

## Denetimli L Kazanç ile İyi Olasılıkların Tahmin Edilmesi

Alexandru Niculescu-Mizil  
Rich Caruana

ALEXN@CS.CORNELL.EDU  
CARUANA@CS.CORNELL.EDU

Bilgisayar Bilimleri Bölümü, Ornell Üniversitesi, Ithaca NY 14853

### Özet

Farklı öğrenme algoritmaları tarafından yapılan tahminler ile gerçek sonsal olasılıklar arasındaki ilişkiyi inceliyoruz. Güçlendirilmiş ağaçlar ve güçlendirilmiş kütükler gibi maksimum marj yöntemlerinin olasılık kütesini 0 ve 1'den uzaklaştırarak tahmin edilen olasılıklarda karakteristik sigmoid şekilli bir bozulmaya yol açtığını gösteriyoruz. Naive Bayes gibi gerçekçi olmayan bağımsızlık varsayımları yapan yöntemler olasılıkları 0 ve 1'e doğru itmektedir. Sinir ağları ve torbalanmış ağaçlar gibi diğer modellerde bu sapmalar yoktur ve iyi kalibre edilmiş olasılıkları tahmin ederler. Bazı öğrenme yöntemleri tarafından öngörülen yanlış olasılıkları düzeltmenin iki yolunu deniyoruz: Platt Ölçekleme ve İzotonik Regresyon. Bu kalibrasyon yöntemlerinin ne tür bozulmalar için uygun olduğunu niteliksel olarak ve etkili olmak için ne kadar veriye ihtiyaç duyduklarını niceliksel olarak inceliyoruz. Ampirik sonuçlar, kalibrasyondan sonra güçlendirilmiş ağaçların, rastgele ormanların ve SV Ms'nin en iyi olasılıkları tahmin ettiğini göstermektedir.

### 1. Giriş

Birçok uygulamada iyi hesaplanmış olasılıkları tahmin etmek önemlidir; iyi doğruluk veya ROC eğrisi altındaki alan yeterli değildir. Bu makale on denetimli öğrenme algoritması tarafından tahmin edilen olasılıkları incelemektedir: SV Ms, sinir ağları, karar ağaçları, bellek tabanlı öğrenme, torbalanmış ağaçlar, rastgele ormanlar, güçlendirilmiş ağaçlar, güçlendirilmiş kütükler, naif bayes ve lojistik regresyon. SV Ms, güçlendirilmiş ağaçlar ve güçlendirilmiş kütükler gibi maksimum marj yöntemlerinin tahmin edilen olasılıkları nasıl 0 ve 1'den uzaklaştırma eğiliminde olduğunu gösteriyoruz. Bu, tahmin ettikleri olasılıkların kalitesine zarar verir ve tahmin edilen olasılıklarda karakteristik sigmoid şekilli bir bozulmaya neden olur. Naif bayes gibi diğer yöntemler ise tam tersi bir önyargıya sahiptir ve tahminleri 0 ve 1'e yaklaştırma eğilimindedir.

Appearing in *Proceedings of the 22<sup>nd</sup> International Conference on Machine Learning*, Bonn, Almanya, 2005. Copyright 2005 yazar(lar)/sahip(ler)i tarafından.

Torbalı ağaçlar ve sinir ağları gibi yöntemlerin önyargıları çok azdır veya hiç yoktur ve iyi kalibre edilmiş olasılıkları tahmin ederler.

Her bir öğrenme yöntemine özgü bozulmayı (veya eksikliğini) inceledikten sonra, bu bozulmaları düzeltmek için iki kalibrasyon yöntemini deniyoruz.

**Platt Ölçeklendirme:** SV M çıktılarını  $[-\infty, +\infty]$ 'dan sonsal olasılıklara dönüştürmek için bir yöntem (Platt, 1999)

**İzotonik Regresyon:** Z adrozny ve E Ikan (2002; 2001) tarafından güçlendirilmiş naif bayes, SV M ve karar ağacı modellerinin tahminlerini kalibre etmek için kullanılan yöntem

Platt Ölçeklemesi, tahmin edilen olasılıklardaki bozulma sigmoid şeklinde olduğunda en etkilidir. İzotonik Regresyon, her türlü monotonik bozulmayı düzeltebilen daha güçlü bir kalibrasyon yöntemidir. Ne yazık ki, bu ekstra gücün bir bedeli vardır. Bir öğrenme eğrisi analizi, İzotonik Regresyonun aşırı uyuma daha yatkın olduğunu ve bu nedenle veri az olduğunda Platt Ölçeklemeden daha kötü performans gösterdiğini göstermektedir.

Son olarak, her bir yöntemin tahminleri kalibre edildikten sonra her bir öğrenme yöntemi tarafından önceden tahmin edilen olasılıkların ne kadar iyi olduğunu inceliyoruz. Sekiz sınıflandırma problemiyle yapılan deneyler, rastgele ormanların, sinir ağlarının ve torbalanmış karar ağaçlarının kalibrasyondan önce iyi kalibre edilmiş olasılıkları tahmin etmek için en iyi öğrenme yöntemleri olduğunu, ancak kalibrasyondan sonra en iyi yöntemlerin güçlendirilmiş ağaçlar, ran-dom ormanları ve SV Ms olduğunu göstermektedir.

### 2. Kalibrasyon Yöntemleri

Bu bölümde, model tahminlerini sonsal olasılıklarla eşleştirmek için iki yöntemi açıklayacağız: Platt Calibration ve İzotonik Regresyon. Ne yazık ki, bu yöntemler ikili sınıflandırma için tasarlanmıştır ve çok sınıflı problemlere genişletilmeleri önemsizdir. Çok sınıflı problemlerle başa çıkmanın bir yolu, onları ikili problemlere dönüştürmek, ikili modelleri kalibre etmek ve ön tahminleri yeniden birleştirmektir (Z adrozny & E Ikan, 2002).

#### 2.1. Platt Calibration

Platt (1999) SV M tahminlerini bir sigmoidden geçirerek

sonsal olasılıklara dönüştürmeyi önermiştir. Bölüm 4'te sigmoid dönüşümün şu şekilde olduğunu göreceğiz

ayrıca güçlendirilmiş ağaçlar ve güçlendirilmiş kütükler için de gerekçelendirilmiştir.

Bir öğrenme yönteminin çıktısı  $f(x)$  olsun. Hesaplanmış olasılıklar elde etmek için çıktığı bir sigmoidden geçirin:

$$P(y = 1|f) = \frac{1}{1 + \exp(Af + B)} \quad (1)$$

B)

Burada  $A$  ve  $B$  parametreleri, uygun bir eğitim setinden  $(f_i, y_i)$  maksimum olabilirlik tahmini kullanılarak uydurulur. Gradyan inişi,  $A$  ve  $B$ 'yi çözüm olacak şekilde bulmak için kullanılır:

$$\argmin_{A,B} \sum_i y_i \log(p_i) + (1 - y_i) \log(1 - p_i) \quad (2)$$

nerede

$$p_i = \frac{1}{1 + \exp(Af_i + B)} \quad (3)$$

İki soru ortaya çıkıyor: sigmoid eğitim seti nereden geliyor? ve bu eğitim setine aşırı uyumdan nasıl kaçınılır?

Kalibre etmek istediğimiz modeli eğitmek için kullanılan veri setinin aynısını kullanırsak istenmeyen yanlışlıklar ortaya çıkar. Örneğin, model eğitim setini tam olarak ayırt etmeyi öğrenirse ve tüm negatif örnekleri pozitif örneklerden önce sıralarsa, sigmoid dönüşümü sadece bir 0,1 fonksiyonu üretecektir. Dolayısıyla, iyi sonsal olasılıklar elde etmek için bağımsız bir kalibrasyon seti kullanmamız gerekir. Ancak bu bir dezavantaj değildir, çünkü aynı küme model ve parametre seçimi için kullanılabilir.

Sigmoid eğitim setine aşırı uyumu önlemek için örneklem dışı bir model kullanılır. Eğitim setinde  $N_+$  pozitif örnek ve  $N_-$  negatif örnek varsa, her eğitim örneği için Platt C alibration  $y_+$  ve  $y_-$  hedef değerlerini kullanır (sırasıyla 1 ve 0 yerine), burada

$$y_+ = \frac{N_+ + 1}{N_+ + 2}; y_- = \frac{1}{N_- + 2} \quad (4)$$

Daha ayrıntılı bir inceleme ve bu özel hedef değerlerin gerekçelendirilmesi için bakınız (Platt, 1999).

## 2.2. İzotonik Regresyon

Sigmoid dönüşümü bazı öğrenme yöntemleri için iyi sonuç verirken diğerleri için uygun değildir. Z adrozny ve E lkan (2002; 2001) SV Ms, Naive Bayes, güçlendirilmiş Naive Bayes ve karar ağaçlarından gelen tahminleri kalibre etmek için İzotonik Regresyona (Robertson vd., 1988) dayanan daha genel bir yöntemi başarıyla kullanmıştır. Bu yöntem daha geneldir çünkü tek kısıtlama eşleme fonksiyonunun izotonik (monotonik olarak artan) olmasıdır. Yani, bir modelden elde edilen tahminler  $f_i$  ve gerçek hedefler  $y_i$  göz önüne alındığında, İzotonik Regresyondaki temel varsayım şudur:

Tablo 1. PAV Algoritması

**Algoritma 1.** Posterior tahmin için PAV algoritması kalibre edilmemiş model tahminlerinden elde edilen olasılıklar.

1 Girdi:  $f'$ 'ye göre sıralanmış eğitim seti  $(f_i, y_i)_{ii}$

2  $m_{i,i}^{\wedge} = y_i, w_{i,i} = 1$  olarak başlatın

3  $\exists i$  s.t.  $m_{k,i-1}^{\wedge} \geq m_{i,i}^{\wedge}$  iken  $w_{k,i} = w_{k,i-1} + w_{i,i}$  olarak ayarlayın

$m_{k,i}^{\wedge} = (w_{k,i-1} m_{k,i-1}^{\wedge} + w_{i,i} m_{i,i}^{\wedge}) / w_{k,i}$  olarak ayarlayın

$m_{k,i-1}^{\wedge}$  ve  $m_{i,i}^{\wedge}$  'i  $m^{\wedge}$  ile değiştirin  $k,i$

4 Kademeli const. fonksiyonunun çıktısını alır:

$$m^{\wedge}(f) = m_{i,i}^{\wedge} \text{ için } f \text{ için } f_i \leq f \leq f_i$$

Burada  $m$  izotonik (monotonik olarak artan) bir fonksiyondur. Daha sonra, bir eğitim seti  $(f_i, y_i)$  verildiğinde, İzotonik Regresyon problemi  $m^{\wedge}$  izotonik fonksiyonunu bulmaktır, öyle ki

$$m^{\wedge} = \argmin_z \sum (y_i - z(f_i))^2 \quad (6)$$

İzotonik Regresyon problemi için kademeli sabit çözüm bulan algoritmalarından biri Tablo 1'de sunulan çift bitişik ihlalciler (PAV) algoritmasıdır (Ayer vd., 1955).

Platt kalibrasyonu durumunda olduğu gibi, İzotonik Regresyon için eğitim setini  $(f(x_i), y_i)$  elde etmek için model eğitim setini  $(x_i, y_i)$  kullanırsak, istenmeyen yanlışlıklar ortaya çıkar. Bu yüzden izotonik fonksiyonu eğitmek için bağımlı bir doğrulama seti kullanırız.

## 3. Veri Setleri

Algoritmaları 8 ikili sınıflandırma problemi üzerinde karşılaştırıyoruz. A DÜLT, C OV TY PE ve L ETTER UC I Repository'den alınmıştır (B lake & Merz, 1998). C OV TY PE, en büyük sınıfı pozitif ve geri kalanını negatif olarak ele alarak ikili bir probleme dönüştürülmüştür. L ETTER'i iki şekilde boolean'a dönüştürdük. L ETTER.p1 "O" harfini pozitif ve geri kalan 25 harfi negatif olarak ele alarak

çok dengesiz bir problem. L ETTER.p2, A -M harflerini pozitifler ve N-Z harflerini negatifler olarak kullanarak zor, ancak

$$y_i = m(f_i) + \epsilon_i \quad (5)$$

iyi dengelenmiş bir problemdir. HS, IndianPine92 veri setidir (Gualtieri vd., 1999) ve zor sınıf olan Soya fasulyesi- nane darısı pozitif sınıftır. SL AC Stanford Linear Accelerator'dan bir problemdir. MEDIS ve MG tıbbi veri setleridir. Veri setleri Tablo 2'de özetlenmiştir.

Tablo 2 Sorunların tanımı

PROBLEM	#ATTR	TRAINSIZE	TESTSIZE	% POZ
ADULT	14/104	4000	35222	25%
COVTYPE	54	4000	25000	36%
LETTER.P1	16	4000	14000	3%
LETTER.P2	16	4000	14000	53%
MEDIS	63	4000	8199	11%
MG	124	4000	12807	17%
SLAC	59	4000	25000	50%
HS	200	4000	4366	24%

#### 4. Tahminlerin Niteliksel Analizi

Bu bölümde farklı öğrenme algoritmalarının kalibrasyonunu niteliksel olarak inceliyoruz. Her bir algoritma için farklı modelleri eğitmek üzere birçok varyasyon ve parametre ayarı kullanıyoruz. Örneğin, on karar ağacı stili, birçok boyutta sinir ağı, birçok çekirdeğe sahip SV Ms vb. kullanarak modelleri eğitiyoruz. Eğitimden sonra, tüm modelleri kalibre etmek için Platt Scal- ing ve İzotonik Regresyon uyguluyoruz. Her model 4000 vakadan oluşan aynı rastgele örneklem üzerinde eğitilir ve 1000 vakadan oluşan bağımsız örneklem üzerinde kalibre edilir. Bu bölümdeki şekillerde, her bir problem ve her bir öğrenme algoritması için, ölçeklendirmeden önce veya sonra en iyi kalibrasyona sahip olan modeli seçiyoruz.

Gerçek koşullu olasılıkların bilinmediği gerçek problemlerde, model kalibrasyonu yeniden sorumluluk diyagramları ile görselleştirilebilir (DeGroot & Fienberg, 1982). İlk olarak, tahmin uzayı on kutuya ayrılır. Tahmin edilen değeri 0 ile 0,1 arasında olan değerler ilk kutuya, 0,1 ile 0,2 arasında olanlar ikinci kutuya, vb. düşer. Her kutu için ortalama tahmin edilen değer, pozitif vakaların gerçek oranına karşı çizilir. Model iyi kalibre edilmişse noktalar köşegen çizginin yakınına düşecektir.

İlk olarak güçlendirilmiş ağaçlar tarafından yapılan tahminleri inceliyoruz. Şekil 1, eğitim veya kalibrasyon için kullanılmayan büyük test setlerindeki sekiz test problemi üzerinde güçlendirilmiş ağaçlar için tahmin edilen değerlerin histogramlarını (üst satır) ve güvenilirlik diyagramlarını (orta ve alt satırlar) göstermektedir. Şekil 1'deki güvenilirlik grafiklerinin ilginç bir yönü, bir sig- sekiz problemin yedisinde sigmoidal şekil<sup>1</sup>, tahminleri kalibre edilmiş hale dönüştürmek için sigmoid kullanımını motive eder olasılıklar. Şeklin orta satırındaki güvenilirlik grafikleri Platt yöntemi kullanılarak uydurulan sigmoidleri göstermektedir. Şeklin alt kısmındaki güvenilirlik grafikleri İzotonik Regresyon ile uydurulan fonksiyonu göstermektedir.

Tahmin edilen değerlerin histogramları incelendiğinde (Şekil 1'de en üst satır), güçlendirilmiş ağaçlar tarafından tahmin edilen neredeyse tüm değerlerin 0 veya 1'e yaklaşan çok az tahminle orta bölgede yer aldığına dikkat edin. Bunun tek istisnası, yalnızca %3 pozitif sınıfa sahip olan ve oldukça çarpık bir veri seti olan L ETTER.P1'dir. Bu problemde tahmin edilen bazı değerler 0'a yaklaşmaktadır, ancak histogramın dikkatli bir şekilde incelenmesi, bu problemde bile 0'a yakın olasılığa sahip olduğu tahmin edilen vakaların sayısında keskin bir düşüş olduğunu göstermektedir. Tahminlerin histogramın merkezine doğru kayması, güçlendirilmiş ağaçların sigmoid şekilli güvenilirlik grafiklerine neden olmaktadır.

Kalibrasyonun tahminleri nasıl dönüştürdüğünü göstermek için sekiz problem için histogramlar ve güvenilirlik diyagramları çiziyoruz

gösterdiğinden, en iyi performans sadece dört iterasyondan sonra elde edilir. Güçlendirmenin daha fazla iterasyon için devam etmesine izin verilirse diğer şekillerde olduğu gibi A DULT üzerinde de aynı sigmoidal şekle sahiptir.

<sup>1</sup>A DULT probleminde boosting aşırı performans

Platt Kalibrasyonu (Şekil 2) ve İzotonik Regresyon (Şekil 3) sonrasında güçlendirilmiş ağaçlar için. Şekiller, kalibrasyonun artırmanın neden olduğu olasılık kütesindeki kaymayı geri aldığını göstermektedir: kalibrasyondan sonra çok daha fazla vaka 0 ve 1'e yakın olasılık tahminine sahiptir. Güvenilirlik diyagramları köşegene daha yakındır ve artırılmış ağaç tahminlerinin S şekli özelliği ortadan kalkmıştır. Her problemde, Platt Ölçekleme veya İzotonik Regresyon kullanarak öngörülerini dönüştürmek, tahmin edilen olasılıklarda önemli bir iyileşme sağlar ve çok daha düşük karesel hata ve log-kaybına yol açar. İzotonik Regresyon ve Platt Ölçekleme arasındaki bir fark histogramlarda belirgindir: İzotonik Regresyon parçalı sabit bir fonksiyon ürettiği için histogramlar kabadır, Platt Ölçekleme tarafından üretilen histogramlar ise daha yumuşaktır. İyi kalibre edilmiş olasılıkların tahmin edilmesi açısından güçlendirmenin daha kapsamlı bir analizi için (Niculescu-Mizil & C aruana, 2005)'e bakınız.

Şekil 6, kalibrasyondan önce ve Platt'ın yöntemiyle kalibrasyondan sonra SL AC problemiindeki on öğrenme yöntemi için tahmin histogramlarını göstermektedir. Platt'ın yöntemi ve Iso- için uygun fonksiyonları gösteren güvenilirlik diyagramları

tonik Regresyon da gösterilmiştir. B oosted stumps ve SV  $Ms^2$  ayrıca belirgin sigmoid şekilli güvenilirlik grafikleri sergilemektedir (Şekil 6'nın sırasıyla ikinci ve üçüncü satırları).

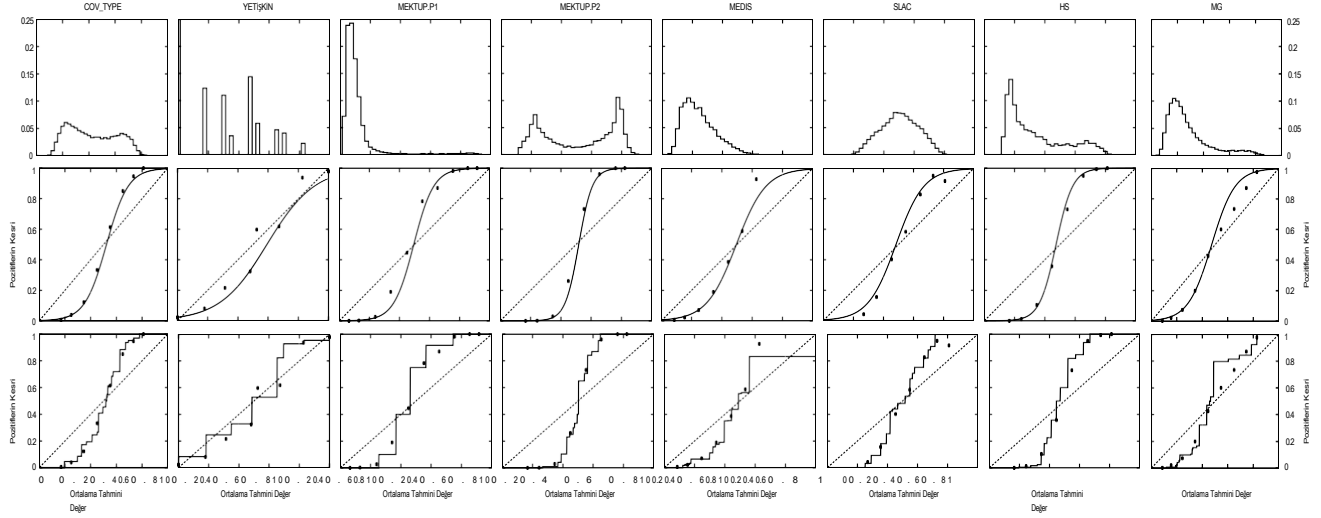
Desteklenmiş kütükler ve SV  $Ms$  diğer yedi problemde benzer davranış sergilemektedir. Güçlendirilmiş ağaçlarda olduğu gibi, güvenilirlik grafiklerinin sigmoidal şekli, tahmin edilen değerlerin histogramlarının merkezindeki kütle yoğunluğu ile birlikte ortaya çıkmaktadır ve güçlendirilmiş kütükler en belirgin olanıdır. Bu davranışı sergileyen öğrenme yöntemlerinin maksimum marjin yöntemleri olduğunu belirtmek ilginçtir. Öngörülerin 0 ve 1'den uzaklaşması sonucu ortaya çıkan sigmoid şekilli güvenilirlik grafiği, maksimum marjin yöntemlerinin karakteristik özelliği olarak görünmektedir.

Sinir ağları için tahmin edilen değerlerin histogramlarını ve güvenilirlik grafiklerini gösteren Şekil 4 çok farklı bir hikaye anlatmaktadır. Güvenilirlik grafikleri, sinir ağlarının başlangıçta iyi kalibre edildiğini ve eğitim sonrası kalibrasyona ihtiyaç duymadığını gösteren diyagonal çizgiyi yakından takip etmektedir. Sadece C OV TY PE problemi kalibrasyondan biraz fayda sağlıyor gibi görünmektedir. Diğer problemlerde her iki kalibrasyon yöntemi de diyagonal çizgiye yaklaşmak için çabılıyor gibi görünmektedir ki bu her ikisi için de doğal olmayan bir görevdir. Bu nedenle, ölçeklendirme sinir ağı kalibrasyonuna biraz zarar verebilir. Platt'ın yöntemiyle eğitilen sigmoidler, Şekil 5'teki histogramlarda görülebileceği gibi, tahminleri 0 ve 1'den uzağa iterek kuyrukları düzgün bir şekilde uydurmakta sorun yaşamaktadır. Şekil 4'teki kalibrasyonsuz sinir ağları için histogramlar, Şekil 2'deki Platt Ölçeklemesinden *sonra* güçlendirilmiş ağaçlar için histogramlara benzer

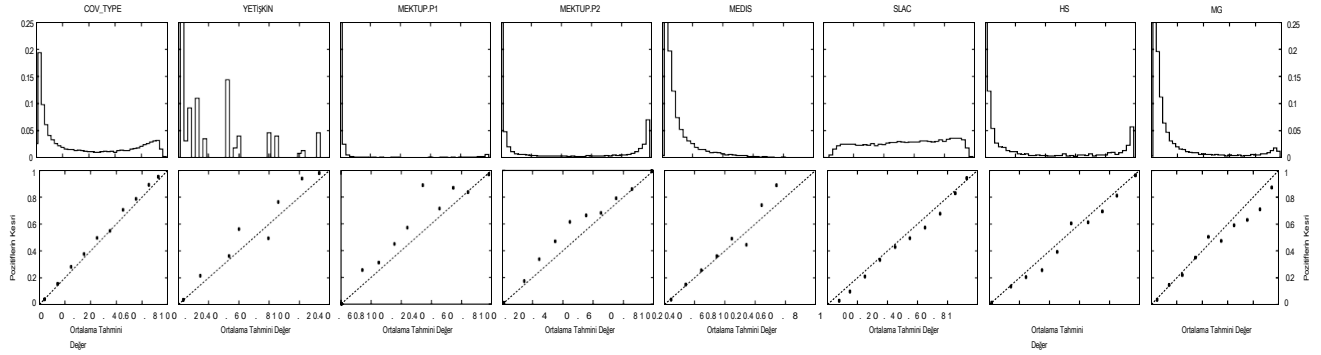
görünmektedir, bu da bize histogramların altta yatan yapıyı yansıttığına dair güven vermektedir.

$^2SV$   $M$  tahminleri  $(x - min)/(maks -)$  ile  $[0,1]$ 'e ölçeklendirilir. *dakika*).

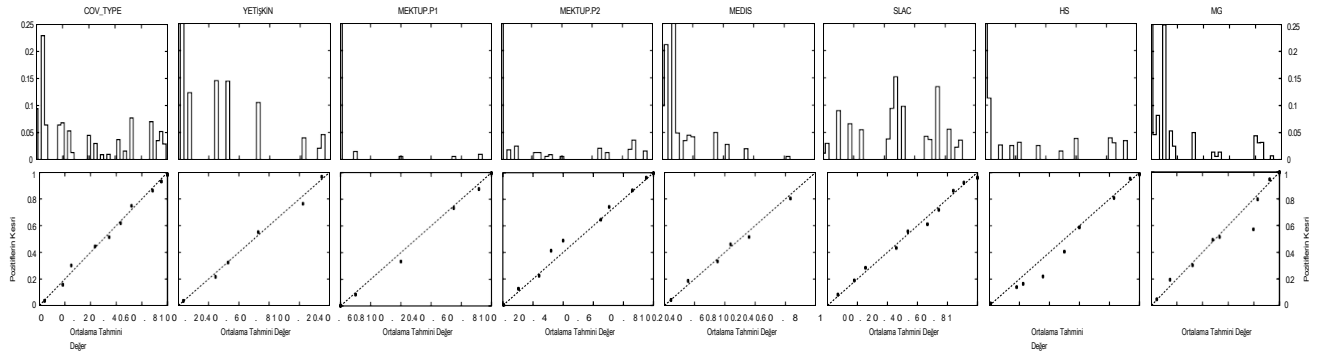
## Denetimli L Kazanç ile İyi Olasılıkların Tahmin Edilmesi



**Şekil 1.** Tahmin edilen değerlerin histogramları ve güçlendirilmiş karar ağaçları için güvenilirlik diyagramları.



**Şekil 2.** Platt yöntemiyle kalibre edilen güçlendirilmiş ağaçlar için tahmin edilen değerlerin histogramları ve güvenilirlik diyagramları.



**Şekil 3.** İzotonik Regresyon ile kalibre edilen güçlendirilmiş ağaçlar için tahmin edilen değerlerin histogramları ve güvenilirlik diyagramları.

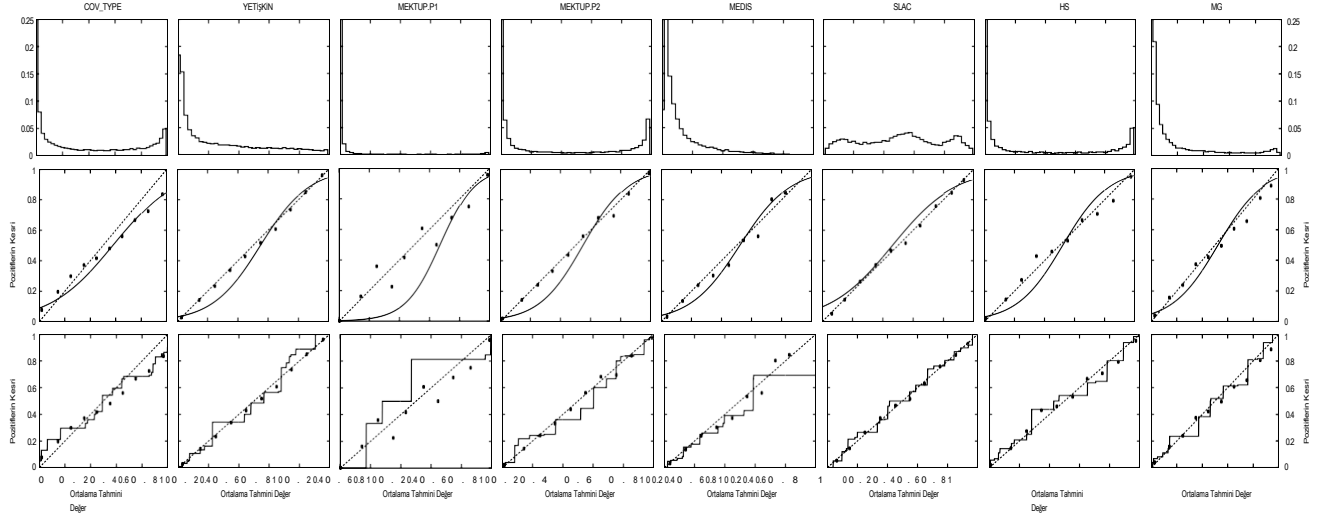
problemlerin yapısı. Örneğin, *mevcut özellikler göz önüne alındığında* L ETTER ve HS problemlerinin "gri" bölgede az sayıda vaka ile iyi tanımlanmış sınıflara sahip olduğu, SL AC probleminde ise iki sınıfın çoğu vaka için önemli belirsizlikle birlikte yüksek örtüşmeye sahip olduğu sonucuna varabiliriz. Tek bir sigmoid çıkış birimine sahip sinir ağlarının, tahminleri kalibre eden çıkışta bir sigmoid ile doğrusal bir sınıflandırıcı (gizli birimlerinin aralığında) olarak görülebileceğini belirtmek ilginçtir. Bu bakımdan

sinir ağları, Platt'ın yöntemi kullanılarak kalibre edildikten sonra SV Ms ve güçlendirilmiş ağaçlara benzer.

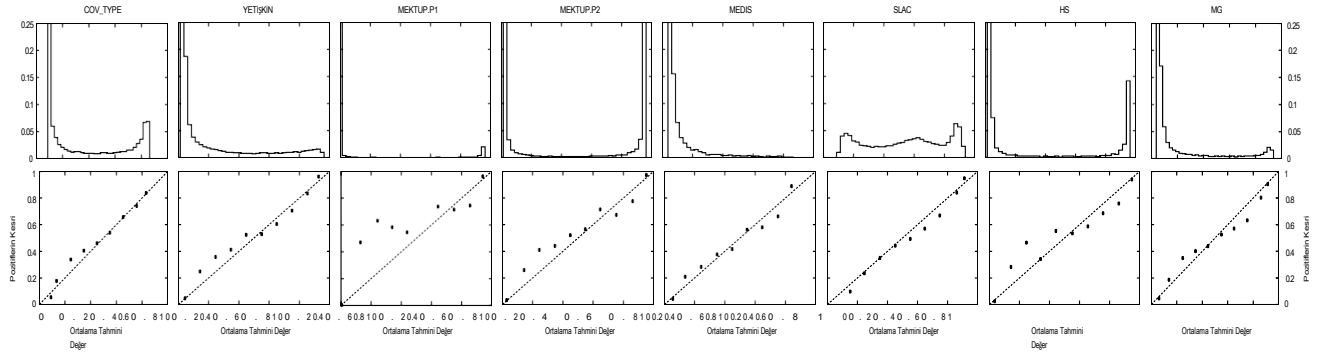
Lojistik regresyon ve torbalanmış ağaçlar için histogramlar ve güvenilirlik diyagramları incelendiğinde sinir ağlarına benzer oldukları görülmektedir. Her iki öğrenme algoritması da başlangıçta iyi kalibre edilmiştir ve sonradan kalibrasyon çoğu problemde onlara yardımcı olmamaktadır. Torbalı ağaçlar MEDİS ve L ETTER.P2 problemlerinde kalibrasyon sonrası biraz yardımcı olmaktadır. Lojistik regresyonun başlangıçta iyi kalibre

edilmesi şaşırtıcı olmasa da





Şekil 4. Sınır ağları için tahmin edilen değerlerin histogramları ve güvenilirlik diyagramları.



Şekil 5. Platt yöntemiyle kalibre edilen sınır ağları için tahmin edilen değerlerin histogramları ve güvenilirlik diyagramları.

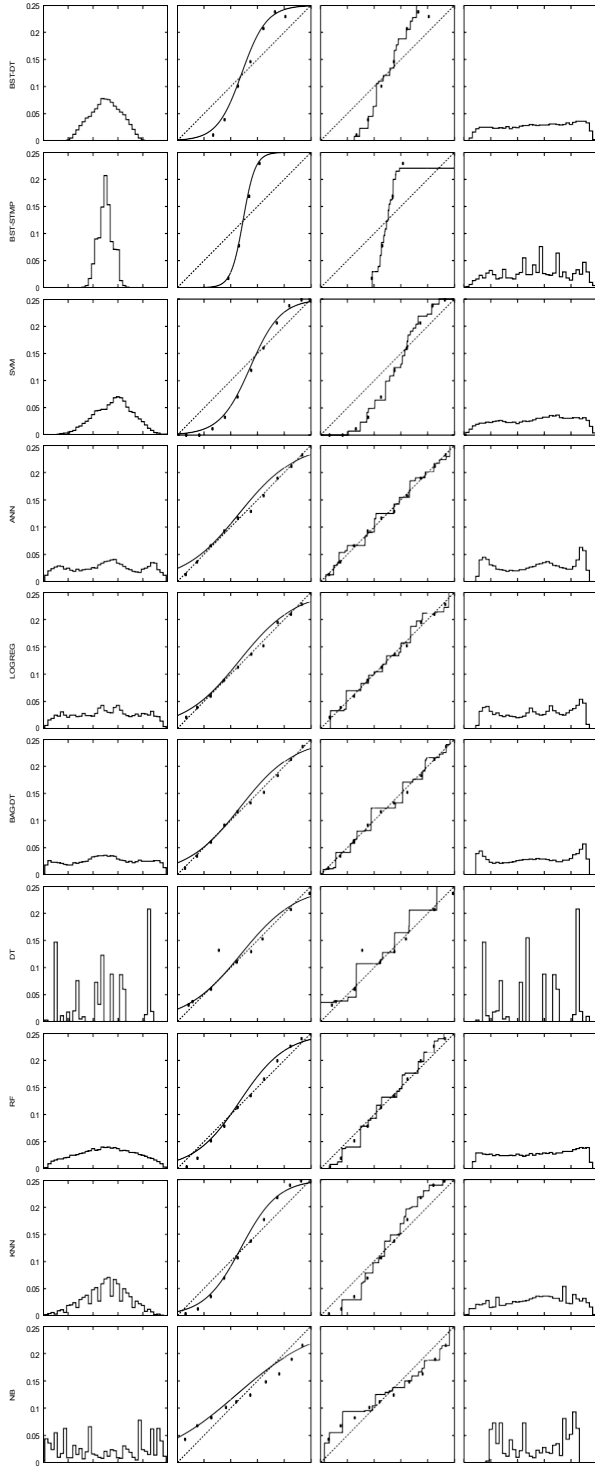
Bu tür iyi kalibre edilmiş olasılıkları gösterdiğinden, karar ağaçlarının torbalanmasının da iyi kalibre edilmiş modeller vermesi ilginçtir. Torbalanmış ağaçların iyi kalibre edildiği göz önüne alındığında, normal karar ağaçlarının da ortalama olarak iyi kalibre edildiği sonucuna varabiliriz, yani karar ağaçları verilerin farklı örnekleri üzerinde eğitilir ve tahminlerinin ortalaması alınırsa, ortalama iyi kalibre edilecektir. Ne yazık ki, tek bir karar ağacı yüksek varyansa sahiptir ve bu varyans kalibrasyonunu etkiler. Platt Ölçekleme bu yüksek varyansla başa çıkamaz, ancak İzotonik Regresyon varyansın yarattığı bazı sorunları çözmeye yardımcı olabilir. Şekil 6'daki beş, altı ve yedinci satırlar SL AC probleminde lojistik regresyon, torbalanmış ağaçlar ve karar ağaçları için histogramları (kalibrasyon öncesi ve sonrası) ve güvenilirlik diyagramlarını göstermektedir.

Rastgele ormanlar daha az nettir. RF modelleri bazı problemlerde iyi kalibre edilmiştir, ancak L ETTER.P2'de kötü kalibre edilmiştir ve HS, C OV TY PE, MEDİS ve L ETTER.P1'de iyi kalibre edilmemiştir. İlginçtir ki bu problemlerde RF'ler, daha düşük bir oranda da olsa, maksimum marj yöntemleriyle aynı davranışı sergilemektedir: önceden tahmin edilen değerler histogramın ortasına doğru hafifçe itilmekte ve güvenilirlik grafikleri sigmoidal bir şekil göstermektedir

(L ETTER problemlerinde daha fazla, C OV TY PE, MEDİS ve HS'de daha az vurgulanmıştır). Bir temel model kümesinden gelen tahminlerin ortalamasını alan torbalama ve rastgele ormanlar gibi yöntemler 0 ve 1'e yakın tahminler yapmakta zorluk çekebilir çünkü temel modellerdeki varyans, sıfıra veya bire yakın olması gereken tahminleri bu değerlerden uzaklaştıracaktır. Tahminler  $[0,1]$  aralığıyla sınırlandırıldığından, varyansın neden olduğu hatalar sıfır ve bir yakınlarında tek taraflı olma eğilimindedir. Örneğin, bir modelin bir vaka için  $p = 0$  tahmininde bulunması gerekiyorsa, torbalama işleminin bunu başarabilmesinin tek yolu *tüm* torbalanmış ağaçların sıfır tahmininde bulunmasıdır. Torbalama işleminin ortalamasını aldığı ağaçlara gürültü eklersek, bu gürültü bazı ağaçların bu vaka için 0'dan büyük değerler tahmin etmesine neden olacak ve böylece torbalanmış topluluğun ortalama tahminini 0'dan uzaklaştıracaktır. Bu etkiyi en güçlü şekilde rastgele ormanlarda gözlemliyoruz çünkü rastgele ormanlarla eğitilen temel düzeydeki ağaçlar, özellik alt kümeleme nedeniyle nispeten yüksek varyansa sahiptir. Kalibrasyon sonrası bu sorunu hafifletmeye yardımcı oluyor gibi görünüyor.

Naive Bayes, sınıf göz önüne alındığında niteliklerin koşullu olarak bağımsız olduğuna dair gerçekçi olmayan bir varsayım yaptığından, tahmin edilen değerleri 0 ve 1'e doğru itme eğilimindedir. Bu, maksimum marj

yöntemlerinin tam tersi bir davranıştır ve



Şekil 6. SL AC için histogramlar ve güvenilirlik diyagramları. SL AC için histogramlar ve güvenilirlik diyagramları.

ters sigmoid şekline sahip güvenilirlik grafikleri oluşturur. Platt Ölçeklendirme kalibrasyonun iyileştirilmesine yardımcı olmaya devam ederken, sigmoidin Naive Bayes modellerini kalibre etmek için doğru dönüşüm olmadığı açıktır. İzotonik Regresyon bu modelleri kalibre etmek için daha iyi bir seçimdir.

Şekil 6'ya dönersek, on farklı modelden kalibrasyon öncesi tahmin edilen değerlerin histogramlarının (ilk sütun) geniş bir varyasyon gösterdiğini görürüz. Maksimum marj yöntemleri (SV M, güçlendirilmiş ağaçlar ve güçlendirilmiş kütükler) tahmin edilen değerleri histogramların merkezinde toplayarak güvenilirlik grafiklerinde sigmoidal bir şekle neden olmaktadır. Hem Platt Ölçekleme hem de İzotonik Regresyon bu sigmoidal şekle uymada etkilidir. Kalibrasyondan sonra tahmin histogramları, tahmin edilen 0 ve 1 değerlerinin yakınındaki kuyruklara doğru uzanır.

İyi kalibre edilmiş yöntemler için (sinir ağları, torbalanmış ağaçlar, rastgele ormanlar ve lojistik regresyon), Platt Ölçekleme ile kalibrasyon aslında olasılık kütlelerini 0 ve 1'den uzaklaştırır. Bu yöntemlerin güvenilirlik diyagramlarına bakıldığında, sigmoidin bu iyi kalibre edilmiş yöntemlerin kuyruklarındaki tahminlere uymakta zorluk çektiği açıktır.

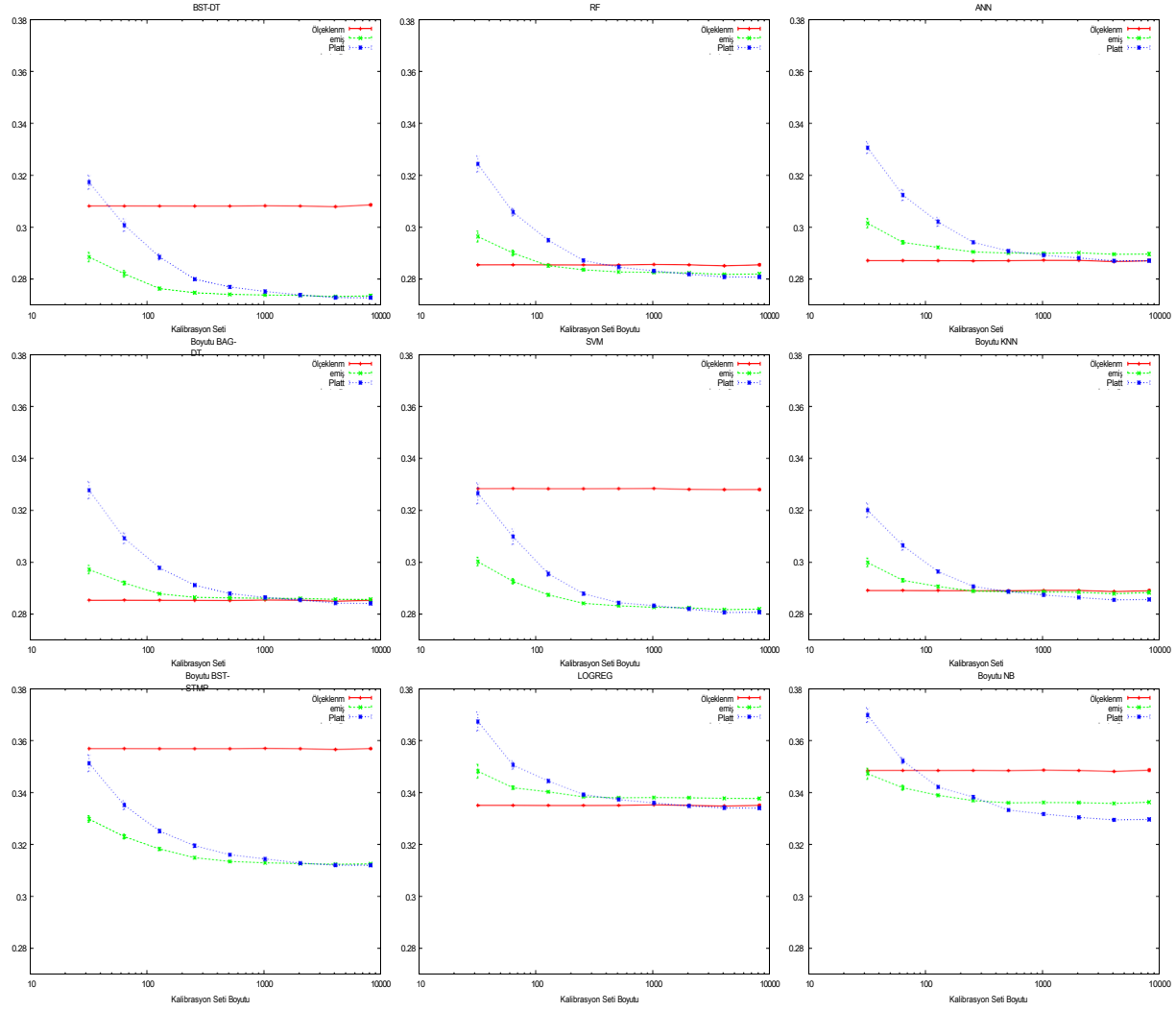
Genel olarak, kalibrasyon öncesi ve sonrası olasılık histogramları incelendiğinde, Platt Ölçeklendirmesi sonrasında histogramların birbirine çok daha benzer olduğu görülmektedir. Kalibrasyon, farklı modeller tarafından tahmin edilen olasılıklar arasındaki farkları önemli ölçüde azaltmaktadır. Elbette kalibrasyon, karar ağaçları ve naif bayes gibi daha eski modellerden gelen tahminleri tam olarak düzeltememektedir.

## 5. L kazanç C urve Analizi

Bu bölümde iki kalibrasyon yöntemi olan Platt Ölçekleme ve İzotonik Regresyon için bir öğrenme eğrisi analizi sunuyoruz. Amaç, kalibrasyon için mevcut veri miktarı değiştiğinde bu kalibrasyon yöntemlerinin ne kadar etkili olduğunu belirlemektir. Bu analiz için Bölüm 4'teki aynı modelleri kullanıyoruz, ancak burada kalibrasyon setinin boyutunu 32 vakadan 8192 vakaya kadar iki kat değiştiriyoruz. Kalibrasyon performansını ölçmek için modellerin karesel hatalarını inceliyoruz.

Şekil 7'deki grafikler sekiz test problemi üzerindeki ortalama karesel hatayı göstermektedir. Her bir problem için on deneme yapılmıştır. Hata çubukları grafiklerde gösterilmiştir, ancak o kadar azdır ki görülmeleri zor olabilir. On öğrenme yönteminden dokuzu için kalibrasyon öğrenme eğrileri gösterilmiştir (karar ağaçları hariç tutulmuştur).

Grafiklerdeki neredeyse yatay çizgiler kalibrasyondan önceki karesel hatayı göstermektedir. Bu çizgilerin tam yatay olmamasının tek nedeni kalibrasyon setlerine daha fazla veri taşındıkça test setlerinin değişmesidir. Her bir grafik, kalibrasyon setinin boyutu küçükten büyüğe değiştiğinde Platt yöntemi veya İzotonik Regresyon ile kalibrasyondan sonraki karesel hatayı göstermektedir. Kalibrasyon seti küçük olduğunda (yaklaşık 200-1000 vakadan az), Platt Ölçekleme dokuz öğrenme yönteminin tümüyle İzotonik Regresyondan daha iyi performans gösterir. Bunun nedeni, İzotonik Regresyonun, İzotonik Regresyona göre daha az kısıtlanmış olmasıdır.



Şekil 7. Platt Ölçekleme ve Platt Ölçeklendirme ve İzotonik Regresyon için L kazanç C urves (8 problemin ortalaması).

Platt Ölçeklendirme, bu nedenle kalibrasyon seti küçük olduğunda aşırı uyum sağlaması daha kolaydır. Platt'ın yönteminde ayrıca bazı aşırı uyum kontrolleri de bulunmaktadır (bkz. Bölüm 2). Kalibrasyon setinin boyutu arttıkça, Platt Ölçekleme ve İzotonik Regresyon için öğrenme eğrileri birleşir, hatta kesişir. Kalibrasyon setinde 1000 veya daha fazla nokta olduğunda, İzotonik Regresyon her zaman Platt Ölçekleme kadar iyi veya ondan daha iyi performans sağlar.

Sinir ağları, torbalanmış ağaçlar ve lojistik regresyon gibi iyi kalibre edilmiş tahminler yapan öğrenme yöntemleri için ne Platt Ölçekleme ne de İzotonik Regresyon, kalibrasyon seti çok büyük olsa bile performansta çok fazla iyileşme sağlamaz. Bu yöntemlerde kalibrasyon faydalı değildir ve aslında kalibrasyon setleri küçük olduğunda performansa zarar verir.

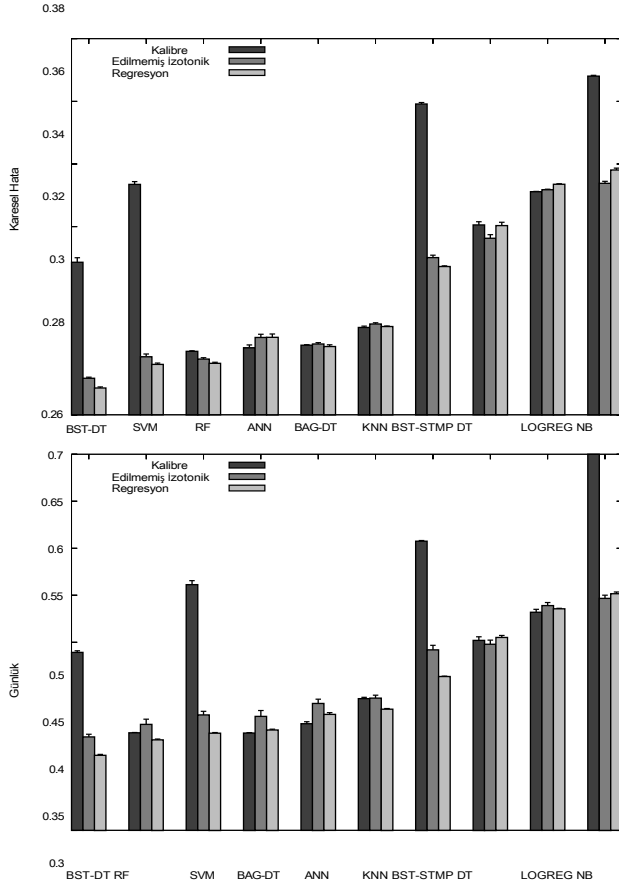
Maksimum marj yöntemleri, güçlendirilmiş ağaçlar, güçlendirilmiş kütükler ve SV Ms için kalibrasyon bir iyileşme sağlar

Kalibrasyon seti küçük olduğunda bile. Bölüm 4'te sigmoidin güçlendirilmiş ağaçlar, güçlendirilmiş kütükler ve SV Ms için iyi bir eşleşme olduğunu gördük. Beklendiği gibi, bu yöntemler için Platt Ölçekleme, küçük ve orta ölçekli kalibrasyonlarda (1000'den az vaka) İzotonik Regresyondan daha iyi performans gösterir ve daha büyük kalibrasyon setleri için neredeyse ayırt edilemez.

Beklendiği gibi, kalibrasyon Naive Bayes modellerinin performansını neredeyse tüm kalibrasyon seti boyutları için artırırken, daha fazla veri olduğunda İzotonik Regresyon Platt Ölçekleme'den daha iyi performans göstermektedir. Modellerin geri kalanı için: K NN, RF ve DT (gösterilmemiştir) kalibrasyon sonrası, kalibrasyon setleri yeterince büyük olduğunda yardımcı olur.

## 6. Ampirik Karşılaştırma

Daha önce olduğu gibi, her bir öğrenme algoritması için farklı parametre ayarları kullanarak farklı modelleri eğitiyor ve kalibre ediyoruz



Şekil 8. Öğrenme algoritmaları Öğrenme algoritmalarının performansı

her model İzotonik Regresyon ve Platt Ölçeklendirme ile. Modeller 4k örnek üzerinde eğitilmiş ve bağımsız 1k örnek üzerinde kalibre edilmiştir. Her veri seti, öğrenme algoritması ve kalibrasyon yöntemi için, kalibrasyon için kullanılan aynı 1 bin noktayı kullanarak en iyi performansa sahip modeli seçiyoruz ve büyük nihai test setindeki performansını rapor ediyoruz.

Şekil 8'de kalibrasyon öncesi ve sonrası her bir öğrenme yöntemi için karesel hata (üstte) ve log-kayıp (altta) gösterilmektedir. Her bir çubuk, sekiz problemin her biri üzerinde yapılan beş denemenin ortalamasını almaktadır. Ortalamalar için 1 standart sapmayı temsil eden hata çubukları gösterilmiştir. Dört öğrenme yöntemi - güçlendirilmiş ağaçlar, SV Ms, güçlendirilmiş kütükler ve naif bayes - tarafından tahmin edilen olasılıklar kalibrasyon ile önemli ölçüde iyileştirilmiştir. Kalibrasyon torbalı ağaçlara yardımcı olmaz ve aslında sinir ağlarına zarar verir. Kalibrasyondan önce en iyi modeller rastgele ormanlar, torbalanmış ağaçlar ve sinir ağlarıdır. Ancak kalibrasyondan sonra, güçlendirilmiş ağaçlar, rastgele ormanlar ve SV Ms en iyi olasılıkları tahmin etmektedir.

## 7. Sonuçlar

Bu makalede on farklı öğrenme yöntemi tarafından tahmin edilen olasılıkları inceledik. Boosting ve SV Ms gibi maksimum marj metotları tahminlerinde karakteristik

bayes tam tersi bir bozulma ile tahminler yapar. Sinir ağları ve torbalanmış ağaçlar gibi yöntemler ise iyi kalibre edilmiş olasılıklar öngörmektedir. Farklı öğrenme yöntemleri tarafından yapılan tahminleri kalibre etmek için Platt Ölçekleme ve İzotonik Regresyonun etkinliğini inceledik. Platt Ölçekleme en çok veri küçük olduğunda etkilidir, ancak İzotonik Regresyon aşırı uyumu önlemek için yeterli veri olduğunda daha güçlüdür. Kalibrasyondan sonra, en iyi olasılıkları önceden belirleyen modeller güçlendirilmiş ağaçlar, rastgele ormanlar, SV Ms, kalibre edilmemiş torbalı ağaçlar ve kalibre edilmemiş sinir ağlarıdır.

## ACKNOWLEDGMENTS

B. Z adrozny ve C. E Ikan'a İzotonik Re- gresyon kodu için, C. Young ve arkadaşlarına, SL AC verileri için Stanford Linear Accelerator'a ve Indian Pines Verileri ile ilgili yardımları için Goddard Space Center'dan A. Gualtieri'ye teşekkür ederiz. Bu çalışma NSF Award 0412930 tarafından desteklenmiştir.

## Referanslar

- Ayer, M., Brunk, H., Ewing, G., Reid, W., & Silverman, E. (1955). Eksik bilgi ile örnekleme için bir n ampirik dağılım fonksiyonu. *Annals of Mathematical Statistics*, 5, 641-647.
- B lake, C., & Merz, C. (1998). UC I makine deposu farklılıklar ortaya koymaktadır. Naif öğrenme gibi diğer yöntemler

öğrenme veritabanları.

- DeGroot, M., & Fienberg, S. (1982). Tahmincilerin karşılaştırılması ve değerlendirilmesi. *Statistician*, 32, 12-22.
- Gualtieri, A., C hettri, S. R., C romp, R., & Johnson, L. (1999). Aviris verilerine uygulanan destek vektör makinesi sınıflandırıcıları. *Proc. E ighth JPL Airborne Geoscience Workshop*.
- Niculescu-Mizil, A., & C aruana, R. (2005). Güçlendirmeden kalibre edilmiş olasılıklar elde etme. *Proc. 21th C onfer- ence on Uncertainty in Artificial Intelligence (UAI '05)*. AUA I Press.
- Platt, J. (1999). Destek vektör ma- kineleri için olasılıksal çıktılar ve düzenlenmiş olabilirlik yöntemleriyle karşılaştırma. *Advances in Large Margin C lassifiers* (s. 61-74).
- Robertson, T., Wright, F., & Dykstra, R. (1988). *Sıra kısıtlı istatistiksel çıkarım*. New York: John Wiley and Sons.
- Z adrozny, B., & E lkan, C. (2001). Karar ağaçları ve naif bayes sınıflandırıcılarından kalibre edilmiş olasılık tahminlerinin elde edilmesi. *IC ML* (s. 609-616).
- Z adrozny, B., & E lkan, C. (2002). Sınıflandırma puanlarının doğru çok sınıflı olasılık tahminlerine dönüştürülmesi. *K DD* (s. 694-699).