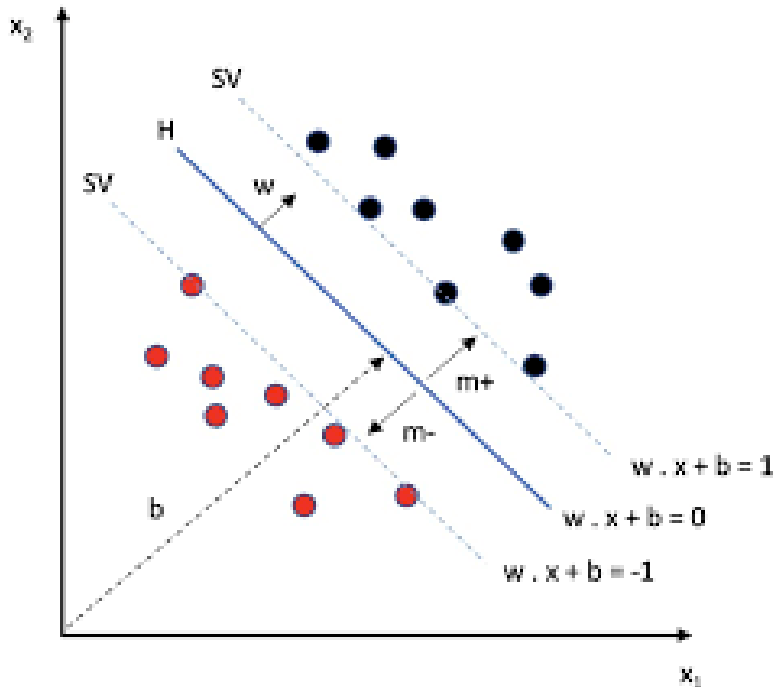




Destek Vektör Makineleri (Support Vector Machines)

Destek vektör makinesi, eğitim verilerindeki herhangi bir noktadan en uzak olan iki sınıf arasında bir karar sınırı bulan vektör uzayı tabanlı makine öğrenme yöntemi olarak tanımlanabilir.



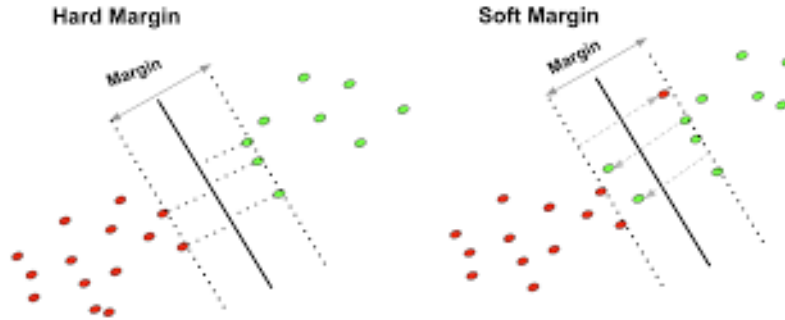
Bazı veriler Linearly seperable (Bir doğru ile ikiye ayrılabilen) verilerdir. Ancak bir veri seti bireden fazla doğru ile ikiye ayrılabilir. Bu doğrulardan en kullanışlı olanını bulmak için destek vektörleri kullanılır. Destek vektör makineleri hem sınıflandırma hem regresyon için kullanılabilir. Verilerden eşit uzaklıkta olacak şekilde en yakındaki verinin üzerinden ana vektöre paralel olacak şekilde çekilen destek vektörlerde amaç en büyük marjin aralığına ulaşabilmektir. Marjin aralığı iki destek vektör arasında ki uzaklığa verilen isimdir.

Hard Marjin ve Soft Marjin

Marjin aralığı iki farklı biçimde tanımlanabilir.

Hard (Sert) marjin ve Soft (Yumuşak) marjin çeşitleridir.

Soft Marjin ve Hard Marjin arasındaki dengeyi sağlayan bir c hiperparametresi bulunur, c ne kadar büyükse marjin aralığı o kadar küçük olur. Eğer model overfittinge uğrarsa c değeri azaltılmalıdır.



Hard (Sert) Marjin

Hard marjin, marjin aralığına giren verilere karşı aşırı duyarlıdır Marjin aralığını mümkün olduğu derecede boş tutmaya çalışır.

Soft (Yumuşak) Marjin

Soft marjin, marjin alacağına giren verilere karşı tolerans gösterebilir. Örneklerden bazıları marjin aralığına girebilir.

Destek Vektör Makinelerinde Düzenleme

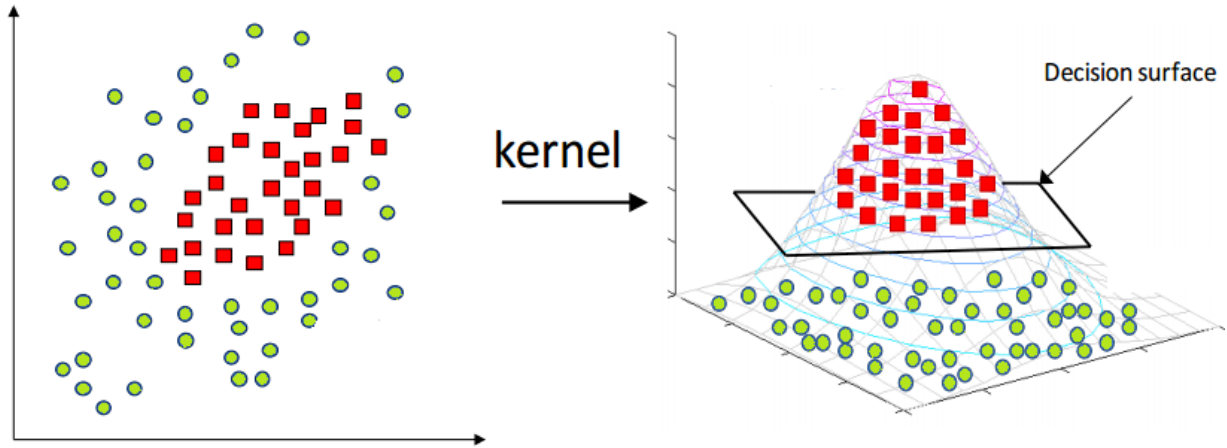
Düzenleme (Regularization) Overfittingin önüne geçmek için yapılan işlemlerdir. Düzenleme parametreleri SVM sınıfından oluşturulan nesnede belirtilir. penalty ve c düzenleme parametreleridir.

Destek Vektör Makinelerinde Kernel Taktiği

Not linearly seperable (Doğru ile ayrılamayan) veriler için destek vektör makinesi normal koşullarda kullanılamaz. Seçeceğimiz karar sınırının doğrusal olması gerekir.

Bu durumun çözümünde kernel taktiği kullanılabilir.

Eğer kullandığımız veri seti doğru ile ayrılamıyorsa bu durumda veri setini 3 boyutlu olarak ayırıp doğru yerine bir düzlemle sınıflandırma yapılabilir.



Gaussian kernel ve rbf (radial basis function) yaygın kernel taktikleridir.

Pythonda SVM

```
#Öncelikle gerekli kütüphane dahil edilir
from sklearn.svm import LinearSVC
#Ardından LinearSVC sınıfından bir nesne oluşturulur
Lr = LinearSVC(penalty = "l2" , c = 10.0, kernel="rbf")
"""
Penalty değeri Regularizasyon(Düzenleme işlemi için atanmış bir değerdir,
Algoritmanın Over fitting (Aşırı öğrenme) 'e yakalanmasını engeller.
c ise regularizasyon katsayısıdır, marjin aralığını belirler.
"""
#Eğitim işlemi yapılır
Lr.fit(xtrain,ytrain)
```

```
#Eğitim yapılan veri seti üzerinden tahmin işlemi yapılır  
tahmin = Lr.predict(xtest)  
#Karşılaştırma için karmaşıklık matrisi kullanılabilir  
from sklearn.metrics import confusion_matrix  
print(confusion_matrix(tahmin,ytest))
```

Mert Aydoğan - Gazi ME