

Kolektif Öğrenme Metotları

- Kolektif öğrenme algoritmalarına genel bakış
- 1-Bagging
- 2-Ardışık Topluluklarla Öğrenme (Boosting)
- 3-Rastsal Altuzaylar
- 4-Rastsal Ormanlar
- 5-Aşırı Rastsal Ormanlar
- 6-Rotasyon Ormanları
- 7-Hata düzeltlenen kod tabanlı topluluklar
- 8-Yerel Modellerin Birlesimi
- 9-İki Katmanlı Kolektif Öğrenme – Model sonuçlarını modelleme

YILDIZ TEKNİK ÜNİVERSİTESİ
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



Kolektif öğrenme algoritmalarına genel bakış

- Amaç: Temel öğrencilerin karar çeşitliliğini sağlamak
- Nasıl ?

YILDIZ TEKNİK ÜNİVERSİTESİ
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



Farkı nasıl sağlarız?

- Farklı temel öğreniciler
- Aynı öğrencinin farklı hiperparametreli halleri (KNN'de farklı K'lar, YSA'da farklı başlangıçlar ve gizli katmana farklı sayıarda nöron koymak vb.)
- Aynı öğrenciyi farklı eğitim setleri ile eğitmek

YILDIZ TEKNİK ÜNİVERSİTESİ
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



Farklı eğitim setlerini nasıl üretiriz?

- Neler değişimdir?
- Örnekler
 - Bagging
 - AdaBoost
 - Rastsal Ormanlar (Random Forest)
- Özellikler
 - Rastsal Altzaylar (Random Subspaces)
 - Rastsal Ormanlar (Random Forest)
 - Rotasyon Ormanları (Rotation Forest)
- Sınıflar
 - Ensemble of Nested Dichotomies
 - ECOC
- Bölgeler
 - Uzman karışıntıları
 - Random Linear Oracle
- Model sonuçlarını modelleme (Stacking)

YILDIZ TEKNİK ÜNİVERSİTESİ
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



1-Bagging

- Yerine koyarak örnekleme

Data ID	1	2	3	4	5	6	7	8	9	10
Original Data	1	2	3	4	5	6	7	8	9	10
Bagging (Round 1)	7	8	10	8	2	5	10	10	5	9
Bagging (Round 2)	1	4	9	1	2	3	2	7	3	2
Bagging (Round 3)	1	8	5	10	5	5	9	6	3	7

Training Data

- Her seçilen n örnekle temel öğrenciyi eğit
- Seçilen n örnek tüm örneklerin yaklaşık %63.2 sini içerir ?

$$\left(1 - \frac{1}{n}\right)^n \approx e^{-1} = 0.368$$



5

Algorithm: Bagging

Input:

- Training data S with correct labels $\omega_i \in \Omega = \{\omega_1, \dots, \omega_C\}$ representing C classes
- Weak learning algorithm **WeakLearn**,
- Integer T specifying number of iterations.
- Percent (or fraction) F to create bootstrapped training data

Do $t = 1, \dots, T$

1. Take a bootstrapped replica S_t by randomly drawing F percent of S .
2. Call **WeakLearn** with S_t and receive the hypothesis (classifier) h_t .
3. Add h_t to the ensemble, Ξ .

End

Test: Simple Majority Voting – Given unlabeled instance x

1. Evaluate the ensemble $\Xi = [h_1, \dots, h_T]$ on x .
2. Let $v_{t,j} = \begin{cases} 1, & \text{if } h_t \text{ picks class } \omega_j \\ 0, & \text{otherwise} \end{cases}$ (8)

be the vote given to class ω_j by classifier h_t .

3. Obtain total vote received by each class

$$V_j = \sum_{t=1}^T v_{t,j}, \quad j = 1, \dots, C \quad (9)$$

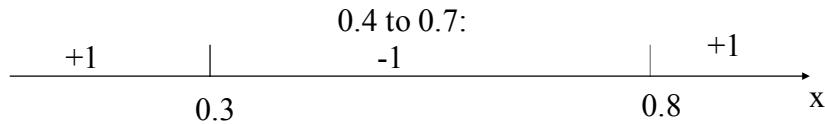
4. Choose the class that receives the highest total vote as the final classification.

BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



Bagging Örnek

Egitim verisi:



Hedef: 10 temel sınıflandırıcıdan oluşan Bagging Ensemble.

-Her temel sınıflandırıcı bir eşik (K) değerine sahip:

$(x \leq K \rightarrow \text{class} = +1 \text{ ya da } -1;$

K toplam entropinin minimizasyonu ile seçilir)



Bagging Round 1:												$x \leq 0.35 \Rightarrow y = 1$	$x > 0.35 \Rightarrow y = -1$
x	0.1	0.2	0.2	0.3	0.4	0.4	0.5	0.6	0.9	0.9			
y	1	1	1	1	-1	-1	-1	-1	1	1			

Bagging Round 2:												$x \leq 0.65 \Rightarrow y = 1$	$x > 0.65 \Rightarrow y = -1$
x	0.1	0.2	0.3	0.4	0.5	0.8	0.9	1	1	1			
y	1	1	1	-1	-1	1	1	1	1	1			

Bagging Round 3:												$x \leq 0.35 \Rightarrow y = 1$	$x > 0.35 \Rightarrow y = -1$
x	0.1	0.2	0.3	0.4	0.4	0.5	0.7	0.7	0.8	0.9			
y	1	1	1	-1	-1	-1	-1	-1	1	1			

Bagging Round 4:												$x \leq 0.3 \Rightarrow y = 1$	$x > 0.3 \Rightarrow y = -1$
x	0.1	0.1	0.2	0.4	0.4	0.5	0.5	0.7	0.8	0.9			
y	1	1	1	-1	-1	-1	-1	-1	1	1			

Bagging Round 5:												$x \leq 0.35 \Rightarrow y = 1$	$x > 0.35 \Rightarrow y = -1$
x	0.1	0.1	0.2	0.5	0.6	0.6	0.6	1	1	1			
y	1	1	1	-1	-1	-1	-1	1	1	1			

Bagging Round 6:												$x \leq 0.75 \Rightarrow y = -1$	$x > 0.75 \Rightarrow y = 1$
x	0.2	0.4	0.5	0.6	0.7	0.7	0.7	0.8	0.9	1			
y	1	-1	-1	-1	-1	-1	-1	1	1	1			

Bagging Round 7:												$x \leq 0.75 \Rightarrow y = -1$	$x > 0.75 \Rightarrow y = 1$
x	0.1	0.4	0.4	0.6	0.7	0.8	0.9	0.9	0.9	1			
y	1	-1	-1	-1	-1	1	1	1	1	1			

Bagging Round 8:												$x \leq 0.75 \Rightarrow y = -1$	$x > 0.75 \Rightarrow y = 1$
x	0.1	0.2	0.5	0.5	0.5	0.7	0.7	0.8	0.9	1			
y	1	1	-1	-1	-1	-1	-1	1	1	1			

Bagging Round 9:												$x \leq 0.75 \Rightarrow y = -1$	$x > 0.75 \Rightarrow y = 1$
x	0.1	0.3	0.4	0.4	0.6	0.7	0.7	0.8	1	1			
y	1	1	-1	-1	-1	-1	-1	1	1	1			

Bagging Round 10:												$x \leq 0.05 \Rightarrow y = -1$	$x > 0.05 \Rightarrow y = 1$
x	0.1	0.1	0.1	0.1	0.3	0.3	0.8	0.8	0.9	0.9			
y	1	1	1	1	1	1	1	1	1	1			

Figure 5.35. Example of bagging.



Bagging (kolektif kararın üretimi)

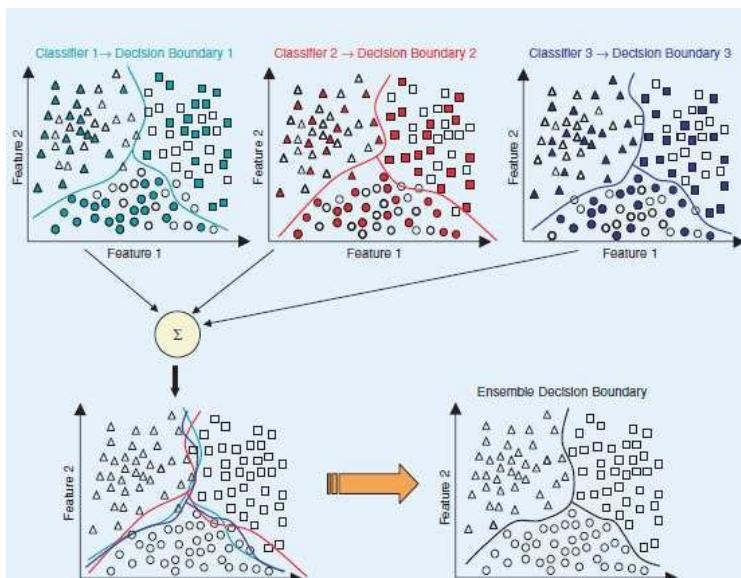
Round	$x=0.1$	$x=0.2$	$x=0.3$	$x=0.4$	$x=0.5$	$x=0.6$	$x=0.7$	$x=0.8$	$x=0.9$	$x=1.0$
1	1	1	1	-1	-1	-1	-1	-1	-1	-1
2	1	1	1	1	1	1	1	1	1	1
3	1	1	1	-1	-1	-1	-1	-1	-1	-1
4	1	1	1	-1	-1	-1	-1	-1	-1	-1
5	1	1	1	-1	-1	-1	-1	-1	-1	-1
6	-1	-1	-1	-1	-1	-1	-1	1	1	1
7	-1	-1	-1	-1	-1	-1	1	1	1	1
8	-1	-1	-1	-1	-1	-1	1	1	1	1
9	-1	-1	-1	-1	-1	-1	1	1	1	1
10	1	1	1	1	1	1	1	1	1	1
Sum	2	2	2	-6	-6	-6	-6	2	2	2
Sign	1	1	1	-1	-1	-1	-1	1	1	1
True Class	1	1	1	-1	-1	-1	-1	1	1	1

Figure 5.36. Example of combining classifiers constructed using the bagging approach.

ensemble basarisi: 100% 😊



2 özellikte Bagging



YILDIZ TEKNİK ÜNİVERSİTESİ
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



Bagging- Bulgular

- Temel öğreniciler eğitim verisine duyarlı iseler iyi çalışır. 1NN ? DT? ANN?
- Spesifik örneklerde yoğunlaşmadığından gurultulu verilere karşı gürbüz.

11



Ağaçları tamamen büyütürsek düşük başarı, yüksek çeşitlilik, daha genel bırakırsak yüksek başarı, az çeşitlilik

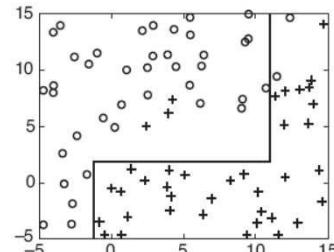
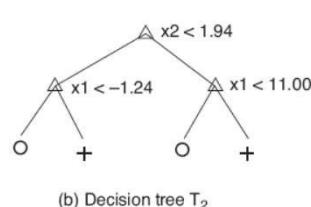
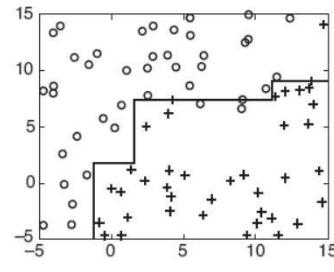
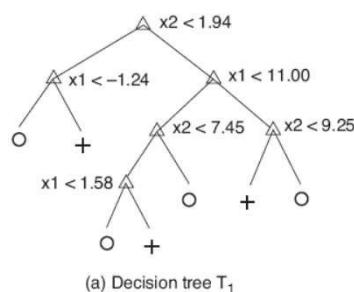


Figure 5.33. Two decision trees with different complexities induced from the same training data.

12



Bagging Classification Trees

MISCLASSIFICATION RATES (%)

Data Set	\bar{e}_S	\bar{e}_B	Decrease
waveform	29.1	19.3	34%
heart	4.9	2.8	43%
breast cancer	5.9	3.7	37%
ionosphere	11.2	7.9	29%
diabetes	25.3	23.9	6%
glass	30.4	23.6	22%
soybean	8.6	6.8	21%

S : tek ağaç

B : çok ağaç

MISCLASSIFICATION RATES FOR NEAREST NEIGHBOR

Data Set	\bar{e}_S	\bar{e}_B
waveform	26.1	26.1
heart	5.1	5.1
breast cancer	4.4	4.4
ionosphere	36.5	36.5
diabetes	29.3	29.3
glass	30.1	30.1

YILDIZ TEKNİK ÜNİVERSİTESİ
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ

2-Ardışık Topluluklarla Öğrenme (Boosting)

- Temel öğrenicilerin eğitileceği örneklerin seçiminde önceki temel öğrenicilerin hata yaptıkları örneklerle öncelik veriyor.
 - Başlangıçta tüm örneklerin eğitim kümesine seçilme olasılıkları aynı
 - Her iterasyonda seçilme olasılıkları güncelleniyor (Bagging'te aynı kalıyordu)



AdaBoost

- Adaboost iterasyonlarında 4 numaralı örneğin sınıflandırılması zor ise, iterasyon sayısı arttıkça 4 numaralı örneğin seçilme şansı (temel sınıflandırıcının eğitiminde kullanılma) artıyor.

Original Data	1	2	3	4	5	6	7	8	9	10
Boosting (Round 1)	7	3	2	8	7	9	4	10	6	3
Boosting (Round 2)	5	4	9	4	2	5	1	7	4	2
Boosting (Round 3)	4	4	8	10	4	5	4	6	3	4

15



AdaBoost

Training:

For all $\{x^t, r^t\}_{t=1}^N \in \mathcal{X}$, initialize $p_1^t = 1/N$

For all base-learners $j = 1, \dots, L$

Randomly draw \mathcal{X}_j from \mathcal{X} with probabilities p_j^t

Train d_j using \mathcal{X}_j

For each (x^t, r^t) , calculate $y_j^t \leftarrow d_j(x^t)$

Calculate error rate: $\epsilon_j \leftarrow \sum_t p_j^t \cdot 1(y_j^t \neq r^t)$

If $\epsilon_j > 1/2$, then $L \leftarrow j - 1$; stop

$$\beta_j \leftarrow \epsilon_j / (1 - \epsilon_j)$$

For each (x^t, r^t) , decrease probabilities if correct:

$$\text{If } y_j^t = r^t \text{ then } p_{j+1}^t \leftarrow \beta_j p_j^t \text{ Else } p_{j+1}^t \leftarrow p_j^t$$

Normalize probabilities:

$$Z_j \leftarrow \sum_t p_{j+1}^t; \quad p_{j+1}^t \leftarrow p_{j+1}^t / Z_j$$

Testing:

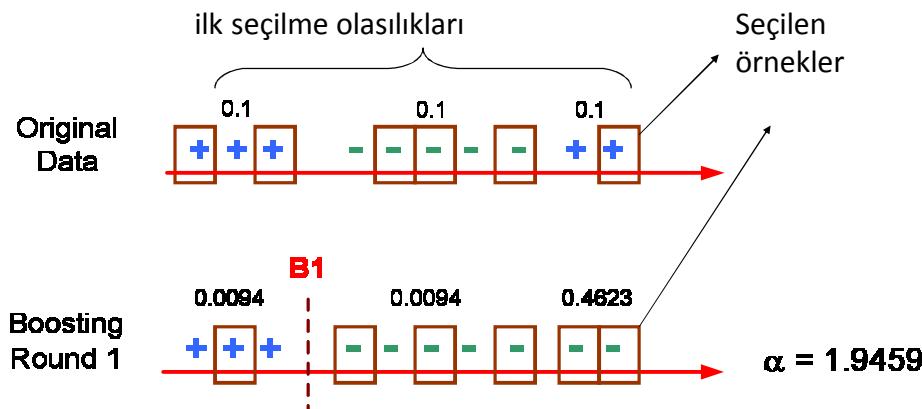
Given x , calculate $d_j(x), j = 1, \dots, L$

Calculate class outputs, $i = 1, \dots, K$:

$$y_i = \sum_{j=1}^L \left(\log \frac{1}{\beta_j} \right) d_{ji}(x)$$



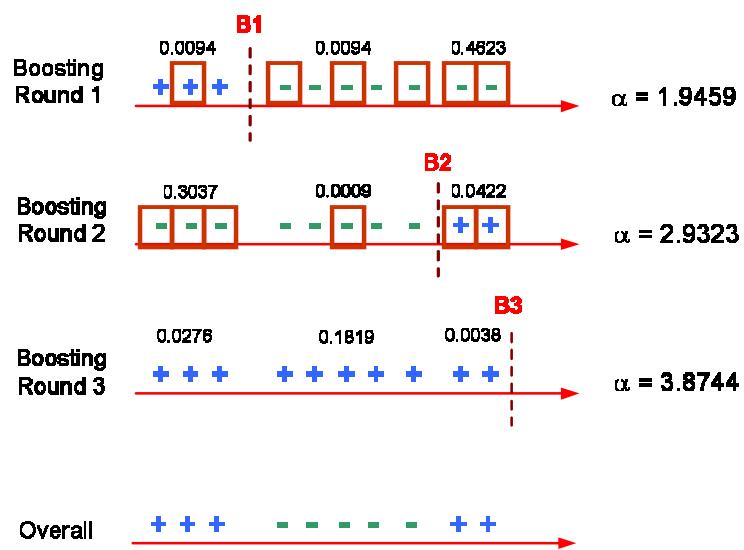
AdaBoost Örnek



17



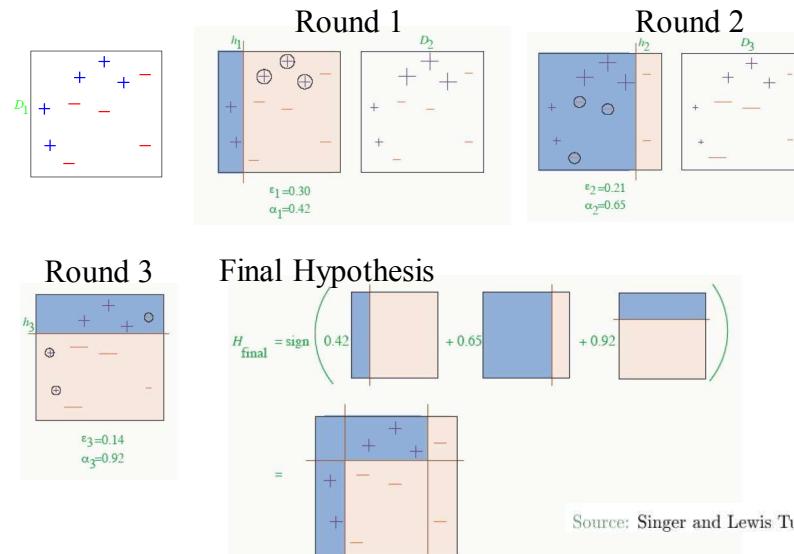
AdaBoost Örnek



18



2 boyutta Adaboost



Source: Singer and Lewis Tutorial



Description of data sets

Name	Cases	Classes	Attributes	
			Cont	Discr
anneal	898	6	9	29
audiology	226	6	-	69
auto	205	6	15	10
breast-w	699	2	9	-
chess	551	2	-	39
colic	368	2	10	12
credit-a	690	2	6	9
credit-g	1,000	2	7	13
diabetes	768	2	8	-
glass	214	6	9	-
heart-c	303	2	8	5
heart-h	294	2	8	5
hepatitis	155	2	6	13
hypo	3,772	5	7	22
iris	150	3	4	-
labor	57	2	8	8
letter	20,000	26	16	-
lymph	148	4	-	18
phoneme	5,438	47	-	7
segment	2,310	7	19	-
sick	3,772	2	7	22
sonar	208	2	60	-
soybean	683	19	-	35
splice	3,190	3	-	62
vehicle	846	4	18	-
vote	435	2	-	16
waveform	300	3	21	-

YILDIZ TEKNİK ÜNİVERSİTESİ
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ

C4.5, and its bagged and boosted versions

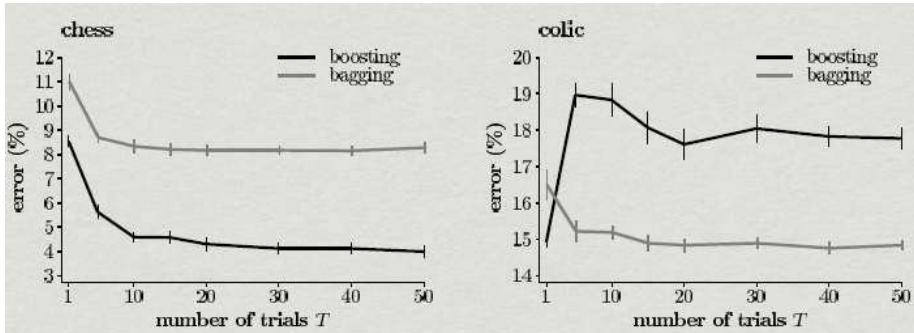
	C4.5	Bagged C4.5 vs C4.5			Boosted C4.5 vs C4.5			Boosting vs Bagging		
		err (%)	err (%)	w-l ratio	err (%)	w-l ratio	w-l ratio	err (%)	w-l ratio	err (%)
anneal	7.67	6.25	10-0	.814	4.73	10-0	.617	10-0	.758	
audiology	22.12	19.29	9-0	.872	15.71	10-0	.710	10-0	.814	
auto	17.66	19.66	2-8	1.113	15.22	9-1	.862	9-1	.774	
breast-w	5.28	4.23	9-0	.802	4.09	9-0	.775	7-2	.966	
chess	8.55	8.33	6-2	.975	4.59	10-0	.537	10-0	.551	
colic	14.92	15.19	0-6	1.018	18.83	0-10	1.262	0-10	1.240	
credit-a	14.70	14.13	8-2	.962	15.64	1-9	1.064	0-10	1.107	
credit-g	28.44	25.81	10-0	.908	29.14	2-8	1.025	0-10	1.129	
diabetes	25.39	23.63	9-1	.931	28.18	0-10	1.110	0-10	1.192	
glass	32.48	27.01	10-0	.832	23.55	10-0	.725	9-1	.872	
heart-c	22.94	21.52	7-2	.938	21.39	8-0	.932	5-4	.994	
heart-h	21.53	20.31	8-1	.943	21.05	5-4	.978	3-6	1.037	
hepatitis	20.39	18.52	9-0	.908	17.68	10-0	.867	6-1	.955	
hypo	.48	.45	7-2	.928	.36	9-1	.746	9-1	.804	
iris	4.80	5.13	2-6	1.069	6.53	0-10	1.361	0-8	1.273	
labor	19.12	14.39	10-0	.752	13.86	9-1	.725	5-3	.963	
letter	11.99	7.51	10-0	.626	4.66	10-0	.389	10-0	.621	
lymphography	21.69	20.41	8-2	.941	17.43	10-0	.804	10-0	.854	
phoneme	19.44	18.73	10-0	.964	16.36	10-0	.842	10-0	.873	
segment	3.21	2.74	9-1	.853	1.87	10-0	.583	10-0	.684	
sick	1.34	1.22	7-1	.907	1.05	10-0	.781	9-1	.861	
sonar	25.62	23.80	7-1	.929	19.62	10-0	.766	10-0	.824	
soybean	7.73	7.58	6-3	.981	7.16	8-2	.926	8-1	.944	
splice	5.91	5.58	9-1	.943	5.43	9-0	.919	6-4	.974	
vehicle	27.09	25.54	10-0	.943	22.72	10-0	.839	10-0	.889	
vote	5.06	4.37	9-0	.864	5.29	3-6	1.046	1-9	1.211	
waveform	27.33	19.77	10-0	.723	18.53	10-0	.678	8-2	.988	
average	15.66	14.11		.905	13.36		.847		.930	

Kırmızı çerçeveler tek ağacın daha iyi olduğu datasetleri gösteriyor.

Son sütundaki 1'den büyükler Bagging'in Boosting'den daha iyi olduklarını gösteriyor.



YILDIZ TEKNİK ÜNİVERSİTESİ
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



YILDIZ TEKNİK ÜNİVERSİTESİ
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ

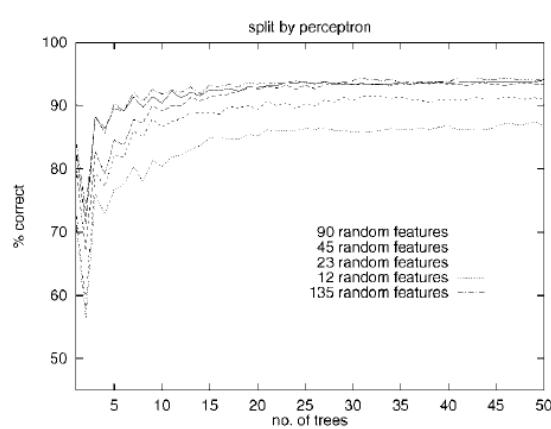
3-Rastsal Altuzaylar - Random Subspace

- Tin Kam Ho, 1998
- Topluluktaki her bir öğreniciyi tüm özellikler yerine bir bölümüyle eğitme.
- Her öğrenicinin eğitim seti oluştururken rastgele d adet özellik seçilip sadece bu özellikler eğitim kümesine dahil edilir.
- Sonuçlar demokrasi usulü birleştirilir.

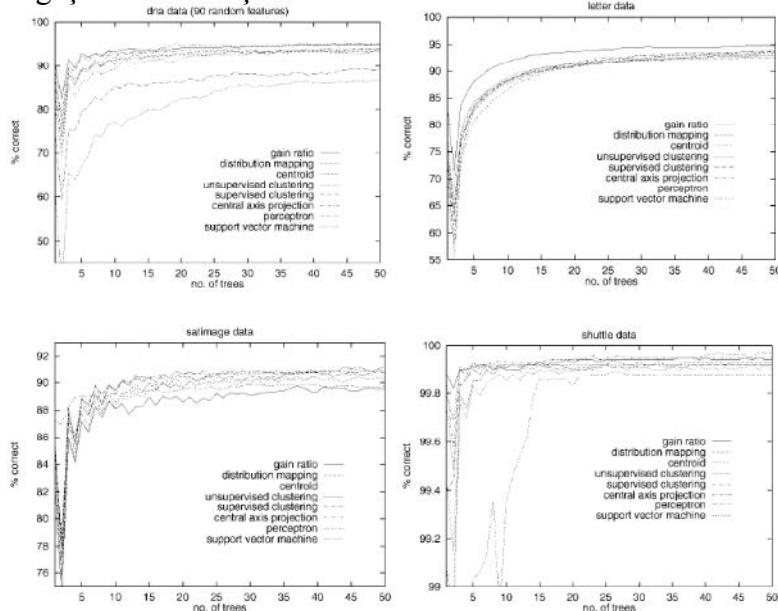
23



Kaç özellik kullanılacak?



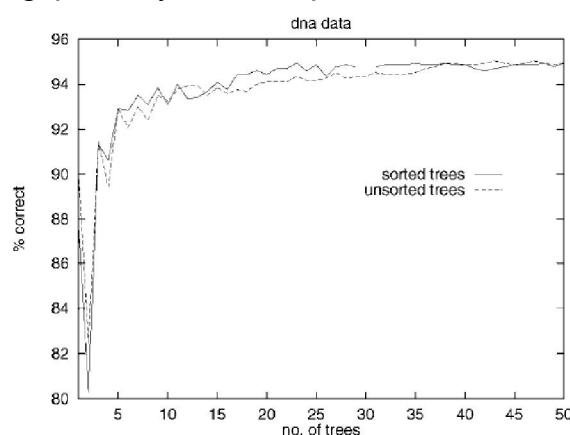
Ağaçları nasıl oluştursak?



YILDIZ TEKNİK ÜNİVERSİTESİ
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



Tüm ağaçları mı yoksa en başarılıları mı kullansak?



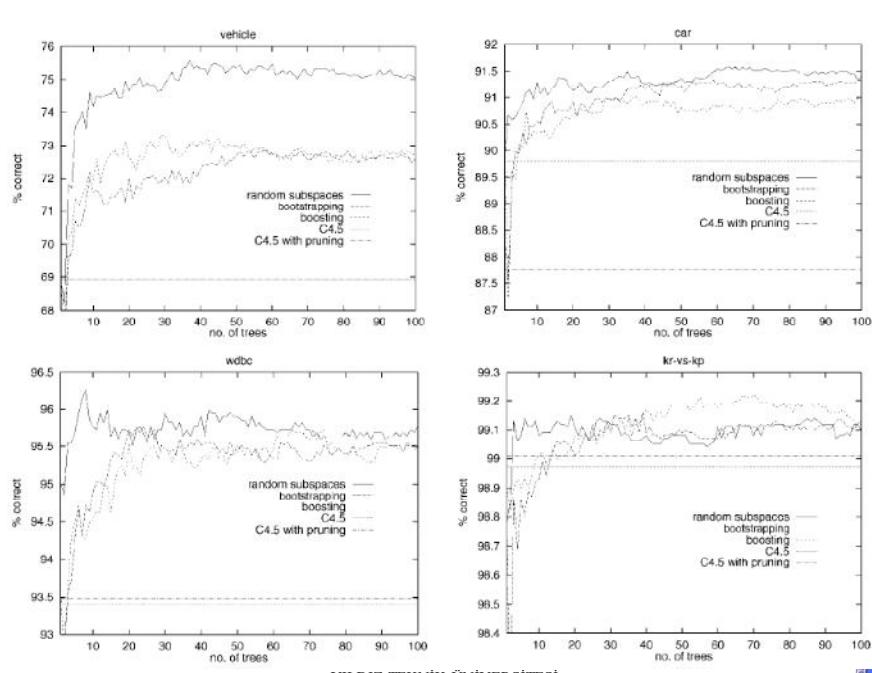
YILDIZ TEKNİK ÜNİVERSİTESİ
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



Denemelerde kullanılan veri kümeleri:
denemeler $10 * 2$ fold cv,
öğrenici C4.5

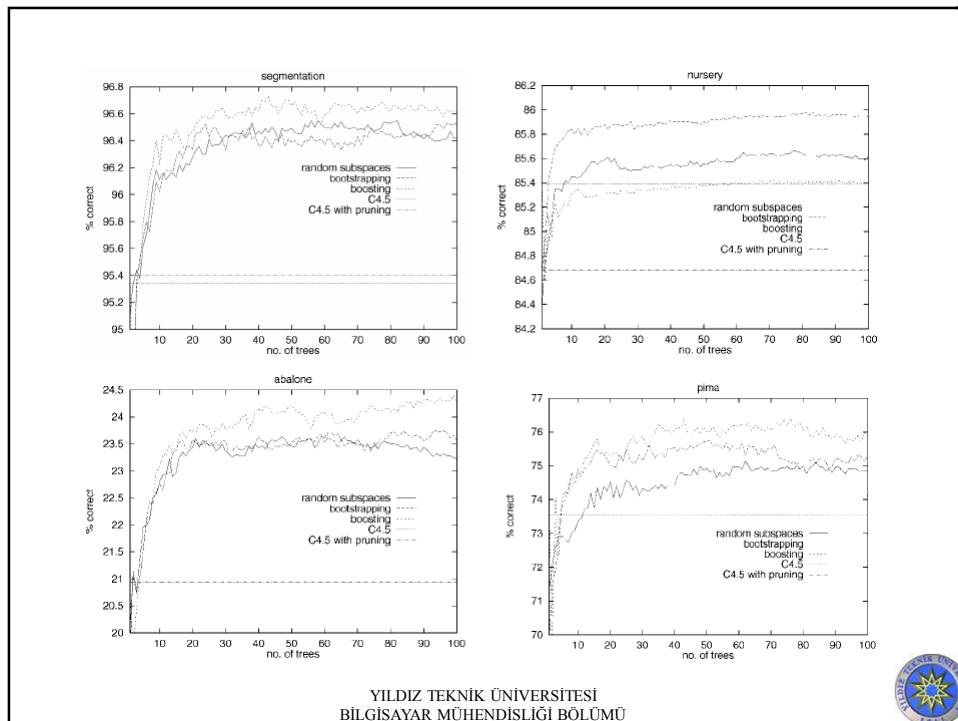
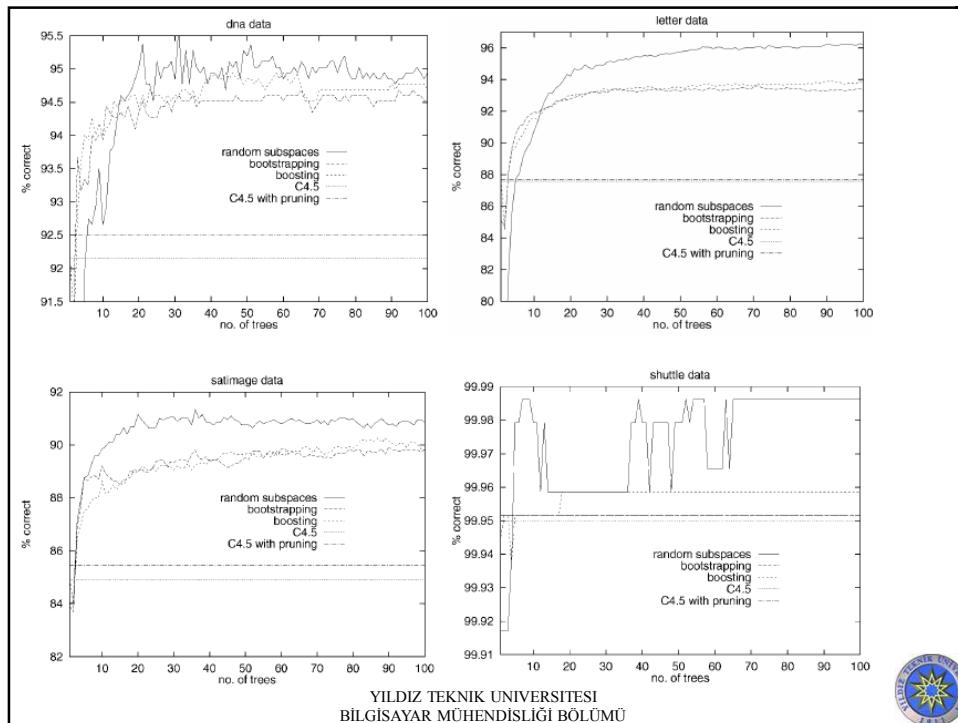
data set	no. of samples	no. of features	no. of classes	feature type(s)
letter	20,000	16	26	numeric
splice	3,190	60	3	categorical
vehicle	846	18	4	numeric
car	1,728	6	4	mixed
wdbc	569	30	2	numeric
kr-vs-kp	3,196	36	2	categorical
lrc	531	93	10	numeric
german	1,000	24	2	numeric
segmentation	2,310	19	8	numeric
nursery	12,961	8	5	mixed
abalone	4,177	8	29	mixed
pima	768	8	2	numeric
tic-tac-toe	958	9	2	categorical
yeast	1,484	8	10	numeric

YILDIZ TEKNİK ÜNİVERSİTESİ
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



YILDIZ TEKNİK ÜNİVERSİTESİ
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ





4-Rastsal Ormanlar - Random Forests

- Öncekilerin aksine sadece karar ağaçları için tasarlanmış bir Ensemble yöntemi
- 2 tür rastgelelik kullanır
 - **Bagging** : her ağacın eğitimi için orijinal eğitim kümesiyle aynı boyutta resampling yapar
 - **Rastgele Özellikler**: Her karar düğümünde tun özellikler yerine rastgele seçilmiş F tanesi (RI) ya da adet özelliklerin lineer kombinasyonlarından F adedi (RC) kullanır

31



Rastsal Ormanlar - Random Forests

- Budanmamış ağaçları kullanır
- Sonuçlar demokrasi usulü birleştirilir.
- Eğitim hızı ?
- F kaç olmalı ?

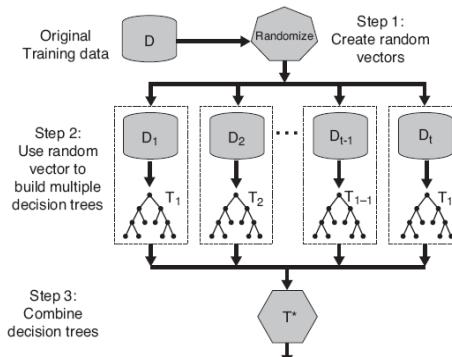


Figure 5.40. Random forests.

2



Forest-RI (ozellik secimi)

Data set	Adaboost	Selection	Forest-RI single input	One tree
Glass	22.0	20.6	21.2	36.9
Breast cancer	3.2	2.9	2.7	6.3
Diabetes	26.6	24.2	24.3	33.1
Sonar	15.6	15.9	18.0	31.7
Vowel	4.1	3.4	3.3	30.4
Ionosphere	6.4	7.1	7.5	12.7
Vehicle	23.2	25.8	26.4	33.1
German credit	23.5	24.4	26.2	33.3
Image	1.6	2.1	2.7	6.4
Ecoli	14.8	12.8	13.0	24.5
Votes	4.8	4.1	4.6	7.4
Liver	30.7	25.1	24.7	40.6
Letters	3.4	3.5	4.7	19.8
Sat-images	8.8	8.6	10.5	17.2
Zip-code	6.2	6.3	7.8	20.6
Waveform	17.8	17.2	17.3	34.0
Twonorm	4.9	3.9	3.9	24.7
Threenorm	18.8	17.5	17.5	38.4
Ringnorm	6.9	4.9	4.9	25.7



Forest-RC (ozelliklerin lineer kombinasyonu)

Data set	Adaboost	Forest-RC		
		Selection	Two features	One tree
Glass	22.0	24.4	23.5	42.4
Breast cancer	3.2	3.1	2.9	5.8
Diabetes	26.6	23.0	23.1	32.1
Sonar	15.6	13.6	13.8	31.7
Vowel	4.1	3.3	3.3	30.4
Ionosphere	6.4	5.5	5.7	14.2
Vehicle	23.2	23.1	22.8	39.1
German credit	23.5	22.8	23.8	32.6
Image	1.6	1.6	1.8	6.0
Ecoli	14.8	12.9	12.4	25.3
Votes	4.8	4.1	4.0	8.6
Liver	30.7	27.3	27.2	40.3
Letters	3.4	3.4	4.1	23.8
Sat-images	8.8	9.1	10.2	17.3
Zip-code	6.2	6.2	7.2	22.7
Waveform	17.8	16.0	16.1	33.2
Twonorm	4.9	3.8	3.9	20.9
Threenorm	18.8	16.8	16.9	34.8
Ringnorm	6.9	4.8	4.6	24.6



Gürültüye dayanıklı mı?

- Eğitim kümесinin etiketlerine eklenen gurultunun hata oranına etkileri

Data set	Adaboost	Forest-RI	Forest-RC
Glass	1.6	.4	-.4
Breast cancer	43.2	1.8	11.1
Diabetes	6.8	1.7	2.8
Sonar	15.1	-6.6	4.2
Ionosphere	27.7	3.8	5.7
Soybean	26.9	3.2	8.5
Ecoli	7.5	7.9	7.8
Votes	48.9	6.3	4.6
Liver	10.3	-.2	4.8



5-Aşırı Rastsal Ormanlar – Extremely randomized trees

- Pierre Geurts · Damien Ernst · Louis Wehenkel, 2006
- Öğrenmede rastgelelik ne kadar fazla olursa, temel öğrenicilerin kararları birbirlerinden o kadar farklı olur.
- Fikir: Ağacın karar düğümlerini rastgele üretelim.

Table 1 Extra-Trees splitting algorithm (for numerical attributes)**Split_a_node(S)***Input:* the local learning subset S corresponding to the node we want to split.*Output:* a split $[a < a_c]$ or nothing.

- If **Stop_split(S)** is TRUE, then return nothing.
- Otherwise select K attributes $\{a_1, \dots, a_K\}$ among all non constant (in S) candidate attributes;
- Draw K splits (s_1, \dots, s_K) , where $s_i = \text{Pick_a_random_split}(S, a_i), \forall i = 1, \dots, K$;
- Return a split s_* such that $\text{Score}(s_*, S) = \max_{i=1, \dots, K} \text{Score}(s_i, S)$.

Pick_a_random_split(S, a)*Inputs:* a subset S and an attribute a *Output:* a split

- Let a_{\max}^S and a_{\min}^S denote the maximal and minimal value of a in S ;
- Draw a random cut-point a_c uniformly in $[a_{\min}^S, a_{\max}^S]$;
- Return the split $[a < a_c]$.

Stop_split(S)*Input:* a subset S *Output:* a boolean

- If $|S| < n_{\min}$, then return TRUE;
- If all attributes are constant in S , then return TRUE;
- If the output is constant in S , then return TRUE;
- Otherwise, return FALSE.

YILDIZ TEKNİK ÜNİVERSİTESİ
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



- Seçilen K adet özelliğin her biri için rastgele bir eşik değerleri belirleniyor. Bu K adet özellik-eşik ikilisinden en iyisi karar düğümünü oluşturuyor.
- 3 parametre:
- M : temel öğrenici sayısı
- K : rastgele seçilecek özellik sayısı (default : sınıflandırma = karekök(orijinal özellik sayısı), regresyon =orijinal özellik sayısı)
- n_{\min} : Bölmeye devam etmek için gerekli minimum örnek sayısı (default : sınıflandırma = 2, regresyon =5)



- Avantajları:
 - Rastgelelik sayesinde diversity yüksek
 - En iyi özellik-eşik ikilisini bulmak için tüm olasılıkların çok azını denediginden eğitimi hızlı
- Dezavantajları:
 - Temel öğrenicilerin ağaçları çok büyük (fazla hafıza gereksinimi)
 - Test işlemi uzun

39



Düger alg. başarı karşı. Win/draw/loss

Classification problems					Regression problems				
PST	TB	RS*	RF*	ET ^d	PST	TB	RS*	RF*	ET ^d
–	8/4/0	11/1/0	11/1/0	10/2/0	–	10/2/0	8/4/0	10/2/0	10/2/0
TB	0/4/8	–	7/5/0	7/5/0	7/5/0	0/2/10	–	0/9/3	1/11/0
RS*	0/1/11	0/5/7	–	0/8/4	2/10/0	0/4/8	3/9/0	–	4/8/0
RF*	0/1/11	0/5/7	4/8/0	–	5/7/0	0/2/10	0/11/1	0/8/4	4/7/1
ET ^d	0/2/10	0/5/7	0/10/2	0/7/5	–	0/2/10	0/8/4	1/7/4	3/7/2

PST: Tekil ağaç

TB: Bagging

RS: Random Subspaces

RF: Random Forest

ET: Extremly Randomized Forest



6-Rotasyon Ormanları - Rotation Forests

- J.J. Rodriguez, L.I Kuncheva, 2006
- Orijinal eğitim kümesine yakın eğitim kümeleriyle eğitilen temel öğrencilerin kişisel başarıları yüksek ancak kararlarının birbirlerinden farklılıklarını az olur.
- Fikir: Tüm örnek ve özelliklerin sayılarını aynı bırakalım. Datayı döndürelim. Ama nasıl?

41



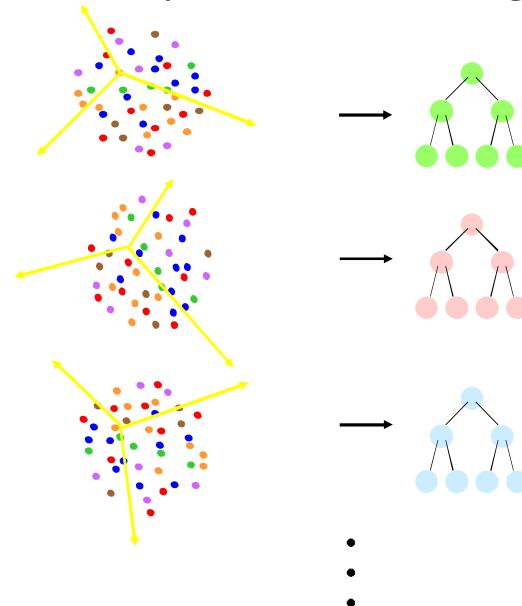
Rotation Forest

- PCA kullanalım ama farklı transformasyon matrislerini nasıl bulacağız? olasılıklar:
 - örneklerden bir kısmı
 - sınıflardan bir kısmı
 - özelliklerden bir kısmı

42



Rotasyon + Karar Ağacıları



43



Input

- \mathcal{L} : a training set, $\mathcal{L} = \{(\mathbf{x}_i, y_i)\}_{i=1}^N = [X \ Y]$ where X is an $N \times p$ matrix containing the input attribute values and Y is an N -dimensional column vector containing the class labels.
- K : number of attribute subsets (or M : number of input attributes contained in each subset).
- \mathcal{W} : a base learning algorithm.
- T : number of iterations.
- \mathbf{x} : a data point to be classified.

Training Phase

For $t = 1, 2, \dots, T$

- Calculate the rotation matrix R_t^* for the t th classifier C_t .
- 1. Randomly split the attribute set F into K subsets $F_{t,k}$ ($k = 1, 2, \dots, K$).
 - Select the columns of X that correspond to the attributes in $F_{t,k}$ to compose a submatrix $X_{t,k}$.
 - Draw a bootstrap sample $X'_{t,k}$ (with sample size generally smaller than that of $X_{t,k}$) from $X_{t,k}$.
 - Apply PCA to $X'_{t,k}$ to obtain a matrix $D_{t,k}$ whose i th column consists of the coefficients of the i th principal component.
- 3. EndFor
- 4. Arrange the matrices $D_{t,k}$ ($k = 1, 2, \dots, K$) into a block diagonal matrix R_t .
- 5. Construct the rotation matrix R_t^* by rearranging the rows of R_t so that they correspond to the original attributes in F .

- Provide $[XR_t^* \ Y]$ as the input of \mathcal{W} to build a classifier C_t .

EndFor

Output

- the class label for \mathbf{x} predicted by the ensemble classifier C^* as

$$C^*(\mathbf{x}) = \operatorname{argmax}_{y \in \Phi} \sum_{t=1}^T I(C_t(\mathbf{x}R_t^*) = y),$$

where $I(\cdot)$ is an indicator function.

BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



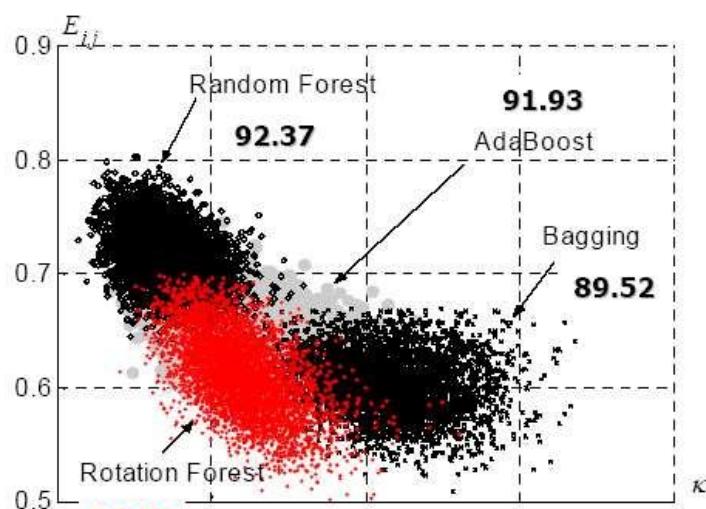
Data Set	Rotations J48	J48	Bagging J48	Boosting J48	Random Forest
anneal	99.01±0.93	98.62±1.01	98.98±0.93	99.54±0.68	99.38±0.78
audiology	79.83±6.93	76.33±7.45	81.12±7.35	83.30±6.99	76.58±7.94
autos	82.56±8.66	82.86±9.25	84.12±8.42	84.61±7.93	81.95±7.85
balance-scale	90.26±2.62	79.43±4.01	81.39±3.70	76.82±4.14	80.28±3.80
breast-cancer	72.07±6.54	68.00±7.43	69.48±7.17	66.12±7.81	69.00±7.31
cleveland-14-heart	82.61±6.12	76.49±6.91	79.70±6.01	79.20±7.25	80.34±6.47
credit-rating	86.00±3.90	82.50±4.24	85.17±4.34	84.02±3.98	85.15±4.23
glass	74.33±8.06	67.77±9.70	73.85±9.34	76.23±9.09	75.65±8.42
german-credit	73.87±3.89	67.89±3.95	72.08±3.63	71.95±4.32	73.57±3.38
heart-statlog	82.37±6.45	76.69±7.51	80.44±6.84	79.38±7.40	80.86±6.53
hepatitis	82.92±8.88	78.95±9.27	80.68±8.89	82.45±8.17	83.04±8.07
horse-colic	84.80±5.35	82.16±5.89	84.80±5.96	81.05±6.20	84.96±5.43
hungarian-14-heart	79.57±6.45	78.85±7.30	78.74±6.65	79.08±7.00	79.28±6.31
hypothyroid	99.57±0.33	99.51±0.37	99.59±0.30	99.65±0.30	99.18±0.46
ionosphere	93.88±3.76	89.97±4.55	92.29±3.79	93.01±3.97	92.84±3.89
iris	95.73±5.20	94.93±4.99	94.58±5.15	94.36±5.22	94.13±5.18
labor	91.69±11.89	79.84±14.57	84.31±14.44	87.20±13.81	87.00±13.45
letter	95.54±0.47	88.02±0.75	92.85±0.65	95.44±0.50	94.52±0.49
lymphography	84.27±8.35	75.64±11.12	78.97±10.32	82.40±9.73	81.28±8.58
pendigits	99.21±0.25	96.46±0.57	97.99±0.44	99.01±0.28	98.81±0.29
pima-diabetes	76.39±4.43	73.85±4.94	75.59±4.54	72.49±5.08	74.78±4.42
primary-tumor	44.37±6.56	42.42±7.57	42.79±6.92	41.64±6.94	41.56±6.50
segment	98.05±0.95	96.81±1.26	97.58±1.05	98.25±0.80	97.71±1.06
sonar	83.49±7.88	73.82±8.71	78.34±9.14	79.95±9.51	80.75±7.84
soybean	94.17±2.47	90.67±3.34	91.88±3.15	92.44±2.76	91.92±2.83
splice	95.49±1.13	92.20±1.37	94.25±1.20	94.11±1.23	90.07±1.79
vehicle	77.95±3.74	72.38±4.25	74.70±4.07	76.44±4.01	74.37±4.43
vote	96.08±2.88	95.71±2.93	96.43±2.47	95.22±3.19	95.74±2.75
vowel-c	96.87±1.76	81.26±4.18	91.72±2.89	94.15±2.42	95.59±2.23
vowel-n	95.77±1.94	79.22±4.59	89.52±3.27	91.93±2.72	92.37±2.73
waveform	83.94±1.72	75.14±1.99	81.78±1.74	81.45±1.71	81.89±1.74
wisconsin-breast-cancer	97.02±1.93	94.30±2.74	95.82±2.54	95.97±2.11	95.75±2.14
zoo	92.35±8.04	93.42±6.93	93.50±7.11	97.04±5.21	95.83±6.02
(Win/Tie/Loss)	(0/13/21)	(0/24/10)	(1/25/8)	(0/24/10)	

○ Rotation Forest is significantly worse, • Rotation Forest is significantly better, level of significance 0.05

YILDIZ TEKNİK ÜNİVERSİTESİ
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



Kappa-error diagrams for vowel data



YILDIZ TEKNİK ÜNİVERSİTESİ
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



7- Hata düzeltken kod tabanlı topluluklar (Error-Correcting Output Codes)

- *Daha basit problemler*
 - K sınıf; L problem (Dietterich and Bakiri, 1995)
 - Kod matrisi \mathbf{W} oluşturma yöntemleri (3 adet)
 - 1-Bir sınıfı diğer tüm sınıflara karşı $L=K$
- $$\mathbf{W} = \begin{bmatrix} +1 & -1 & -1 & -1 \\ -1 & +1 & -1 & -1 \\ -1 & -1 & +1 & -1 \\ -1 & -1 & -1 & +1 \end{bmatrix}$$

Her sütunun ifade ettiği eğitim setini bir öğrenci öğrenir. Yeni eğitim setinde sütunda $+1$ ile ifade edilen sınıfa ait örneklerin çıkışı $+1$, diğer tüm sınıflara ait örneklerin çıkışı -1 yapılır.

Lecture Notes for E Alpaydin
2004 Introduction to Machine
Learning © The MIT Press (V1.1)

47



- 2-Sınıf ikilileri (Pairwise)

$$L=K(K-1)/2$$

$$\mathbf{W} = \begin{bmatrix} +1 & +1 & +1 & 0 & 0 & 0 \\ -1 & 0 & 0 & +1 & +1 & 0 \\ 0 & -1 & 0 & -1 & 0 & +1 \\ 0 & 0 & -1 & 0 & -1 & -1 \end{bmatrix}$$

Her sütunun ifade ettiği eğitim setini bir öğrenci öğrenir. Yeni eğitim setinde sütunda $+1$ ile ifade edilen sınıfa ait örneklerin çıkışı $+1$, -1 ile ifade edilenlerinki -1 yapılır. Diğer sınıflara ait örnekler eğitim setine dahil edilmezler.



- 3-Tüm olası alt problemler (Full code) $L=2^{(K-1)}-1$

$$\mathbf{w} = \begin{bmatrix} -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & +1 & +1 & +1 & +1 \\ -1 & +1 & +1 & -1 & -1 & +1 & +1 \\ +1 & -1 & +1 & -1 & +1 & -1 & +1 \end{bmatrix}$$

Her sütunun ifade ettiği eğitim setini bir öğrenici öğrenir. Yeni eğitim setinde sütunda +1 ile ifade edilen sınıflara ait örneklerin çıkışı +1, -1 olan sınıflara ait örneklerin çıkışı -1 yapılır.



Test süreci

- Her bir öğrenici, her test örneğine -1, +1 aralığında bir değeri çıkış olarak verir ($d_{1..L}$).
- Bir örneğin sınıfı ise bu d vektörünün en çok benzediği w satırının indisidir. w'deki 0 lar etkisiz eleman olarak algılanmalıdır.



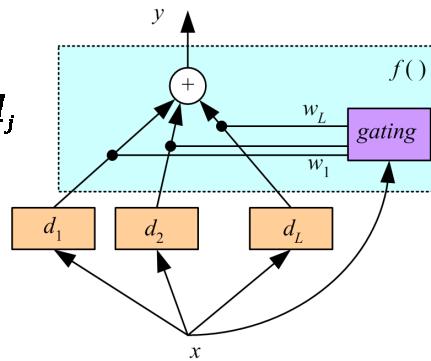
8- Yerel Modellerin Birleşimi (Mixture of Experts)

Her bir öğrenci veri uzayının farklı bir bölgesine uzmanlaşır.

$$(Jacobs \text{ et al., 1991}) \quad \mathbf{y} = \sum_{j=1}^L \mathbf{w}_j \mathbf{d}_j$$

Test örneğinin gating'e de verilmesinin sebebi örneğin uzayın hangi bölgesinde olduğunun bulunmasıdır.

Öğreniciler ve gating fonksiyonu lineer ya da nonlinear olabilir.

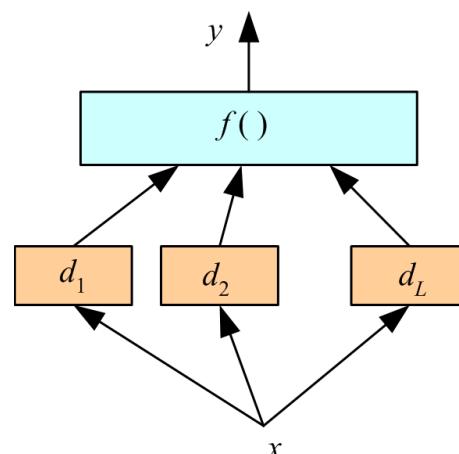


Lecture Notes for E Alpaydin 2004 Introduction to Machine Learning © The MIT Press (V1.1)



9- İki aşamalı öğrenme (Stacking)

- Öğrenicilerin çıktılarının yeni bir veri kümesi olarak görüldüğü yöntemdir.
(Wolpert, 1992)



Lecture Notes for E Alpaydin 2004 Introduction to Machine Learning © The MIT Press (V1.1)



Attrs	Cl.	Classifier ₁			Classifier ₂			Classifier _N			class = a?	
		a	b	c	a	b	c	a	b	c		
$AttrVec_1$	a	$P_{1,a1}$	$P_{1,b1}$	$P_{1,c1}$	$P_{2,a1}$	$P_{2,b1}$	$P_{2,c1}$	\dots	$P_{N,a1}$	$P_{N,b1}$	$P_{N,c1}$	1
$AttrVec_2$	b	$P_{1,a2}$	$P_{1,b2}$	$P_{1,c2}$	$P_{2,a2}$	$P_{2,b2}$	$P_{2,c2}$	\dots	$P_{N,a2}$	$P_{N,b2}$	$P_{N,c2}$	0
$AttrVec_3$	b	$P_{1,a3}$	$P_{1,b3}$	$P_{1,c3}$	$P_{2,a3}$	$P_{2,b3}$	$P_{2,c3}$	\dots	$P_{N,a3}$	$P_{N,b3}$	$P_{N,c3}$	0
$AttrVec_4$	c	$P_{1,a4}$	$P_{1,b4}$	$P_{1,c4}$	$P_{2,a4}$	$P_{2,b4}$	$P_{2,c4}$	\dots	$P_{N,a4}$	$P_{N,b4}$	$P_{N,c4}$	0
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$AttrVec_n$	a	$P_{1,an}$	$P_{1,bn}$	$P_{1,cn}$	$P_{2,an}$	$P_{2,bn}$	$P_{2,cn}$	\dots	$P_{N,an}$	$P_{N,bn}$	$P_{N,cn}$	1

(a) original training set

(c) meta training set for class a , Stacking with MLR

	<i>Classifier_i</i>	
<i>a</i>	<i>b</i>	<i>c</i>
$P_{i,a1} = 0.90$	$P_{i,b1} = 0.05$	$P_{i,c1} = 0.05$
$P_{i,a2} = 0.15$	$P_{i,b2} = 0.70$	$P_{i,c2} = 0.15$
$P_{i,a3} = 0.10$	$P_{i,b3} = 0.80$	$P_{i,c3} = 0.10$
$P_{i,a4} = 0.20$	$P_{i,b4} = 0.20$	$P_{i,c4} = 0.60$
\vdots	\vdots	\vdots
$P_{i,an} = 0.80$	$P_{i,bn} = 0.10$	$P_{i,cn} = 0.10$

(b) sample class probability distribution

<i>Classifier</i> <i>a</i>	<i>Classifier</i> <i>a</i>		<i>Classifier</i> <i>N</i>	<i>class</i> = <i>a</i> ?
$P_1,a1$	$P_2,a1$...	$P_{N,a1}$	1
$P_1,a2$	$P_2,a2$...	$P_{N,a2}$	0
$P_1,a3$	$P_2,a3$...	$P_{N,a3}$	0
$P_1,a4$	$P_2,a4$...	$P_{N,a4}$	0
:	:	⋮	⋮	⋮
$P_{1,n}$	$P_{2,n}$...	$P_{N,n}$	1

(d) meta training set for class a , StackingC with MLR

Figure 1. Illustration of Stacking and StackingC on a dataset with three classes (a , b and c), n examples and N base classifiers. $P_{i,jk}$ refers to the probability given by base classifier i for class j on example number k .



Sorular

- Eğitim datasını bölgelere bölüp kolektif öğrenme uygularsak, temel öğrenicilerin kararlarını demokrasi ile birleştirmek mantıklı olur mu?
 - Boosting, Bagging, RS, RF, RotF hangileri paralelleştirilebilir?
 - Rotasyon Ormanlarında temel öğrenci olarak KNN kullanımı işe yarar mı? Neden karar ağaçları kullanılmış?

