandırma

En büyük karar kenarı?

Positif örnekler: y=1



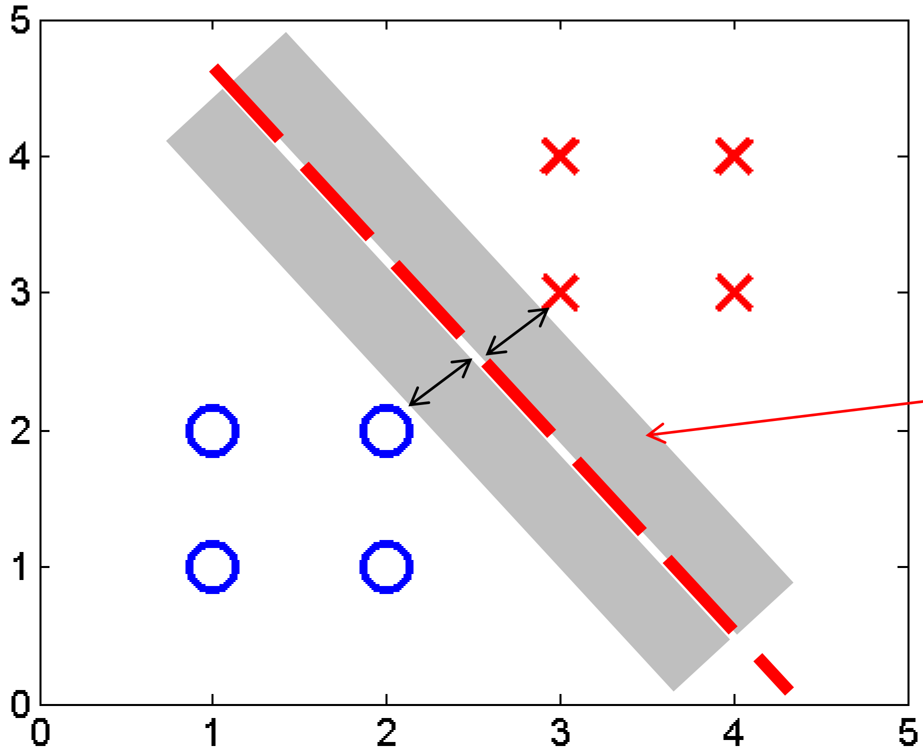
Kenar

Negatif örnekler: y=-1

SVM sınıflandırma

En büyük karar kenarı !

Positif örnekler: y=1



En büyük kenarı!

Negatif örnekler: y=-1

SVM sınıflandırma

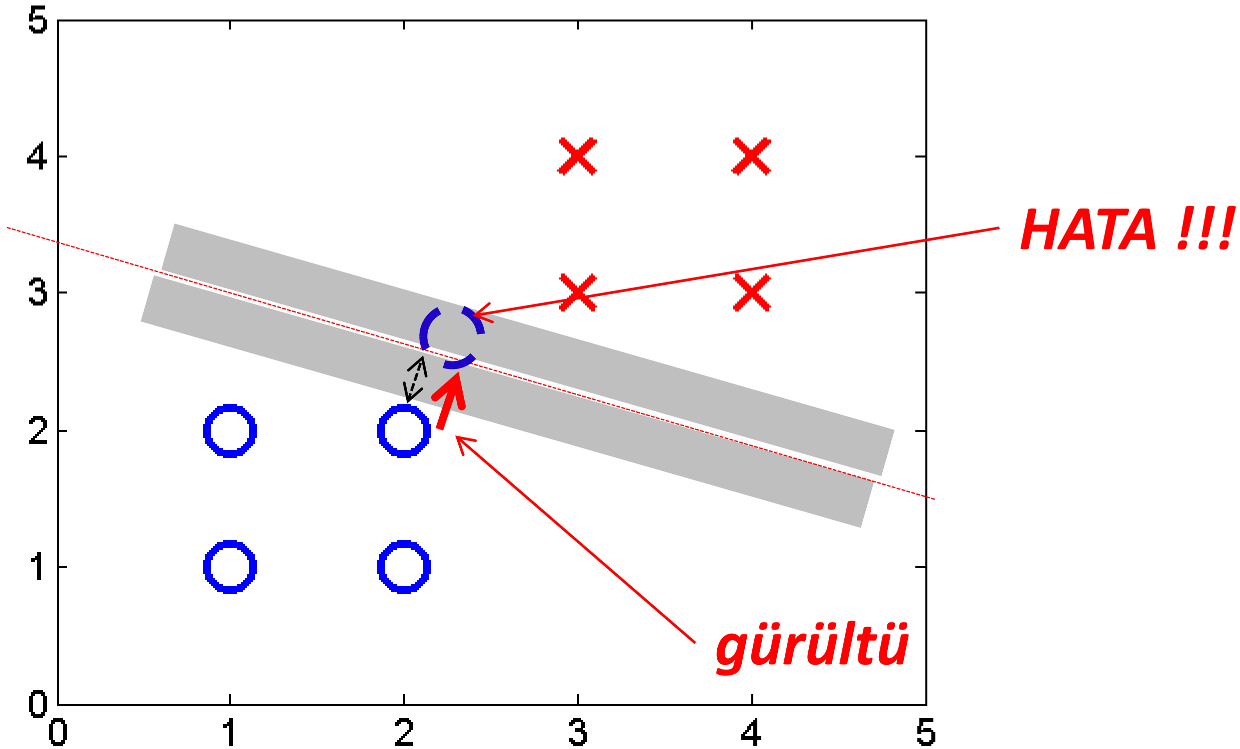
“Büyük Kenar Prensibi”

- En büyük kenarı, en gürbüz kararlar demektir !
- Örneklerde gürültü varsa, en büyük kenarıyla en gürbüz karar verme yapılabilir

SVM sınıflandırma

Az kenarlı karar sınırı, gürültü tarafından kolayca etkilenebilir

Positif örnekler: y=1

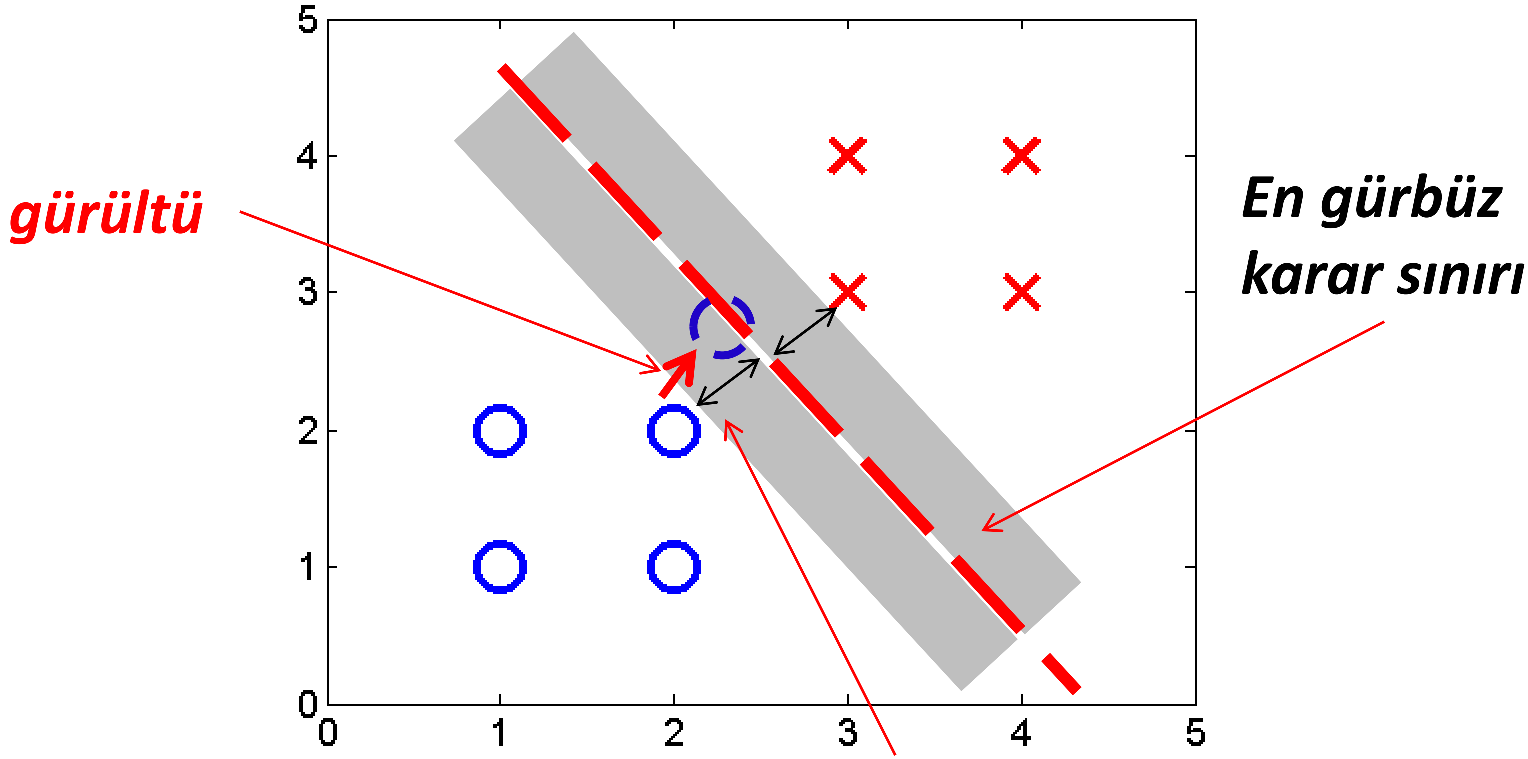


Negatif örnekler: y=-1

SVM sınıflandırma

Büyük kenarlı karar sınırı, en gürbüz karar verme demektir !

Positif örnekler: y=1



Negatif örnekler: y=-1

HATA YOK !

SVM sınıflandırma

- SVM karar sınırı bu şekilde çözülür:

$$\min_{\theta} C \sum_{i=1}^{m} H_{y^i}(\theta^T \cdot x^i) + \frac{1}{2} \sum \theta_j^2,$$

$$H_y(z) = \begin{cases} 0, & y \cdot z \geq 1 \\ 1 - y \cdot z, & y \cdot z < 1 \end{cases}$$

$$y^i = \pm 1$$

SVM sınıflandırma

Fark edilecek önemli nokta:

- Eğer bütün örneklerde $y^i \cdot (\theta^T \cdot x^i) \geq 1$, *H-terimi sıfırdır*

$$\min_{\theta} C \sum_{i=1}^{m} H_{y^i}(\theta^T \cdot x^i) + \frac{1}{2} \sum \theta_j^2,$$

$$H_y(z) = \begin{cases} 0, & y \cdot z \geq 1 \\ 1 - y \cdot z, & y \cdot z < 1 \end{cases}$$

$$y^i = \pm 1$$

SVM sınıflandırma

Fark edilecek önemli nokta:

- Bu durumda sadece $\frac{1}{2}\sum \theta_j^2$ azaltmaya çalışıyoruz

$$\min_{\theta} \frac{1}{2}\sum \theta_j^2,$$

$$H_y(z) = \begin{cases} 0, & y \cdot z \geq 1 \\ 1 - y \cdot z, & y \cdot z < 1 \end{cases}$$

$$y^i = \pm 1$$

SVM sınıflandırma

SVM her zaman, mümkünse, pozitif örnekleri bir tarafa (“+1” üstüne) negatif örnekler diğer tarafa (“-1” altına) çekmeye çalışıyor

$$\min_{\theta} C \cancel{\sum_{i=1}^{m} H_{y^i}}(\theta^T \cdot x^i) + \frac{1}{2}\sum \theta_j^2,$$

$$H_y(z) = \begin{cases} 0, & y \cdot z \geq 1 \\ 1 - y \cdot z, & y \cdot z < 1 \end{cases}$$

$$y^i = \pm 1$$

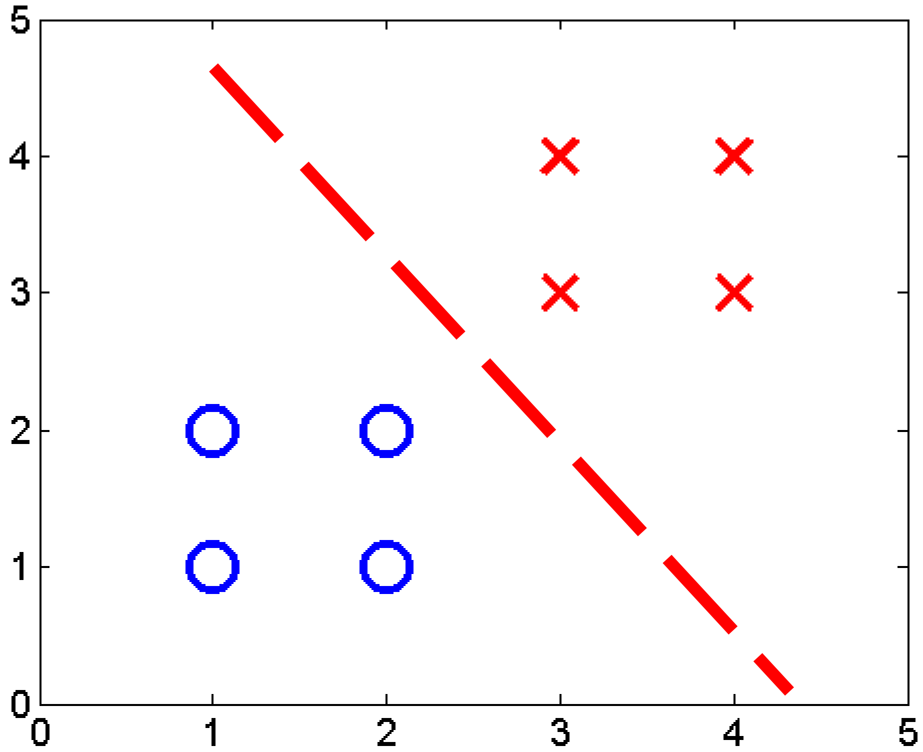
SVM sınıflandırma

- Bütün pozitif örnekler “+1” üstüne ve negatif örnekler “-1” altına çekilebilirse, soruna “doğrusal olarak ayrılabilir bir sorun” diye diyoruz

SVM sınıflandırma

Doğrusal olarak ayrılabilir sorun

Positif örnekler: y=1



Negatif örnekler: y=-1

SVM sınıflandırma

- *Doğrusal olarak ayrılabilir sorunlar için*, SVM'de karar sınırı uygun seçimle, H-terim sıfıra eşit yapılabilir ve sadece $\sum \theta_j^2$ teriminin azaltılması gerekiyor

$$\min_{\theta} C \sum_{i=1}^{m} H_{y^i}(\theta^T \cdot x^i) + \frac{1}{2}\sum \theta_j^2,$$

min

=0

$$H_y(z) = \begin{cases} 0, & y \cdot z \geq 1 \\ 1 - y \cdot z, & y \cdot z < 1 \end{cases}$$

$$y^i = \pm 1$$

SVM sınıflandırma

- Aslında, en düşük $\sum \theta_j^2$ en büyük kenarlı doğrusal karar sınırı vermektedir

SVM sınıflandırma

- Şu gösterilebilir ki, SVM problemin çözümü her zaman bu şekilde olmaktadır;

$$\theta_j = \sum \alpha^i \cdot y^i \cdot x_j^i$$

SVM sınıflandırma

- Doğrusal karar sınırının θ-parametresi, örneklerin x-özellik-vektörlerinden oluşturulur ... ama ...

$$\theta_j = \sum \alpha^i \cdot y^i \cdot x_j^i$$

- Bütün örnekler bu toplama girmez, genellikle sadece birkaç alfa sıfırdan farklı ve sadece o x-özellik-vektörler θ-parametresine girir
  - Tipik olarak, 1000 örneklerinden sadece 20-50 örnek θ-parametresine girebilir

SVM sınıflandırma

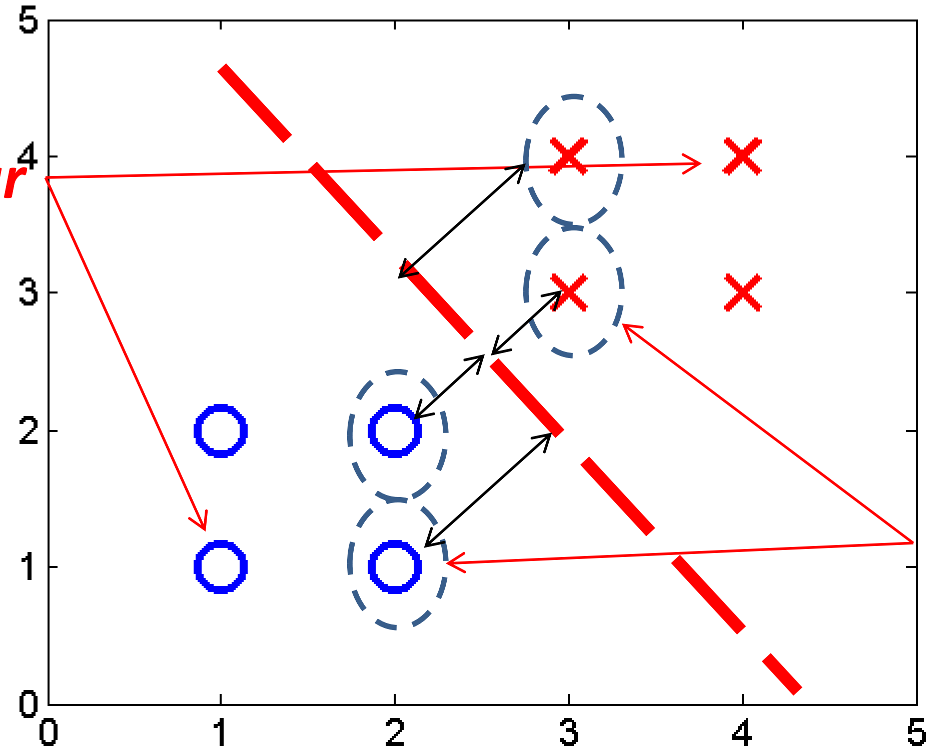
- Bu şekilde katkısı olan orijinal örneklere "destek vektörleri" denir
- Sadece o örnekler θ-parametresini, ve bu anlamda SVM kararlarını, etkiliyor;
- Tüm diğer örnekler karar verme için önemli değil (SVM kararlarına etkilemez)

SVM sınıflandırma

Destek vektörleri:

Positif örnekler: y=1

Önemsiz örnekler; şunlar yoksa, karar sınırı değişmez



Önemli örnekler, bunlar karar sınırı belirtir; destek vektörleri bu

Negatif örnekler: y=-1

SVM sınıflandırma

“Destek vektör makineleri” bu nedenle diyoruz

SVM sınıflandırma

SVM optimizasyon sorununun 2. hali;

- Örnekler doğrusal olarak ayrılamaz
- Bu durumda, H-terim sıfıra konulmaz, ve SVM'de hem H-terimi hem de θ-terimi azaltılması lazım

$$\min_{\theta} C \sum_{i=1}^{m} H_{y^i}(\theta^T \cdot x^i) + \frac{1}{2}\sum \theta_j^2,$$

$\cancel{=0}$

$$H_y(z) = \begin{cases} 0, & y \cdot z \geq 1 \\ 1 - y \cdot z, & y \cdot z < 1 \end{cases}$$

$$y^i = \pm 1$$

SVM sınıflandırma

SVM optimizasyon sorununun 2. şekil;

- Bu durumda, C-sabiti önemli oluyor
- C-sabiti yanlış örneklerin maliyetteki ağırlığını belirtiyor

$$\min_{\theta} C \sum_{i=1}^{m} H_{y^i}(\theta^T \cdot x^i) + \frac{1}{2}\sum \theta_j^2,$$

$$H_y(z) = \begin{cases} 0, & y \cdot z \geq 1 \\ 1 - y \cdot z, & y \cdot z < 1 \end{cases}$$

önemli!

$$y^i = \pm 1$$

SVM sınıflandırma

- Büyük C için, SVM oldukça az yanlış karar yapmaya çalışacak
- Düşük C için, SVM daha çok yanlış karar yapılabilir ama oldukça büyük karar kenarı sağlamaya çalışacak

$$\min_{\theta} C \sum_{i=1}^{m} H_{y^i}(\theta^T \cdot x^i) + \frac{1}{2}\sum \theta_j^2,$$

$$H_y(z) = \begin{cases} 0, y \cdot z \geq 1 \\ 1 - y \cdot z, y \cdot z < 1 \end{cases}$$

$$\theta = \sum \alpha^i y^i x^i$$

SVM sınıflandırma

- SVM, kavramsal basit (yani, == en büyük karar kenarı) aynı zamanda makine öğrenme en güçlü metodlarından biridir
- *El yazısı karakter tanıma* (posta), *yüz tanıma* (facebook), *metin sınıflandırma* (haber seçme) uygulamalarında çok ileri yapay sınır ağları gibi yaklaşımlarla karşılaştırılabilir yada daha iyi perfomansı tipik olarak gösterir

SVM sınıflandırma

- Bu anda SVM yaklaşımı, makine öğrenme yaklaşımları olarak altın standart rolünde bulunmaktadır
- Daha ileri okuma için,
  - SVM, Vladimir Vapnik ve Alexey Chernovenkis tarafından istatistiksel öğrenme teorisi kullanarak türetilmiş
  - İstatistiksel öğrenme teorisi ve SVM ile ilgili İnternette birçok (ingilizce) kaynak var

SVM sınıflandırma

- SVM optimizasyon sorunu çözmek için, dereceli azaltma ve aşağudaki maliyet fonksiyonu kullanılabilir

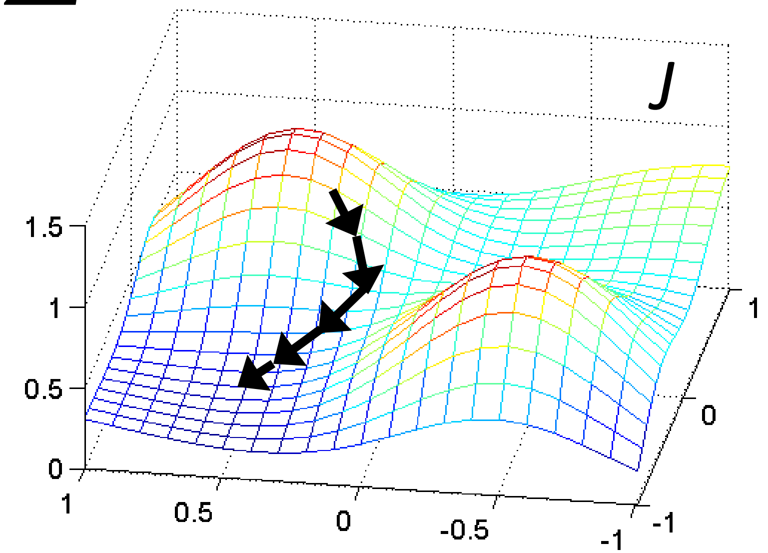
$$J(\theta) = C \sum_{i=1}^{m} H_{y^i}(\theta^T \cdot x^i) + \frac{1}{2} \sum \theta_j^2$$

Yakınsamaya kadar tekrarlayın {

bütün *j*'ler için;

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$$

}



SVM sınıflandırma

- Daha verimli yöntemleri uygulanabilir
- SVM sorununun özel matematiksel yapısına sahip olması için, özel metodlar geliştirilmiş oldu

$$\min_{\theta} C \sum_{i=1}^{m} H_{y^i}(\theta^T \cdot x^i) + \frac{1}{2} \sum \theta_j^2,$$

$$H_y(z) = \begin{cases} 0, y \cdot z \geq 1 \\ 1 - y \cdot z, y \cdot z < 1 \end{cases}$$

$$\theta = \sum \alpha^i y^i x^i$$

SVM sınıflandırma

Bunların arasında;

- Sequential Mimimal Optimization, SMO (Ardışık minimal optimizasyon metodu)
- İnterior point convex optimization metods (Konveks optimizasyonun iç nokta yöntemi)

SVM sınıflandırma

SVM için hazır yazılım paketleri kullanılmalı;

- *SVM-light*
- *SVM-struct*
- *mySVM*
- *LIBSVM*

(www.support-vector-machines.org da daha çok yazılım paketleri bulunabilir)

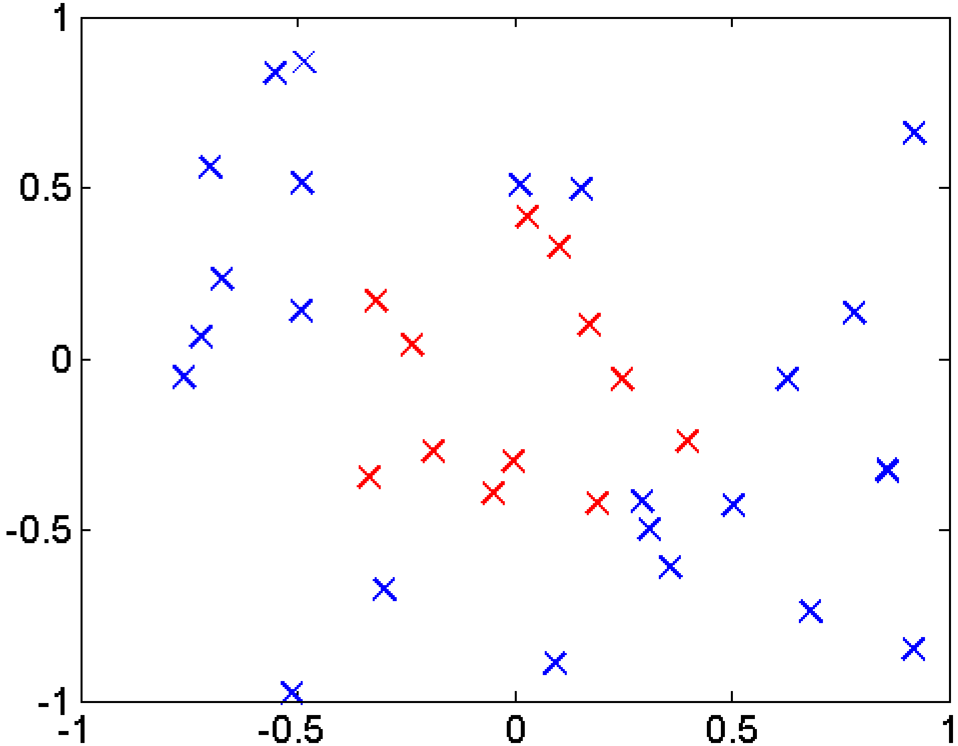
SVM sınıflandırma

Lineer olmayan sınıflandırma için, lojistik regresyon gibi SVM, ek olan lineer olmayan karmaşık özellikler ile kullanılabilir

SVM sınıflandırma

Lineer olmayan ilişki:

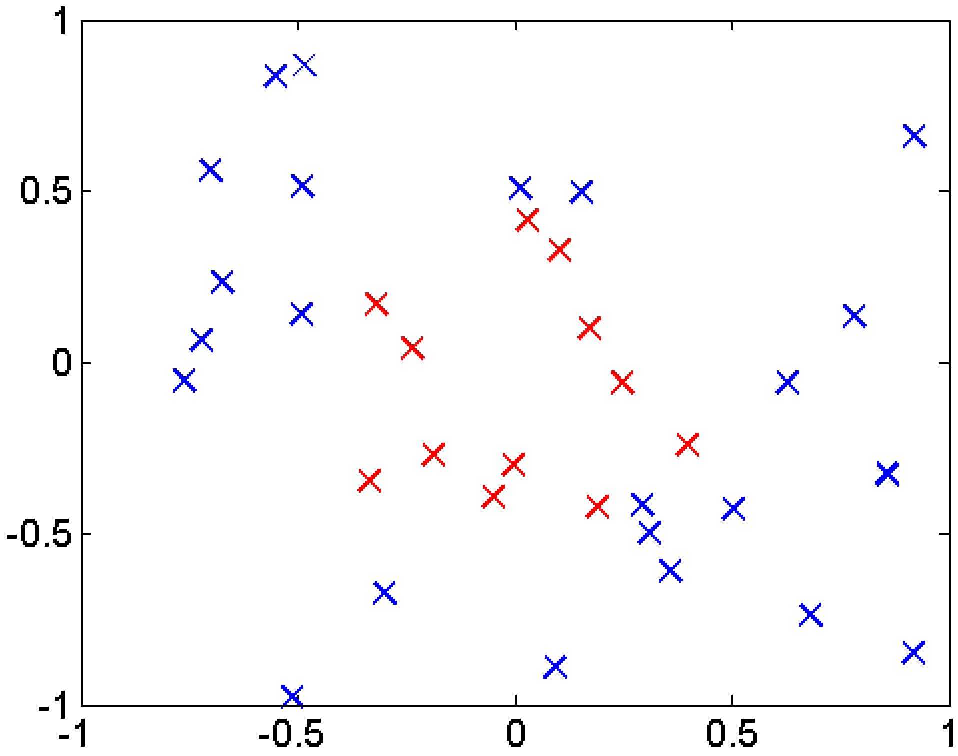
kırmızı – y=1 örnekleri, mavi – y=-1 örnekleri



SVM sınıflandırma

Ek olan lineer olmayan yeni karmaşık özellik:

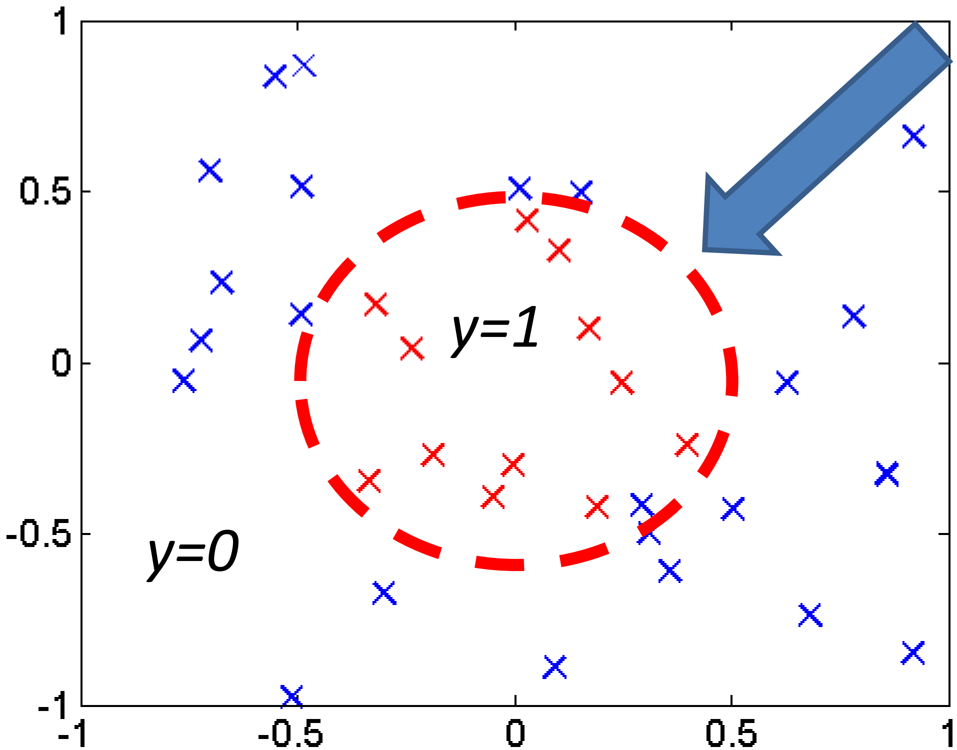
$$x' = g({x_1}^2 + {x_2}^2)$$



SVM sınıflandırma

Bu özelliğe göre, muhtemelen lineer olmayan karar sınırı bu şekilde üretilebilir:

$$x' = x_1^2 + x_2^2 = 1$$



SVM sınıflandırma

- SVM'de, lineer olmayan özellikler için "kernel" (çekirdek) diye terim kullanılır, K(x^i,x^j) sembol ile gösterilir
- Kerneller kullanarak, SVM lineer olmayan sınıflandırma için kullanılabilir

SVM sınıflandırma

- SVM problemi, bu şekilde de yazılabilir (dual şekli adlandırılan şekil);

$$\max \sum \alpha^i - \frac{1}{2} \sum \alpha^i \alpha^j y^i y^j (x^i \cdot x^j)$$

$$\sum \alpha^i y^i = 0$$

$$0 \le \alpha^i \le C$$

SVM sınıflandırma

- Bu SVM problemi, doğrudan α-parametreler (θ-parametreler değil) için yazılır (α, destek vektörlerinin ağırlıkları)

$$\max \sum \alpha^{i} - \frac{1}{2} \sum \alpha^{i} \alpha^{j} y^{i} y^{j} (x^{i} \cdot x^{j})$$

$$\sum \alpha^{i} y^{i} = 0$$

$$0 \leq \alpha^{i} \leq C$$

$$\boxed{\theta = \sum \alpha^{i} y^{i} x^{i}}$$

SVM sınıflandırma

- Bu SVM probleminde, x'ler bir skalar çarpım içinde girmektedir

$$\max \sum \alpha^i - \frac{1}{2} \sum \alpha^i \alpha^j y^i y^j (x^i \cdot x^j)$$

$$\sum \alpha^i y^i = 0$$

$$0 \le \alpha^i \le C$$

$$\boxed{(x^i \cdot x^j) = \sum x_k^i x_k^j}$$

SVM sınıflandırma

- Kernel şeklini sağlamak için, bu skalar çarpım lineer-olmayan bir şekile değiştirilir:

$$\max \sum \alpha^i - \frac{1}{2} \sum \alpha^i \alpha^j y^i y^j K(x^i, x^j)$$

$$\sum \alpha^i y^i = 0$$

$$0 \leq \alpha^i \leq C$$

Skalar çarpım yerine giriyor

$$\boxed{(x^i \cdot x^j) \equiv K(x^i, x^j)}$$

SVM sınıflandırma

- SVM kernel şekli olarak, en çok kullanılan Gauss (radyal) kerneli dir;

$$\max \sum \alpha^i - \frac{1}{2} \sum \alpha^i \alpha^j y^i y^j K(x^i, x^j)$$

$$K(x^i, x^j) = e^{-\frac{(x^i - x^j)^2}{2\sigma^2}}$$

SVM sınıflandırma

- Diğer popüler seçenekleri;
  - Homojen polinom kerneli

$$K(x^i, x^j) = (x^i \cdot x^j)^d$$

  - Homojen olmayan polinom kerneli

$$K(x^i, x^j) = (1 + x^i \cdot x^j)^d$$

  - Hiperbolik tanjant kerneli

$$K(x^i, x^j) = \tanh(c + kx^i \cdot x^j)$$