Öznitelik İzdüşümü Kullanarak Artımlı Biçimde Sınıflandırma Öğrenilmesi

T. Aydın, H. A. Güvenir
Bilkent Üniversitesi Bilgisayar Müh. Bölümü, Ankara
atolga@cs.bilkent.edu.tr, guvenir@cs.bilkent.edu.tr

Özetçe

Sınıflandırma öğrenilmesi makine öğrenmesi ve veri tabanı madenciliği disiplinlerinde önemli bir araştırma konusudur. Bu çalışmada ÖİKS (Öznitelik İzdüşümü Kullanılarak Sınıflandırma) öznitelik izdüşümü tabanlı, artımlı biçimde sınıflandırma öğrenen bir algoritma geliştirilmiştir. Algoritmanın öğretme bölümünde nominal öznitelikler için noktalar ve her bir noktaya düşen öğretme örneklerinin sınıflara dağılımı belirlenir. Nümerik özniteliklerde ise her bir sınıf için gaussian olasılık yoğunluğu fonksiyonları oluşturulur. Algoritmanın sorgulama (test) bölümünde ise her bir öznitelik izdüşümü oyunu sınıflar arasında bölüştürür. Özniteliklerin oy vektörleri belirli oy değerlendirme şekillerine göre kullanılıp, sorgu örneğinin sınıfı tahmin edilmeye çalışılır.

Abstract

Classification learning is an important research topic in machine learning and data mining disciplines. In our study, CUFP (Classification by Using Feature Projections), a feature projections-based incremental classification-learning algorithm, was developed. The training phase of CUFP determines points and the distribution of the training instances at each point to the classes in the case of nominal feature projections. For linear feature projections, gaussian probability density functions are constructed for each class. In the classification phase, each feature projection distributes votes among classes. The vote vectors of features are evaluated according to some vote evaluation strategies and the query instance's class is predicted.

1. Giriş

Sınıflandırma öğrenilmesi makine öğrenmesi ve veri tabanı madenciliğinde önemli bir araştırma konusudur. Öznitelik izdüşümü tabanlı sınıflandırma öğrenen algoritmalar özniteliklerin birbirinden bağımsız olduklarını kabul ederek, her bir öznitelik izdüşümünden gelen sınıflandırma sonuçlarını çeşitli biçimlerde birleştirirler. Literatürde öznitelik izdüşümlerinde çeşitli modeller oluşturarak başarılı sonuçlar elde etmiş öznitelik izdüşümü tabanlı yaklaşımlar mevcuttur [1, 2]. Bu çalışmalardaki başarı öyküleri bu alanda yeni bir algoritma olan ÖİKS algoritmasının geliştirilmesine neden olmuştur.

Önceki çalışmalarda nominal özniteliklerde nokta, nümerik özniteliklerde ise geniş değer aralıkları oluşturularak, veri kümesindeki bilgiler temsil edilmektedir. Herhangi bir nokta değer aralığında bu aralığa düşen öğretme örneklerinin olası sınıflara dağılımı bilgisi tutulur. Geniş değer aralıkları ise tıpkı nokta değer aralığında olduğu gibi aralığa düşen öğretme örneklerinin olası sınıflara dağılımı bilgisini tutacak şekilde oluşturulabileceği gibi her bir sınıf için ayrı ayrı homojen olarak da oluşturulabilir. Homojen geniş değer aralıkları yüksek olasılık ile örtüşen değer aralıkları oluşumuna neden olurlar. [4] numaralı kaynakçada belirtilen çalışma bu tür aralıklar oluşturmayı temel alarak sınıflandırma yapmaktadır. Heterojen geniş değer aralıklarında ise örtüşme olmaz. Bu tip değer aralıkları ise aralarında boşluk olup olmama durumlarına göre ikiye ayrılırlar. Aralarında boşluk bulunmayan heterojen geniş değer aralıkları birbirine tek bir noktada yapışık olup, örtüşmezler. [1] numaralı kaynaktaki çalışmada bu tür geniş değer aralıkları kullanılmıştır. [2] numaralı kaynaktaki çalışmada ise aralarında boşluk bulunan heterojen geniş değer aralıkları kullanılmıştır.

ÖİKS sınıflandırma öğrenme algoritması, literatürde yer alan öznitelik izdüşümü tabanlı sınıflandırma algoritmalarından nümerik özniteliklerde geniş değer aralıkları oluşturmayıp, her bir sınıf değeri için gaussian olasılık yoğunluğu fonksiyonu oluşturması bakımından farklılık göstermektedir. Öğretme örneklerinin gaussian olasılık yoğunluğu fonksiyonuna sahip olma olasılığı çok çok düşük olmasına rağmen bu tür bir modellemeye gidilmiştir. Çeşitli veri kümeleri kullanılarak elde edilen deneysel sonuçlar tatmin edici seviyededir.

Literatürde öznitelik izdüşümü tabanlı olmayıp, artımlı sınıflandırma öğrenen algoritmalar mevcuttur [5, 6]. Öznitelik izdüşümü tabanlı çalışmalarda ise sadece homojen geniş değer aralıkları oluşturan modeller artımlı öğrenme yapmakta, diğer heterojen geniş değer aralıkları oluşturan modellerde ise tasarımları itibariyle bu mümkün olmamaktadır. ÖİKS'de herhangi bir nümerik özniteliğindeki gaussian olasılık yoğunluğu fonksiyonları bütün sınıf değerleri için $(-\infty, +\infty)$ arasında tanımlı olduğu için heterojen bir bilgi tutma yöntemi olduğu açıktır. ÖİKS'nin önemli özelliklerinden biri de heterojen bilgi tutma özelliği ile beraber söz konusu fonksiyonların (modelin) artımlı biçimde öğrenilmesini mümkün kılmasıdır.

Herhangi bir sorgu örneğinin sınıflandırılması sırasında her bir öznitelik toplamları '1' olacak şekilde tüm sınıflara oy dağıtan bir oy vektörü sunar. Bu oy vektörünün içeriği sorgu örneğinin öznitelik izdüşümünde hangi noktaya veya değer aralığına düştüğü ile yakından ilgilidir. Literatürdeki çalışmalarda sorgu örneğinin '*c*' gibi bir sınıf değeri için alacağı nihai oy, özniteliklerin sundukları oy vektörlerinin '*c*' sınıfı için verdikleri oyların toplamı ile bulunmaktadır. ÖİKS'de ise '*c*' sınıfı için nihai oyu bulabilmek için bu klasik yöntem ve ayrıca üç yeni yöntem kullanılmıştır.

a) Öznitelik izdüşümlerinden gelen oyları topla.

b) Öznitelik izdüşümlerinden gelen maksimum oyu seç.

c) Öznitelik izdüşümlerinden gelen ortanca (medyan) oyu seç.

d) En yüksek oyu 'c' sınıfına veren öznitelik izdüşümlerinin sayısını 'c' sınıfına verilen nihai oy olarak kullan.

ÖİKS'nin kullandığı üç yeni yöntemde [3] numaralı kaynakçada anlatılan çalışmadan esinlenilmiştir. Sözkonusu çalışmada Naive Bayes sınıflandırma algoritmasının zayıf yönleri anlatılarak, bu yönlerden kurtulmak için bazı basitleştirici varsayımlar yapılmış ve içerik bakımından olmasa da düşünce olarak bu üç yöntemin benzerleri geliştirilmiştir. Genel anlamda yazarlar farklı sınıflandırıcıların tahminlerini birleştirme konusunda çalışmışlardır. Her bir sınıflandırıcı verilen bir örneği farklı ölçüm vektörleriyle temsil etmektedir.

ÖİKS'de ise her bir öznitelik izdüşümü, verilen bir örneği söz konusu öznitelik için aldığı değer ile temsil eden bir sınıflandırıcı olarak kabul edilebilir.

Makalenin organizasyonu şu şekildedir: 2. bölümde ÖİKS'nin öğretme aşaması açıklanmaktadır. 3. bölüm sorgulama aşamasına adanmıştır. Son bölümde gerçek veri kümeleri kullanılarak elde edilmiş deneysel sonuçlar verilmekte ve makale son bulmaktadır.

2. ÖİKS'de Öğretme

ÖİKS algoritmasında öğretme işi artımlı bir şekilde gerçekleştirilir. Nominal öznitelikler için kavram tanımı (model) noktalar ve her bir noktaya düşen öğretme örneklerinin sınıflara dağılımının belirlenmesi şeklinde öğrenilir. Nümerik özniteliklerde ise kavram tanımı her bir sınıf için oluşturulan gaussian olasılık yoğunluğu fonksiyonlarının belirlenmesi şeklinde öğrenilir.

Öğretme aşaması Şekil-1'de gösterilmiştir. $ÖİKS_{öğretme}$ prosedürü girdi parametresi olarak öğretme kümesine henüz yani en son eklenen öğretme örneği *t*'yi ve *oy değerlendirme şekli*'ni almaktadır. Eğer *t* öğretme örneğinin *f* özniteliğine ait değeri olan t_f kayıp, bilinmeyen bir değer değilse söz konusu öznitelikteki kavram tanımı artımlı olarak güncellenir. Kayıp değer olması durumunda, kayıp değere sahip olunan özniteliklerde herhangi bir güncelleme yapılmaz.

Nominal bir *f* özniteliğinde, *noktayı_bul* prosedürü t_f'in öğretme kümesinin o ana kadarki öğretme örneklerinin *f* izdüşümü değerleri arasında bulunup bulunmadığını belirtir. Eğer t_f, *p* gibi bir noktada bulunursa, *t*'nin sınıfının *s* olduğunu kabul edersek, *f* özniteliği izdüşümünün *p* noktasındaki *s* sınıfına sahip öğretme örneklerinin sayısı, yani *nokta_sınıf_örneği_sayısı* [*f, p, s*], bir artırılır. Eğer t_f *f* özniteliği izdüşümünde bulunamazsa, yeni bir *p'* noktası eklenerek bu noktadaki *s* sınıfına ait öğretme örneklerinin sayısı '1', diğer sınıflara ait öğretme örneklerinin sayısı ise '0' olarak belirlenir.

Nümerik bir *f* özniteliğinde $\mu_{f,c}$, *c* sınıfına sahip öğretme örneklerinin *f* izdüşümü değerlerinin ortalamasını; $\mu^2_{f,c}$, *c* sınıfına sahip öğretme örneklerinin *f* izdüşümü değerlerinin karelerinin ortalamasını ve $\sigma_{f,c}$, *c* sınıfına sahip öğretme örneklerinin *f* izdüşümü değerlerinin standart sapmasını temsil etmektedir. Sözkonusu *f* özniteliğinde, *s* sınıfına sahip t_f değeri kayıp olmayan bir *t* öğretme örneği geldiğinde $\mu_{f,s}$ ve $\sigma_{f,s}$ artımlı bir biçimde güncellenerek, *s* sınıfı için oluşturulan gaussian olasılık yoğunluğu fonksiyonu (*gauss_ol_yoğ_fonk*$_{f,s}$) yeniden belirlenir. Standart sapmanın güncellenebilmesi için $\mu^2_{f,c}$ değerinin de güncellenmesi gerekir. Herhangi bir *f* nümerik özniteliğinde *s* sınıfına sahip öğretme örneği sadece *t* ise, standart sapma değeri ve dolayısıyla da gaussian olasılık yoğunluğu fonksiyonu bu özniteliğin *s* sınıfı için tanımsız olur.

ÖİKS'deki öğretme aşamasını Şekil-2'de verilen örnek veri kümesini inceleyerek daha iyi anlayabiliriz. Veri kümesi 10 adet sınıf değeri bilinen öğretme örneğinden ve 1 adet de sınıf değeri tahmin edilecek olan sorgu örneğinden oluşmaktadır. Veri kümemiz f_1 nominal ve f_2 nümerik özniteliklerini içermektedir. Nominal öznitelik 'A' ve 'B' değerlerini, nümerik öznitelik ise bazı tamsayı değerlerini almaktadır. Örnekler 'c_1' ve 'c_2' olmak üzere toplam 2 sınıf değeri almaktadırlar. Ayrıca veri kümemiz kayıp öznitelik değeri içermemektedir.

ÖİKS$_{\textbf{\textit{öğretme}}}$ **(*t*)** /* *t*: henüz eklenen öğretme örneği */
t'nin sınıfı *s* olsun
s dışındaki sınıfların tümü *diğer* olsun
if (*öğretme kümesi* = {*t*}) ise
  for (her bir *f* özniteliği ve her bir *c* sınıfı)
    sınıfa_ait_örnek_sayısı [*f*][*c*] = 0

for (her bir *f* özniteliği)
  if (t_f kayıp bir değer değilse)
    sınıfa_ait_örnek_sayısı[*f*][*s*] ++

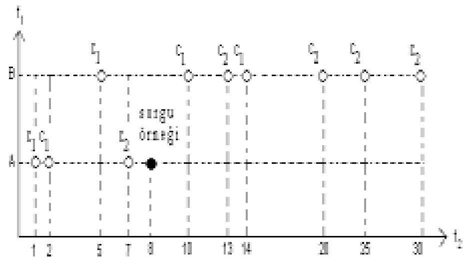
for (her bir *f* özniteliği)
  if (*f* nominal bir öznitelikse ve
    t_f kayıp bir değer değilse)
   p = *noktayı_bul* (*f*,t_f)
   if (*p* gibi bir nokta varsa)
     nokta_sınıf_örneği_sayısı [*f,p,s*] ++
   else /* yeni bir nokta eklenmeli*/
     p' gibi yeni bir nokta ekle
     nokta_sınıf_örneği_sayısı [*f, p',s*] = 1
     nokta_sınıf_örneği_sayısı [*f, p',diğer*] = 0

  else if (*f* nümerik bir öznitelikse ve
    t_f kayıp bir değer değilse)
   if (*sınıfa_ait_örnek_sayısı*[*f*][*s*] = 1) ise
     $\mu_{f,s} = t_f$, $\mu_{f,diğer} = 0$
     $\mu^2_{f,s} = t_f^2$, $\mu^2_{f,diğer} = 0$
     $\sigma_{f,s}$ = *tanımsız*
     gauss_ol_yoğ_fonk$_{f,s}$= *tanımsız*
   else
     n = *sınıfa_ait_örnek_sayısı*[*f*][*s*]
     $\mu_{f,s} = (\mu_{f,s} * (n\text{-}1) + t_f) / n$
     $\mu^2_{f,s} = (\mu^2_{f,s} * (n\text{-}1) + t_f^2) / n$
     $\sigma_{f,s} = \sqrt{\frac{n}{n-1}(\mu^2_{f,s} - (\mu_{f,s})^2)}$
     if ($\sigma_{f,s}$ = 0) ise
      gauss_ol_yoğ_fonk$_{f,s}$= *tanımsız*
     else
      $$gauss\_ol\_yoğ\_fonk_{f,s} = \frac{1}{\sigma_{f,s}\sqrt{2\pi}} e^{-\frac{(x-\mu_{f,s})^2}{2\sigma_{f,s}^2}}$$

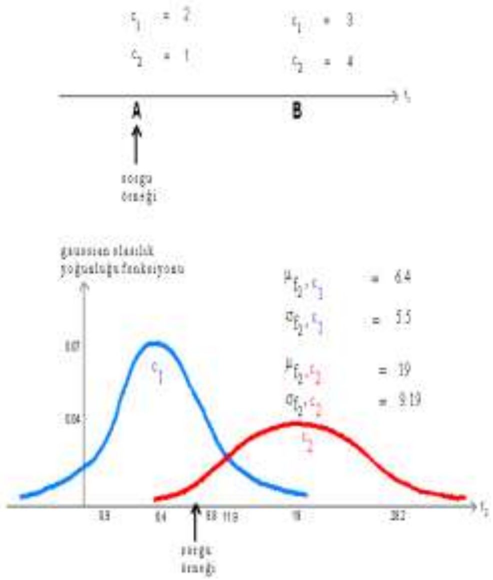
gönder { Nümerik özniteliklerde gauss_ol yoğ_fonk$_{f,c}$ ($\forall$ *f, c*) Nominal özniteliklerde noktalar ve noktalara ait sınıf örneği sayıları }

Şekil-1: ÖİKS'de Artımlı Öğretme

Şekil-3'te sözkonusu veri kümesindeki 10 adet öğretme örneğinden her iki öznitelik için öğrenilen kavram tanımları verilmiştir.



Şekil-2: Örnek Veri Kümesi



Şekil-3: Örnek Veri Kümesi İçin Öğrenilen Kavram Tanımı

3. ÖİKS'de Sorgulama

ÖİKS'nin sorgulama aşamasında her bir öznitelik bağımsız olarak q sorgu örneği için olası sınıf değerleri için oy verir ve özniteliklerden gelen oylar çeşitli oy değerlendirme şekillerine göre değerlendirilerek nihai sınıf oyları elde edilir.

Sorgulama aşaması Şekil-4'te gösterilmiştir. Bu aşama 2 bölüm halinde incelenebilir. 1. bölümde her bir f özniteliğinde ilk önce tüm sınıflara '0' oyu verilir. Eğer q_f kayıp bir değer ise söz konusu f özniteliği q için herhangi bir sınıf tahmininde bulunmaz ve tüm sınıflara daha önce verilen '0' oyları ile süreç sonlandırılır. Eğer q_f kayıp bir değer değilse süreç özniteliğin tipine göre değişiklik gösterir.

Nominal bir f özniteliğinde *noktayı_bul* prosedürü q_f'in öğretme örneklerinin f izdüşümü değerleri arasında bulunup bulunmadığını belirtir. Eğer bulunamazsa aynı şekilde f özniteliği q için herhangi bir sınıf tahmininde bulunmaz ve tüm sınıflara daha önce verilen '0' oyları ile süreç sonlandırılır. Şayet q_f, p gibi bir noktada bulunursa f özniteliği her bir c sınıfı için aşağıdaki denklemde gösterildiği şekilde oy verir:

$$\textit{öznitelik\_oyu}[f,c] = \frac{m}{n}$$

Burada m, f özniteliği izdüşümünün p noktasındaki c sınıfına sahip öğretme örneklerinin sayısını; n ise c sınıfına sahip, f özniteliği değeri bilinen öğretme örneklerinin sayısını temsil etmektedir. Böylelikle '100 * *öznitelik_oyu*[f, c]', f özniteliği değeri bilinen c sınıfına sahip öğretme örneklerinin ne kadarlık bir yüzdesinin f özniteliğinde q_f değerine sahip olduğunu verir. Süreç tüm sınıflara verilen oyların normalize edilmesi ve böylelikle özniteliklerin sınıflandırmada eşit ağırlığa sahip olmasının sağlanması ile sona erer.

Nümerik bir f özniteliğinde öznitelik oyunun hesaplanabilmesi için en azından bir tane c sınıfı için o sınıfa sahip, f özniteliği değeri bilinen öğretme örneklerinin f izdüşümü değerlerinin standart sapmasının tanımlı ve sıfırdan farklı olması yani f özniteliğinde en azından bir tane c sınıfı için gaussian olasılık yoğunluğu fonksiyonu oluşturulabilmesi beklenir. Bu sağlandığı takdirde f özniteliği $\sigma_{f,c}$ değeri tanımlı ve sıfırdan farklı her bir c sınıfı için aşağıdaki denklemde gösterildiği şekilde oy verir:

$$öznitelik\_oyu\,[f,c] = \lim_{\Delta x \to 0} \int_{q_f}^{q_f+\Delta x} g\,dx$$

$$g = \frac{1}{\sigma_{f,c}\sqrt{2\pi}} e^{-\frac{(q_f-\mu_{f,c})^2}{2\sigma_{f,c}^2}}$$

Burada *g*, *f* özniteliği izdüşümünde *c* sınıfı için oluşturulan gaussian olasılık yoğunluğu fonksiyonunda x = q_f için görüntü değerini verir. '$\lim_{\Delta x \to 0} \int_{q_f}^{q_f+\Delta x} g\,dx$' ise söz konusu fonksiyonun x = q_f ve x = q_f +Δx aralığı için Δx sıfıra giderkenki limit alanını, başka bir deyişle *f* özniteliği değeri bilinen *c* sınıfına sahip öğretme örneklerinin ne kadarlık bir oranının *f* özniteliğinde q_f değerine sahip olduğunu verir. Limit alanının sıfır olduğu açıkça görülmektedir. Ancak *f* özniteliğinin bazı sınıf değerlerine verdiği, limit durumunda sıfır olan oylar normalizasyon sırasında Δx terimlerinin sadeleşmesi sonucu gerçek değerlerini alırlar.

Her bir öznitelik her bir sınıf değerine oy verdikten sonra sorgulama aşamasında 2. bölüme geçilir ve herhangi bir *c* sınıfının alacağı nihai oy, dört farklı oy değerlendirme şekline göre belirlenir.

Şekil-5'te nihai oyu bulma prosedürü ve bu prosedür içerisinde de oy değerlendirme şekilleri açıkça gösterilmiştir.

Sorgulama aşamasının bu bölümünde en yüksek nihai oyu sadece tek bir *c* sınıfı almışsa, *q* sorgu örneğinin sınıfı '*c*' olarak tahmin edilir. Aksi durumda sorgu örneği için herhangi bir tahminde bulunulmaz.

Sorgulama aşaması Şekil-2'deki örnek veri kümesindeki öğretme örneklerinden öğrenilen kavram tanımları kullanılarak yine Şekil-2'de gösterilen sorgu örneğinin nasıl sınıflandırıldığının açıklanması ile daha iyi anlaşılabilir.

ÖİKS$_{sorgulama}$ (*q*, *oy değerlendirme şekli*)
/* *q*: sorgu örneği */
for (her bir *f* özniteliği) (1.Bölüm)

for (her bir *c* sınıfı)
öznitelik_oyu[*f*,*c*] = 0

if (*f* nominal bir öznitelikse ve
q_f kayıp bir değer değilse)

p = *noktayı_bul* (*f*,q_f)
if (*p* gibi bir nokta varsa)
for (her bir *c* sınıfı)
m = *nokta_sınıf_örneği_sayısı*[*f*,*p*,*c*]
n = *sınıfa_ait_örnek_sayısı*[*f*][*c*]
if ($n \neq 0$) ise
$öznitelik\_oyu[f,c] = \frac{m}{n}$
öznitelik_oylarını_normalize_et (*f*)
/* öyleki $\sum_C öznitelik\_oyu[f,c] = 1$ */

else if (*f* nümerik bir öznitelikse ve
q_f kayıp bir değer değilse)

if (en azından bir *c* sınıfı için $\sigma_{f,c}$ tanımlı
ve 0'dan farklıysa)

for (her bir $\sigma_{f,c}$'si tanımlı ve 0'dan farklı
c sınıfı)

$$g = \frac{1}{\sigma_{f,c}\sqrt{2\pi}} e^{-\frac{(q_f-\mu_{f,c})^2}{2\sigma_{f,c}^2}}$$

$$öznitelik\_oyu\,[f,c] = \lim_{\Delta x \to 0} \int_{q_f}^{q_f+\Delta x} g\,dx$$

öznitelik_oylarını_normalize_et (*f*)

for (her bir *c* sınıfı) (2.Bölüm)
nihai_oyu_bul(*oy değerlendirme şekli*, *c*)

if (en yüksek nihai oyu tek bir *c* sınıfı almış ise)
q sorgu örneğinin tahmini sınıfı = *c*
else
q sorgu örneği için tahmin yapma

Şekil-4: ÖİKS'de Sorgulama

<A, 8> vektörü ile gösterilen sorgu örneğimizin her iki öznitelik izdüşümünde düştüğü noktalar Şekil 3'de gösterilmiştir.

$$\textit{öznitelik\_oyu}[f_1, c_1] = \frac{2}{5} = 0{,}4$$

$$\textit{öznitelik\_oyu}[f_1, c_2] = \frac{1}{5} = 0{,}2$$

$$\textit{normalleşmiş\_öznitelik\_oyu}[f_1, c_1] = \frac{0{,}4}{0{,}4 + 0{,}2} = 0{,}67$$

$$\textit{normalleşmiş\_öznitelik\_oyu}[f_1, c_2] = \frac{0{,}2}{0{,}4 + 0{,}2} = 0{,}33$$

$$\textit{öznitelik\_oyu}[f_2, c_1] = \lim_{\Delta x \to 0} \int_{8}^{8+\Delta x} \frac{1}{5{,}5\sqrt{2\pi}} e^{-\frac{(8-6{,}4)^2}{2*5{,}5^2}} dx = 0{,}07\Delta x$$

$$\textit{öznitelik\_oyu}[f_2, c_2] = \lim_{\Delta x \to 0} \int_{8}^{8+\Delta x} \frac{1}{9{,}19\sqrt{2\pi}} e^{-\frac{(8-19)^2}{2*9{,}19^2}} dx = 0{,}02\Delta x$$

$$\textit{normalleşmiş\_öznitelik\_oyu}[f_2, c_1] = \frac{0{,}07\Delta x}{0{,}07\Delta x + 0{,}02\Delta x} = 0{,}78$$

$$\textit{normalleşmiş\_öznitelik\_oyu}[f_2, c_2] = \frac{0{,}02\Delta x}{0{,}07\Delta x + 0{,}02\Delta x} = 0{,}22$$

c_1 *sınıfına verilen oylar kümesi* = {0,67, 0,78}

c_2 *sınıfına verilen oylar kümesi* = {0,33, 0,22}

Oy değerlendirme şekli olarak "*Öznitelik İzdüşümlerinden Gelen Oyları Topla"* kullanıldığını kabul edersek:

nihai_oy[c_1] = 0.67 + 0.78 = 1.45

nihai_oy[c_2] = 0.33 + 0.22 = 0.55 olur.

Sorgu örneğimizin tahmin edilen sınıf değeri 'c_1' dir.

---

nihai_oyu_bul (*oy değerlendirme şekli, c*)

ös = öznitelik sayısı
ss = sınıf sayısı

if (*oy değerlendirme şekli* : "*Öznitelik İzdüşümlerinden Gelen Oyları Topla"*) *ise*

$$\textit{nihai\_oy}\ [c] = \sum_{f=1}^{ös} \textit{öznitelik\_oyu}[\,f, c\,]$$

else if (*oy değerlendirme şekli* : "*Öznitelik İzdüşümlerinden Gelen Maksimum Oyu Seç*") ise

$$\textit{nihai\_oy}\ [c] = \max_{f=1}^{ös} \textit{öznitelik\_oyu}[\,f, c\,]$$

else if (*oy değerlendirme şekli* : "*Öznitelik İzdüşümlerinden Gelen Ortanca (Medyan) Oyu Seç*") ise

$$\textit{nihai\_oy}\ [c] = \operatorname{median}_{f=1}^{ös} \textit{öznitelik\_oyu}[\,f, c\,]$$

else if (*oy değerlendirme şekli* : "*En Yüksek Oyu 'c' Sınıfına Veren Öznitelik İzdüşümlerinin Sayısını 'c'Sınıfına Verilen Nihai Oy Olarak Kullan*") ise

for (her bir *f* özniteliği)

$$\textit{maksimum\_oy} = \max_{c=1}^{ss} \textit{öznitelik\_oyu}[\,f, c\,]$$

if (öznitelik_oyu [*f*, *c*] = *maksimum_oy*) ise
nihai_oy [*c*]++

Şekil-5: ÖİKS'de Nihai Oy Bulma

4. Deneysel Sonuçlar

Tablo-1'de ÖİKS öznitelik izdüşümü tabanlı sınıflandırma öğrenme algoritması, bünyesindeki dört farklı oy değerlendirme şeklinden dolayı ODŞ 1, ODŞ 2, ODŞ 3 ve ODŞ 4 adlı versiyonlarla gösterilmektedir. Tabloda bu dört versiyon ve Naive Bayes sınıflandırma algoritması sorgu örneklerinin sınıflarının tahmini sırasında elde edilen doğruluk değerleri bakımından karşılaştırılmaktadır.

Karşılaştırmada 11 adet gerçek hayat veri kümesi kullanılmıştır. Veri kümesinde örneklerin biri sorgu örneği olarak seçilerek, geri kalan örneklerle kavram tanımı öğrenilmiştir. Bu işlem her defasında farklı bir örneği sorgu örneği olarak kullanmak kaydıyla veri kümesindeki örneklerin sayısı kadar tekrarlanarak ortalama bir doğruluk değeri bulunmuştur.

Tablo-1: Veri Kümelerine Ait Doğruluk Tablosu

Veri Küm.	**ODŞ 1**	**ODŞ 2**	**ODŞ 3**	**ODŞ 4**	**Naive Bayes**
Bcan.	**96,14**	93,28	94,13	94,13	95,99
Clev.	82,84	69,64	82,18	82,45	**83,5**
Diab.	75	71,48	71,48	**75,61**	75,26
Echo.	71,62	75,68	75,68	75	**78,38**
Hors.	76,09	**82,34**	70,71	65	77,99
Hung.	84,35	75,51	81,29	83,39	**84,98**
Iris	95,33	**96**	92	94,78	95,33
Sonr.	68,75	58,65	70,67	**70,94**	67,31
Thyr.	97,21	92,09	98,14	**98,52**	96,74
Vote	88,67	**93,33**	87,33	89,55	89
Wine	94,38	90,45	90,45	86,39	**97,19**
Ort. Doğr.	84,58	81,68	83,1	83,25	**85,61**

ÖİKS'nin versiyonları Naive Bayes karşısında yer yer daha başarılı olmuşlardır. Naive Bayes sınıflandırma algoritmasında tek bir öznitelikten gelebilecek bir sıfır olasılık değerinin tüm olasılık değerini sıfır yapma riski vardır. Geliştirilen ÖİKS algoritmasında belirli bir *c* sınıfı için özniteliklerden gelen oylar ile Naive Bayes'in aynı sınıf için elde ettiği olasılık değerleri farklı farklı kavramları temsil etmektedir. Üstelik oylar hiçbir ÖİKS versiyonunda birbirleri ile çarpılmadığı için, Naive Bayes'tekine benzer risklerin taşınmaması büyük bir avantaj olarak ortaya çıkmaktadır.

5. Sonuç

Bu çalışmada ÖİKS adı verilen yeni bir öznitelik izdüşümü tabanlı, artımlı biçimde sınıflandırma öğrenen bir algoritma geliştirilmiş ve başarılı deneysel sonuçlar elde edilmiştir.

ÖİKS sınıflandırma öğrenme algoritması, literatürde yer alan öznitelik izdüşümü tabanlı sınıflandırma algoritmalarından nümerik özniteliklerde geniş değer aralıkları oluşturmayıp, her bir sınıf değeri için gaussian olasılık yoğunluğu fonksiyonu oluşturması yöntemiyle farklılık göstermektedir.

Literatürde yer alan çalışmalarda homojen geniş değer aralığı oluşturan modeller artımlı biçimde oluşturulabilmekte, diğer heterojen geniş değer aralığı oluşturan modellerde ise tasarımları itibariyle bu mümkün olmamaktadır. ÖİKS ise heterojen bilgi tutma özelliği ile beraber söz konusu fonksiyonların (modelin) artımlı biçimde öğrenilmesini mümkün kılmaktadır.

Kaynakça

[1] **Güvenir, H.A. and Demiröz, G.**, 1997. Classification by voting feature intervals, Proceedings of the 9th European Conference on Machine Learning.

[2] **Güvenir, H.A.**, 2003. Benefit Maximization in Classification on Feature Projections, Proceedings of the 3rd IASTED International Conference on Artificial Intelligence and Applications, Malaga, Spain, September.

[3] **Kittler, J., Hatef, M., Duin, R.P.W. and Matas, J.**, 1998. On Combining Classifiers,

IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20.

[4] **Güvenir, H.A. and Koç, H.G.**, 1998. Concept Representation with Overlapping Feature Intervals, Cybernetics and Systems, Vol.29.

[5] **Mandziuk, J. and Shastri, L.**, 1999. Incremental class learning – an approach to longlife and scalable learning, IEEE International Conference on Neural Networks, Vol. 2.

[6] **Diehl, C.P. and Cauwenberghs, G.**, 2003. SVM Incremental Learning, Adaptation and Optimization, International Joint Conference on Neural Networks, Portland OR, July.