

# BM605 PR: Örüntü Tanıma

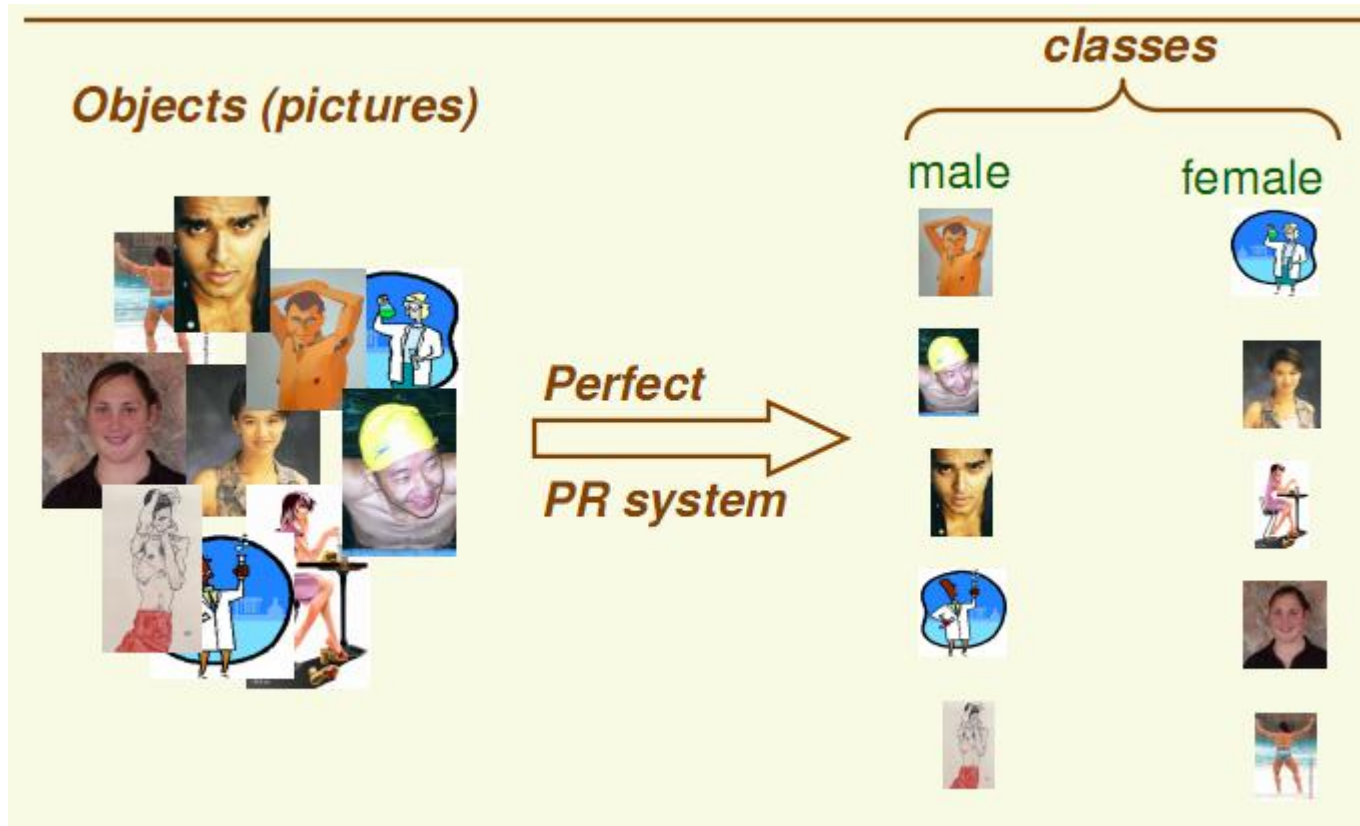
Samsun – 2011

# Örüntü Tanıma Nedir?

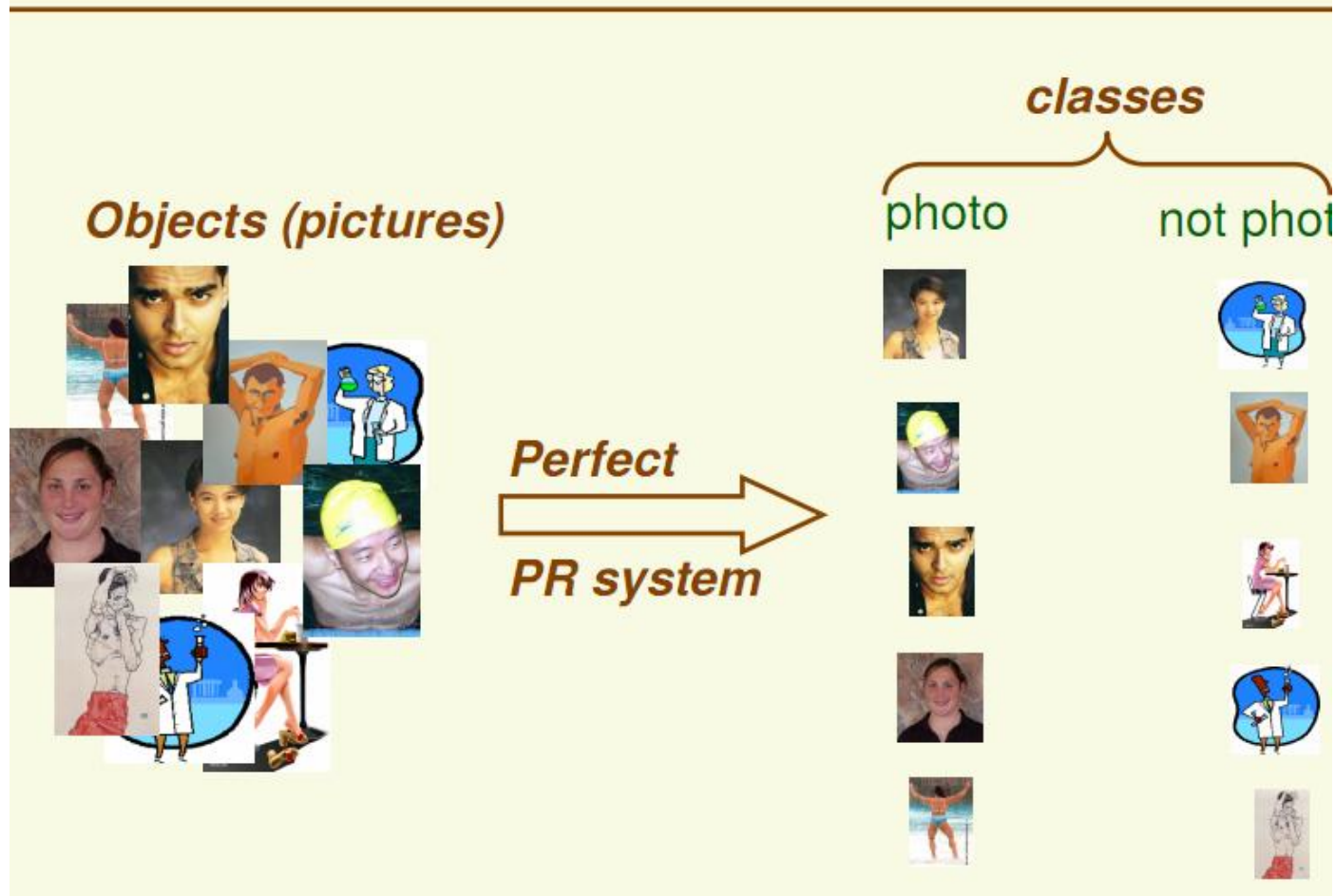
- Veri içerisindeki örüntüyü tanıma
- Nesne/olayı, önceden tanımlanan kategorite/sınıfa atama işi
- Ör. Bardak / Yüz / Telefon?



# Uygulama:Cinsiyet: E/K

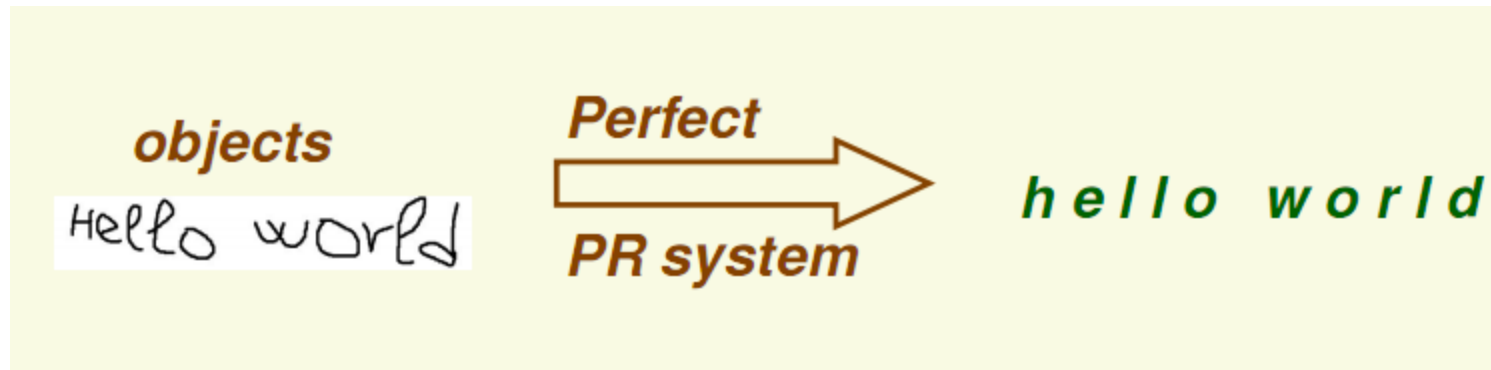


# Uygulama: Foto?

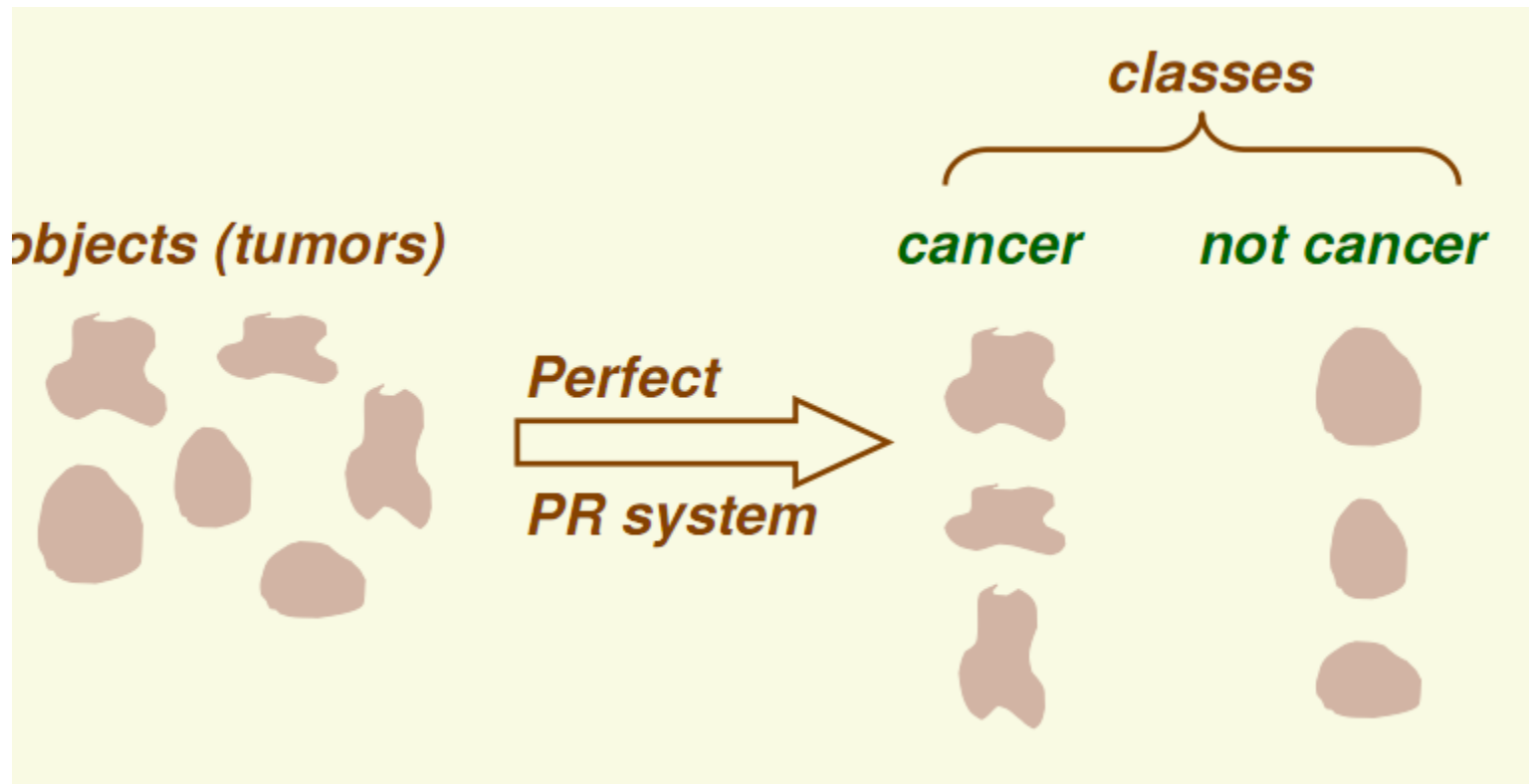


# Uygulama: OCR

- Sınıflar: harfler: a, b, c, ..., z



# Uygulama: Tıbbi Tanı



# Uygulama: ses tanıma

*objects (acoustic signal)*



*Perfect*  
→  
*PR system*

*phonemes*

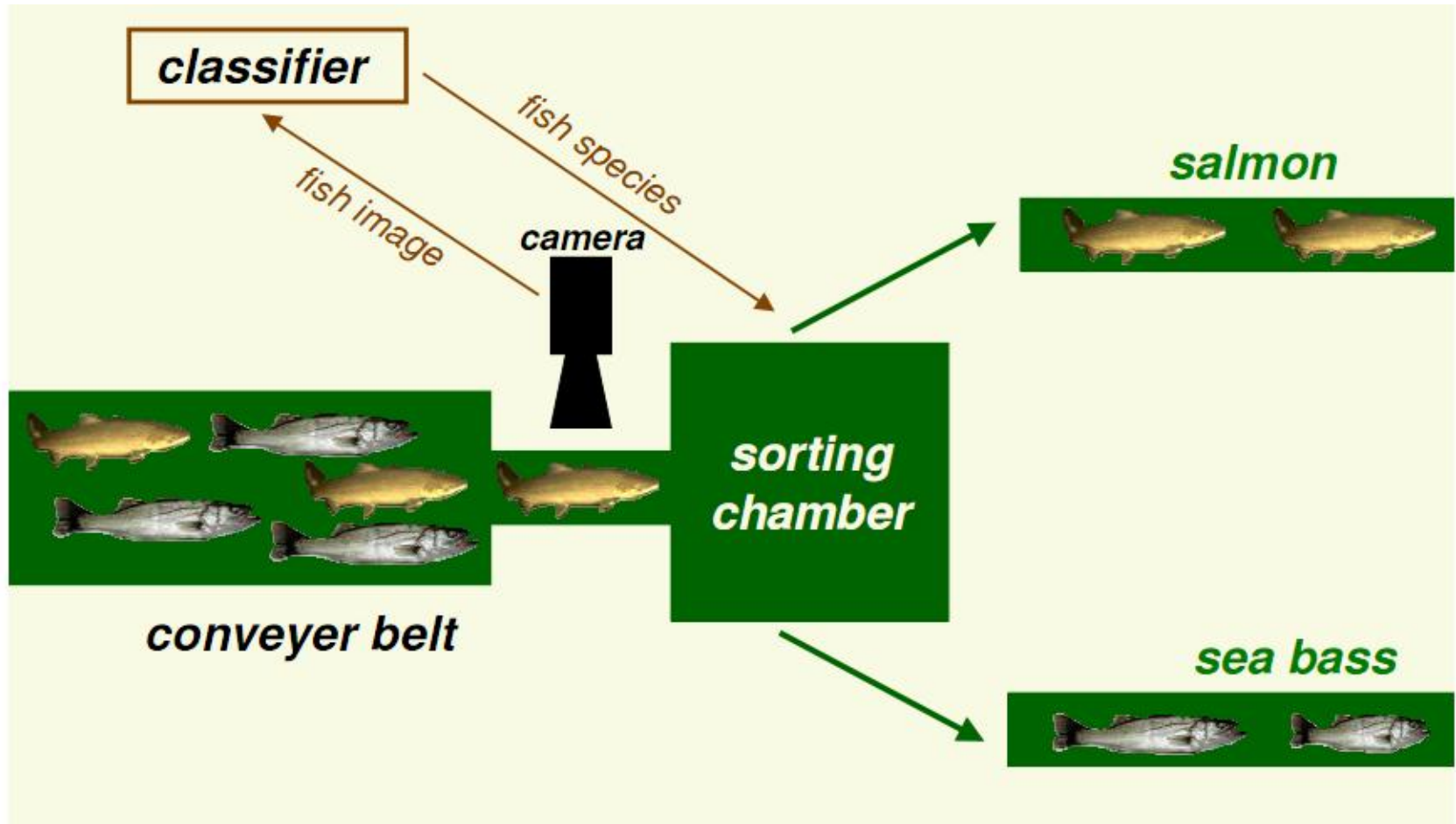
*re-kig-'ni-sh&n*

# Uygulama: kredi

<i>objects (people)</i>					<i>classes</i>	
	income	debt	married	age	approve	deny
John Smith	200,000	0	yes	80		<input checked="" type="checkbox"/>
Peter White	60,000	1,000	no	30	<input checked="" type="checkbox"/>	
Ann Clark	100,000	10,000	yes	40	<input checked="" type="checkbox"/>	
Susan Ho	0	20,000	no	25		<input checked="" type="checkbox"/>



# Anlatımlarda kullanılacak uygulama: Balık Sınıflama



# PR Sistem tasarımı?

- Veriyi topla (eğitim verisi) ve elle sınıfla



- data.x (veri) ve data.y (sınıf)
- Önişle: arkaplandan çıkartma



- Öznitelik çıkart
  - Uzunluk, parlaklık, genişlik, vs

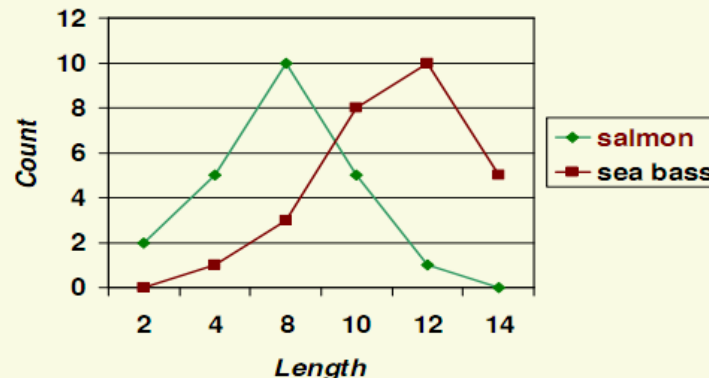
# PR Sistem tasarımı? – 2

- Sınıflandırıcıyı tasarla
  - Modeli seç
  - Eğit (train)
- Test et
  - Yeni verileri sınıflandırma yeteneği nasıl?

# Sınıflandırıcı tasarımı

- Somon (salmon) balığı, levrekten (sea bass) kısa
- Balık uzunluğu ayırt edici
- Her bir uzunluktaki balık sayısı

	2	4	8	10	12	14
bass	0	1	3	8	10	5
salmon	2	5	10	5	1	0



# Balık uzunluğu: ayırtaç olarak

- L'den kısa olanlar somondur
- Uzun olanlar levrektir
- Peki L nedir?
- L=5 alsak (hatalı bir değer)
- Yanlış sınıflandırma oranı =  $17/50 = \text{\textcolor{red}{\%34}}$ 
  - Kırmızılar hatalı (FN,FP), yeşiller doğru (FP,TP)

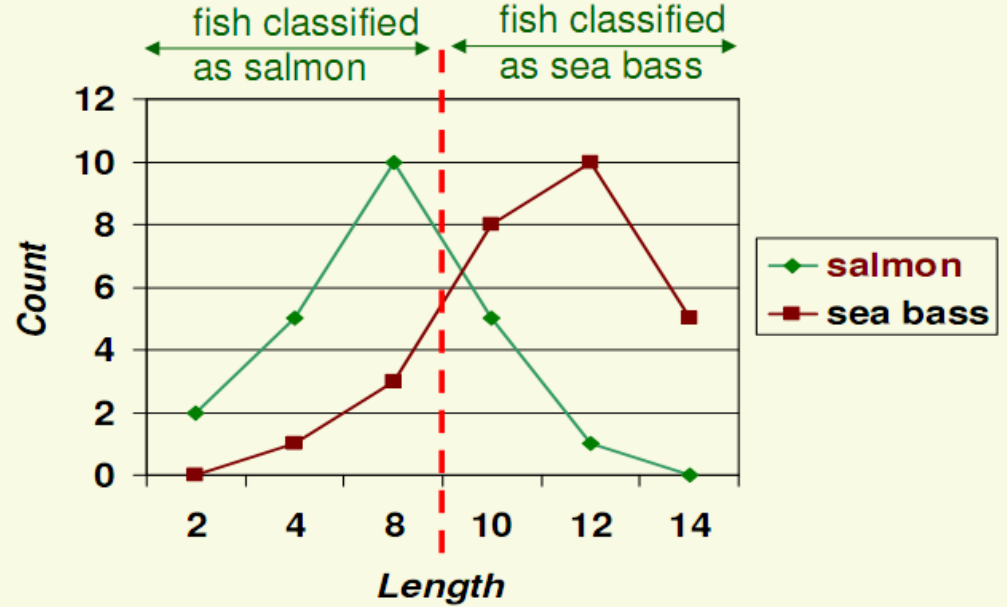
	2	4	8	10	12	14
bass	0	1	3	8	10	5
salmon	2	5	10	5	1	0

# Balık uzunluğu: ayırtaç olarak

En uygun L değeri nedir?

L'nin yaklaşık olarak 9 değeri alması en az hatalı sınıflandırmayla sonuçlanır.

Bu durumda **%20** sınıflandırma hatası alınır.



# Ne yapacağız?

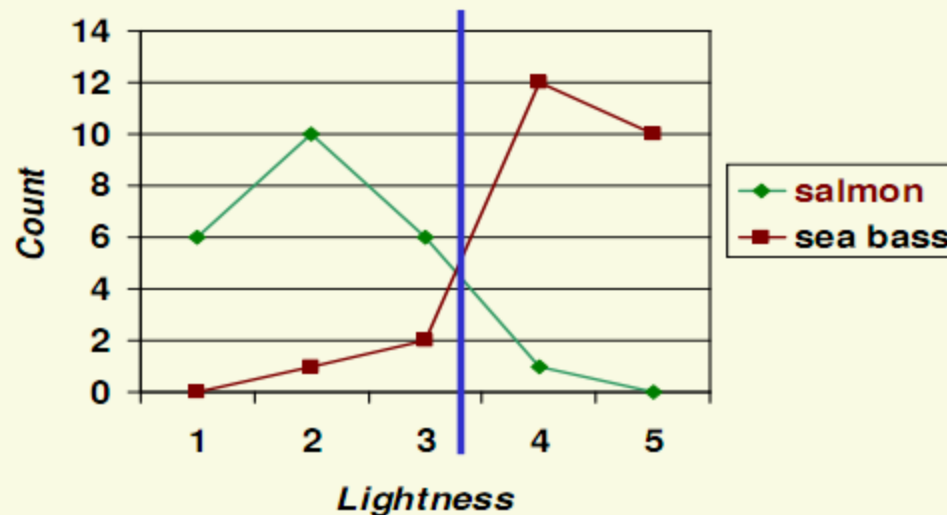
- Uzunluk yetersiz
- Ne yapabiliriz?
  - Başka bir ayırtaç
  - Somon daha parlak
  - Ortalama balık parlaklığını deneyelim



# Balık parlaklık ayırtacı

- Sadece parlaklık ayırtaç olduğunda **%8**'e kadar indik

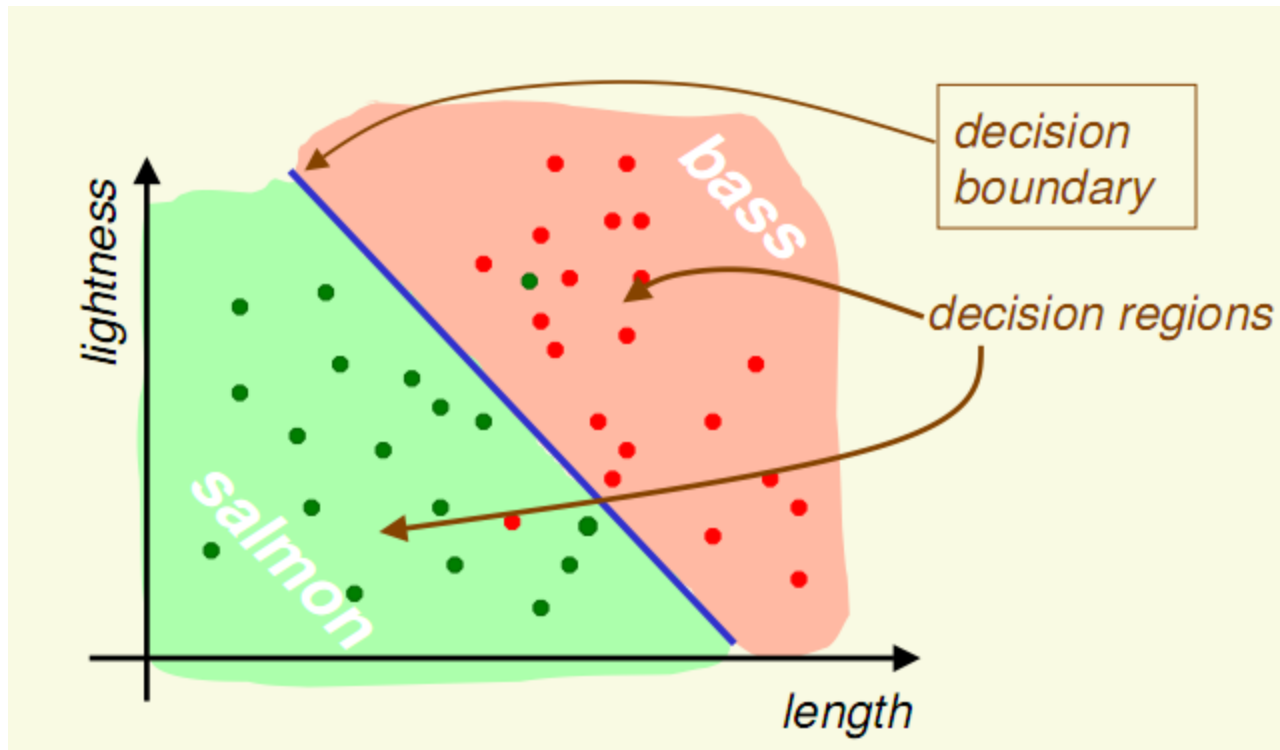
	1	2	3	4	5
bass	0	1	2	10	12
salmon	6	10	6	1	0





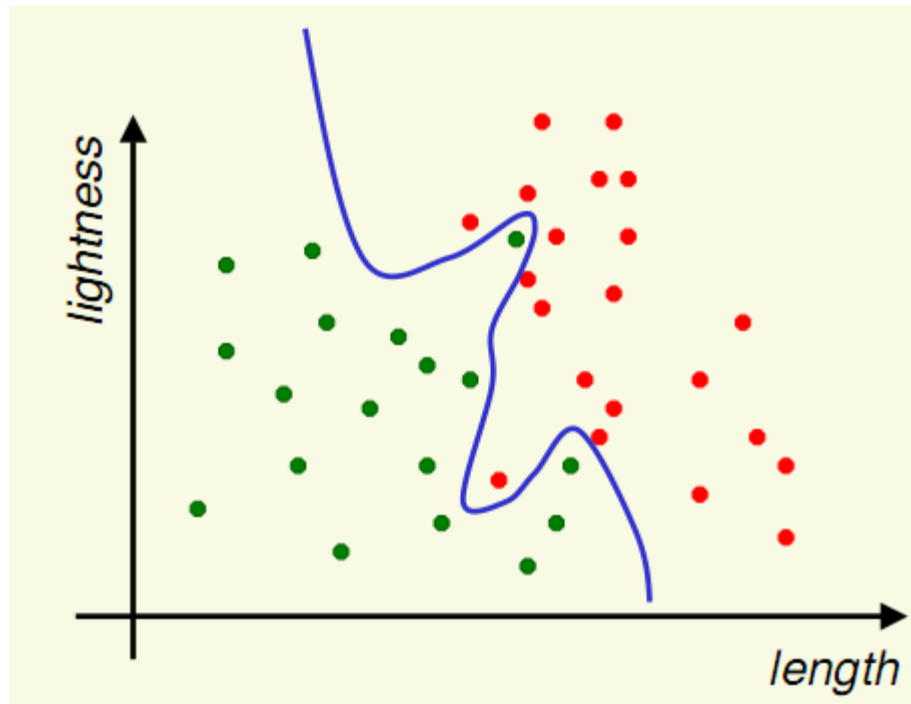
# Öznitelikleri/ayırtaçları birleştir

- Hem parlaklığı hem de uzunluğu kullan
- Öznitelik vektörü: [uzunluk, parlaklık]
- Sınıflandırma hatası: **%4**



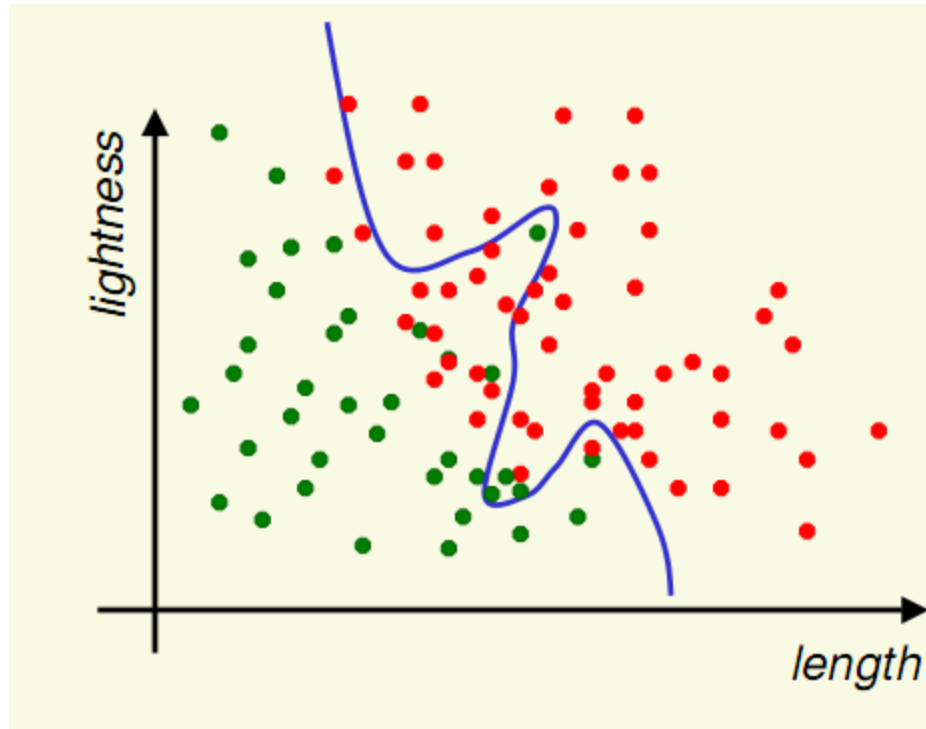
# Daha iyi sınıflandırma sınırı

- Sınıflandırma hatası: **%0**



# Test: yeni verilerle

- Sınıflandırıcınız yeni verilerle de iyi sonuç üretmeli
- “ideal” diye isimlendirdiğimiz %25 hata üretti

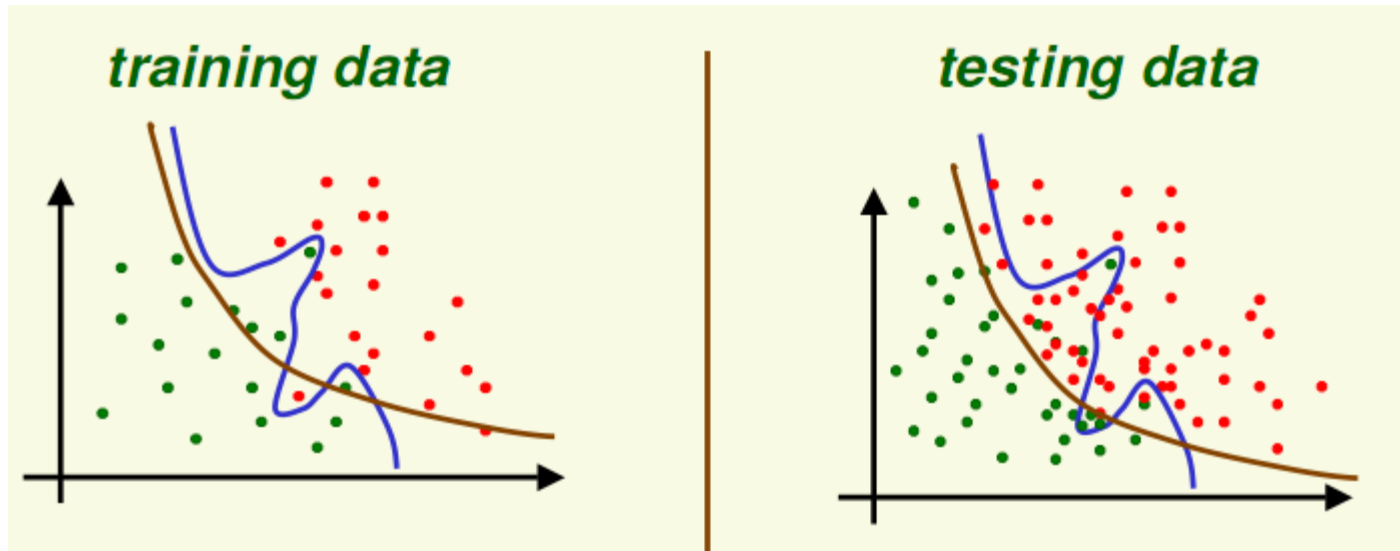


# Nerede hata yaptık?

- Genelleştirme X ezberleme (overfitting)
- Karmaşık sınır bölgesi → ezberleme problemi

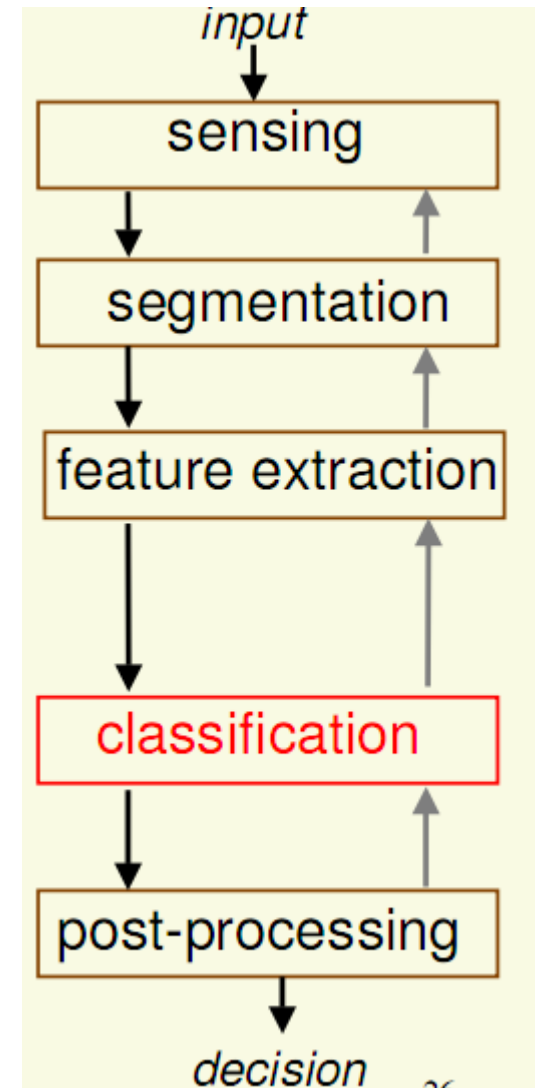
# Genelleştirme

- Basit karar yüzeyi; daha iyi genelleştirme



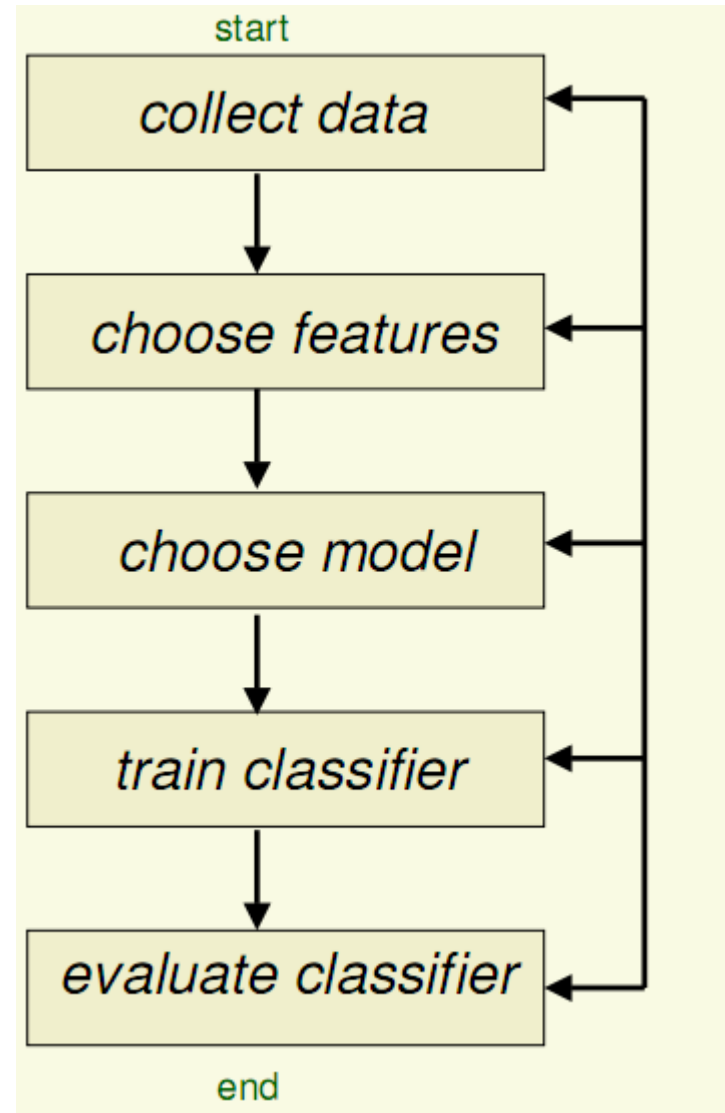
# PR'nin genel yapısı

- Veri topla: kamera vs
- Segmentation: örüntülerle örtüşmemeli
- Öznitelik çıkart: ayırt etme yetisi
- Sınıflandırma: sınıfları belirle, eğit, test et
- Son işleme: “Tne cat” → “The cat”



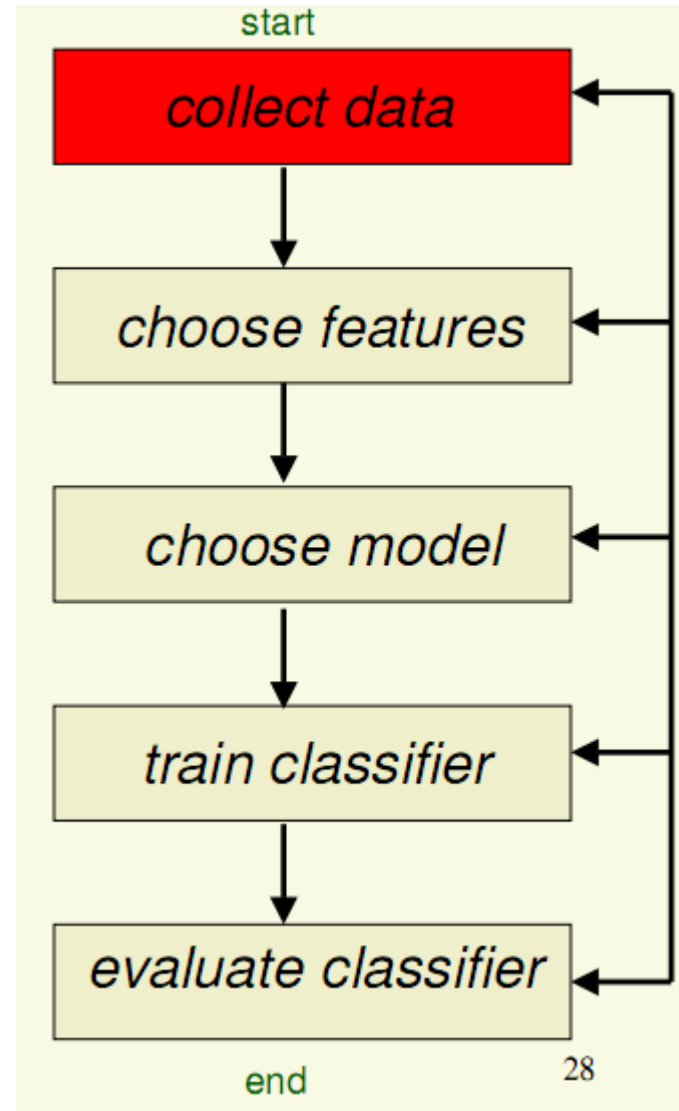
# PR sistemini nasıl tasarlarız?

- Öznitelik ve model ön bilgi gerektirir



# PR sistemini nasıl tasarlarız?

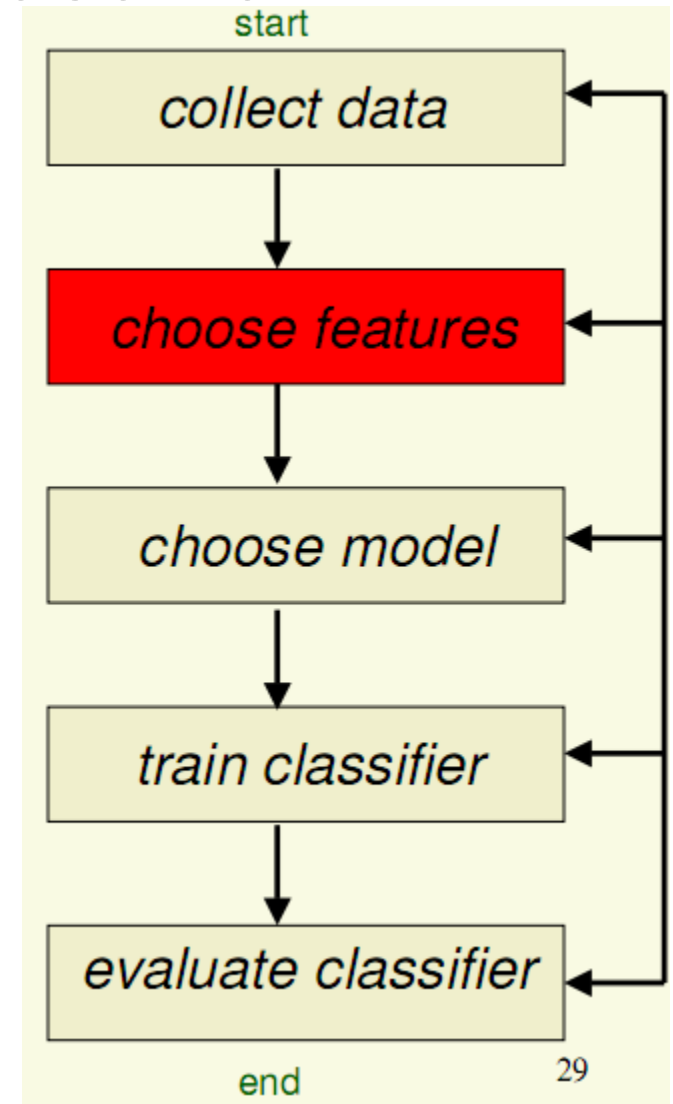
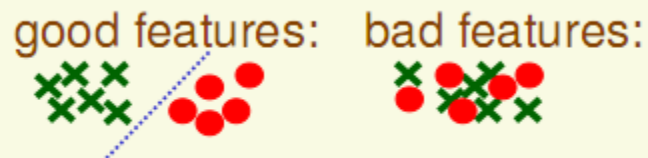
- Veri toplama
  - Maliyetlidir
  - Toplanan veri eğitim ve test örneklerini ne derecede **temsil** ediyor





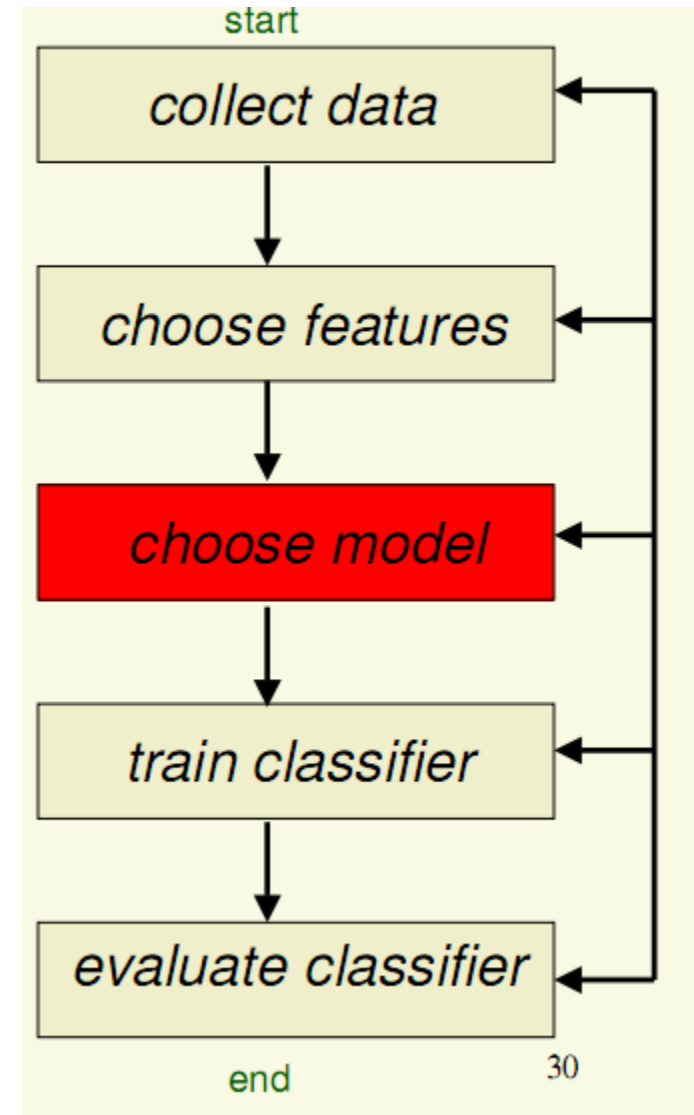
# PR sistemini nasıl tasarlarız?

- Öznitelik çıkartma
  - Ayırt etme yetisi yüksek olan öznitelikler
  - Benzer nesneleri aynı, farklı nesneleri ayrı sınıflandırabilmeli
  - Sınıf-içi varyans küçük ve sınıflar arası varyans büyük olmalı
  - Önbilgi işleri kolaylaştırır



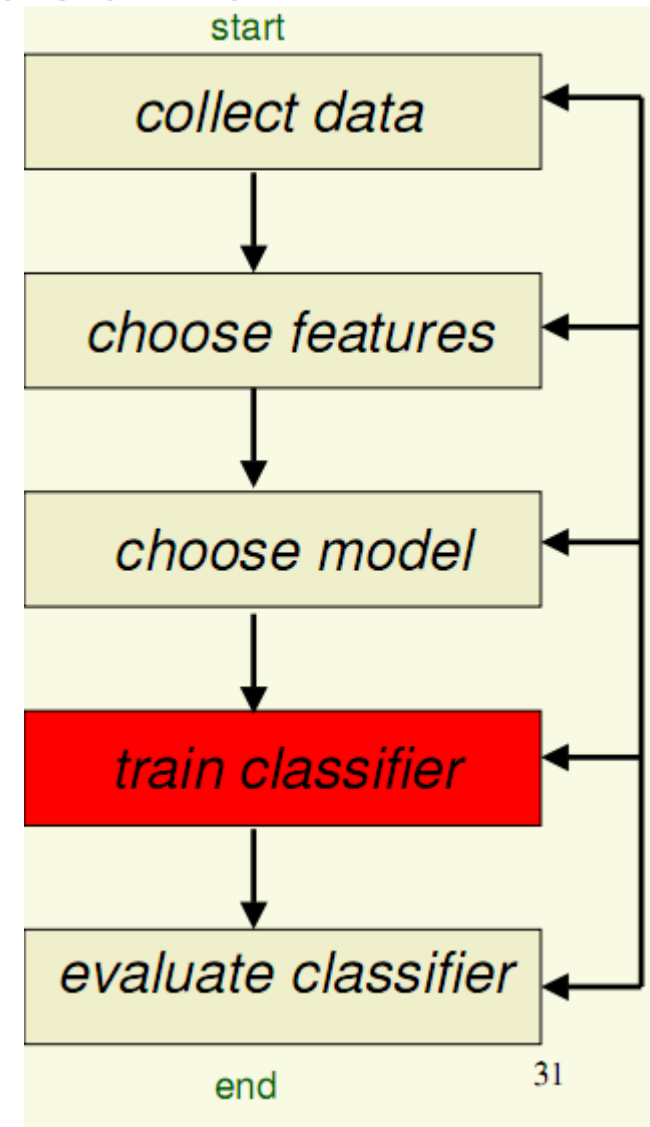
# PR sistemini nasıl tasarlarız?

- Sınıflandırıcının türü
  - Modeli kabul/ret, ne zaman?
  - Problem için en iyi sınıflandırıcı hangisi?



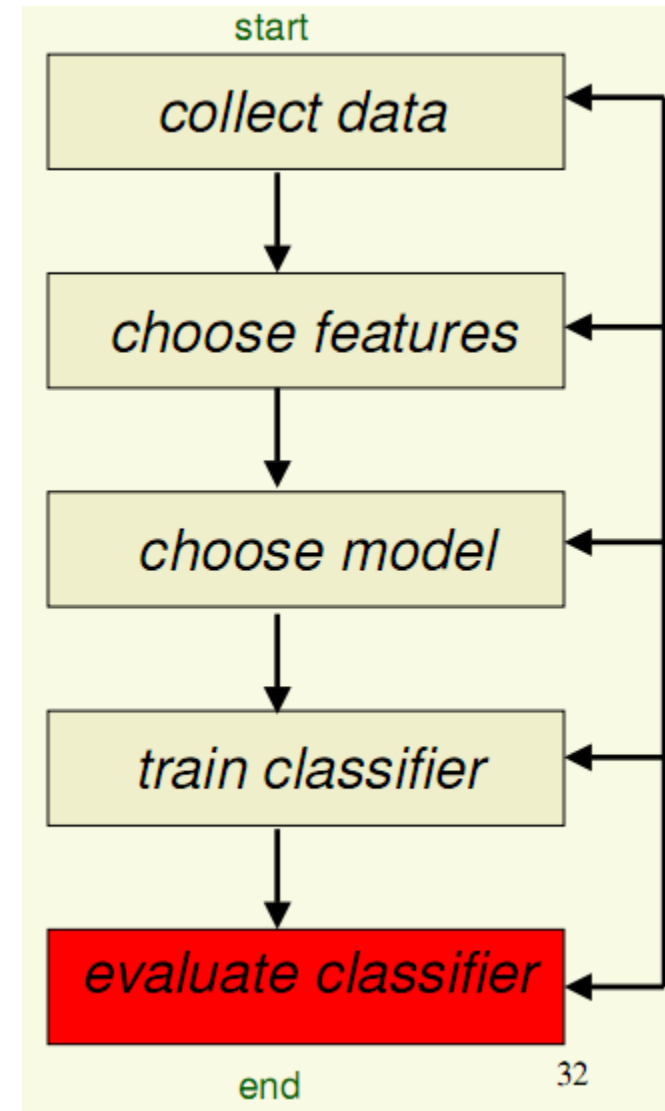
# PR sistemini nasıl tasarlarız?

- Parametre ayarlama
  - Veriye uyacak (fit) şekilde model parametreleri ayarlanır
  - Öğrenme yöntemleri var



# PR sistemini nasıl tasarlarız?

- Öğrendiklerini sınaama
  - Başarım ölçütleri
  - İyileştirme gerekiyor mu?
  - Ezberin önüne geçme
  - Hesapsal karmaşıklık X başarım dengesi

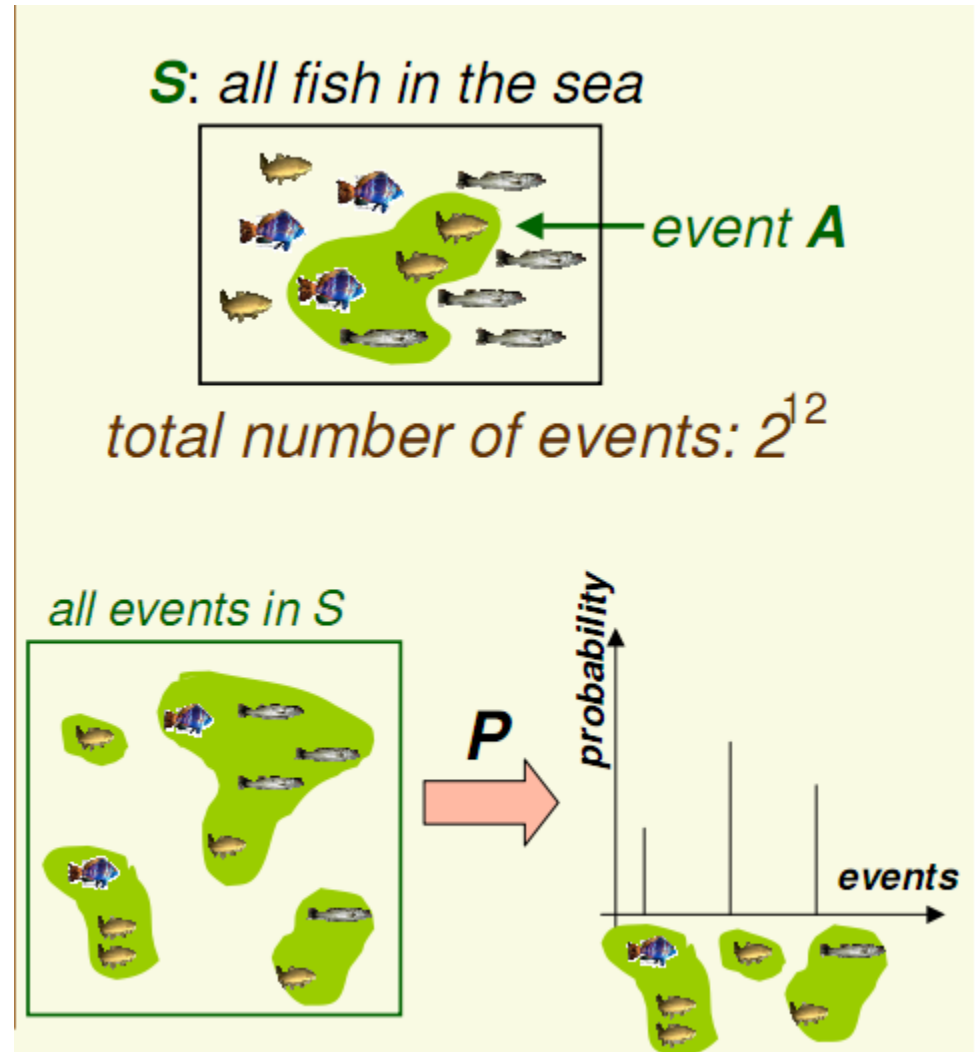


# Özet

- Faydalı
  - Çok sayıda uygulama alanı
- Fakat zor
  - Çözülmesi gereken konular var

# Temel istatistik kavramları

- Seçim öncesi anket
- Anketimiz ne derece sağlıklı
- A olayı (**event**)
- S alt kümesi
- $P(A)$  olasılığı



# Olasılık aksiyomları

- $P(A) \geq 0$
- $P(S) = 1$
- A kesişim B boşsa,  $P(A \cup B) = P(A) + P(B)$

# Olasılık özellikleri

$$P(\emptyset) = 0$$

$$P(A) \leq 1$$

$$P(A^c) = 1 - P(A)$$

$$A \subset B \Rightarrow P(A) < P(B)$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$\{A_i \cap A_j = \emptyset, \forall i, j\} \Rightarrow P\left(\bigcup_{k=1}^N A_k\right) = \sum_{k=1}^N P(A_k)$$

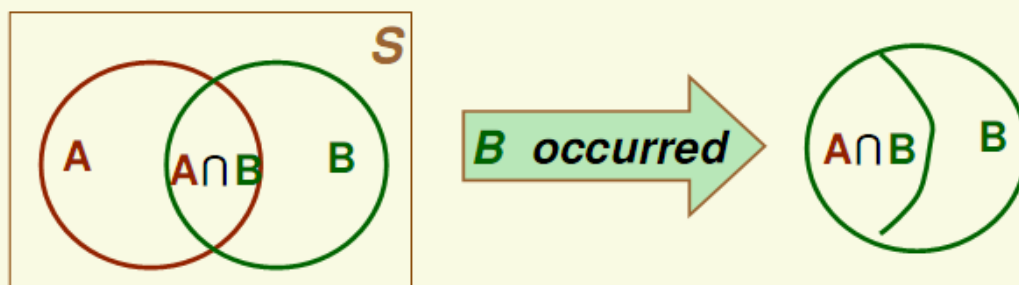
$A^c$ :  
complement  
/eşlenik



# Koşullu olasılık

- If A and B are two events, and we know that event B has occurred, then (if  $P(B) > 0$ )

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$



the “new” sample space is **B**, the “new” **A** is old  $A \cap B$

- multiplication rule

$$P(A \cap B) = P(A/B) P(B)$$

# Koşullu olasılık

- $P(A|B)$ : B biliniyorken, A'nın olasılığı
- $P(A \cap B)$ :
  - Her iki olayın ortak olasılıkları
  - İkisinin de gerçekleşme olasılığı
- $P(A \cap B) / P(B)$ : B olayı olduğunda bunun A olma olasılığı

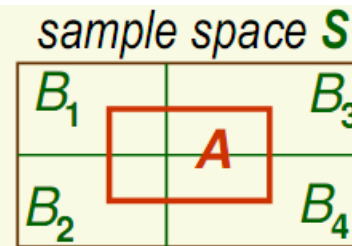
# Bağımsızlık

- Birinin olduğunu bilmek diğerinin olasılığını etkilemez
  - $P(A \cap B) = P(A) P(B)$
  - $P(A \cap B) = P(A | B) P(B)$
- $P(A | B) = P(A) P(B) / P(B) = P(A)$

# Toplam olasılık yasası

- $B_1, B_2, \dots, B_n$  partition  $S$

- Consider an event  $A$



$$A = A \cap B_1 \cup A \cap B_2 \cup A \cap B_3 \cup A \cap B_4$$

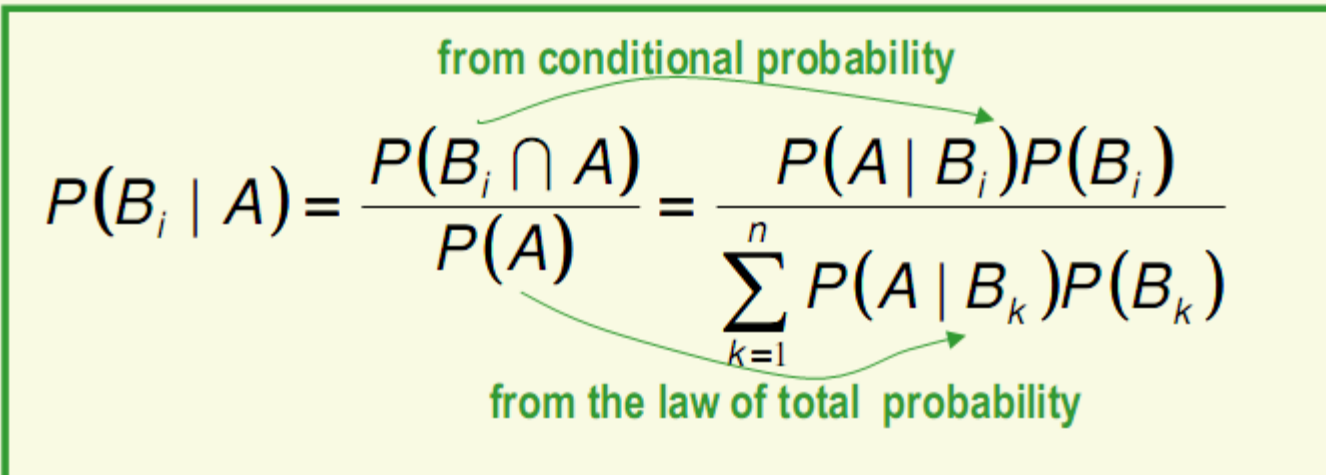
- Thus  $P(A) = P(A \cap B_1) + P(A \cap B_2) + P(A \cap B_3) + P(A \cap B_4)$
- Or using multiplication rule:

$$P(A) = P(A/B_1)P(B_1) + \dots + P(A/B_4)P(B_4)$$

$$P(A) = \sum_{k=1}^n P(A|B_k)P(B_k)$$

# Bayes Teoremi

- A meydana geldiğinde B<sub>i</sub>'leri olasılıkları nedir?
- Cevabı: Bayes Kuralıdır



The diagram shows the derivation of Bayes' Theorem. It features the equation 
$$P(B_i | A) = \frac{P(B_i \cap A)}{P(A)} = \frac{P(A | B_i)P(B_i)}{\sum_{k=1}^n P(A | B_k)P(B_k)}$$
 enclosed in a green rectangular border. A green arrow points from the text "from conditional probability" to the term  $P(A | B_i)$  in the numerator. Another green arrow points from the text "from the law of total probability" to the denominator  $\sum_{k=1}^n P(A | B_k)P(B_k)$ .

$$P(B_i | A) = \frac{P(B_i \cap A)}{P(A)} = \frac{P(A | B_i)P(B_i)}{\sum_{k=1}^n P(A | B_k)P(B_k)}$$

from conditional probability

from the law of total probability

# Bayes Teoremi

- **$P(B|A)$** : A ortaya çıkınca B'nin olasılığı
- **$P(A)$** : A için önsel/marjinal/uç olasılık
- **$P(B)$** : B için önsel/marjinal/uç olasılık
- **Bayes Kuralı**:  $P(A|B)$  ile
  - Eğer **B** gözlemlenmişse, **A** gözlemi hakkındaki inancımızı ne şekilde güncelleriz?
- yanıtını verir

# Örnek

- İki tabak (T1 ve T2) ve bunlar iki tür bisküvi (sade-BS, çikolatalı-BÇ) olsun.
  - T1: 10 BÇ + 30 BS
  - T2: 20 BÇ + 20 BS
- Çocuk rastgele tabaktan rastgele bisküviyi seçsin. Seçilen bisküvi BS ise, bunun T1 olma olasılığı nedir?

# Çözüm

- $P(A|B) = P(B|A) P(A) / P(B)$ 
  - A: T1'den seçim
  - B: sade seçim
  - P(A): çocuğun T1'den seçme olasılığı
    - Eşit seçim hakkı:  $\frac{1}{2} = 0.5$
  - P(B): sade seçme olasılığı
    - T1'den sade seçme + T2'den sade seçme
    - T1:  $P(B|A): \frac{1}{2} * 30 / (10+30) = 0.375$
    - T2:  $P(B|A'): \frac{1}{2} * 20 / (20+20) = 0.25$
    - $0.375 + 0.25 = 0.625$
  - $P(B|A)$ : T1'den seçiliyken, bunda sade seçme olasılığı
    - T1:  $P(B|A): \frac{1}{2} * 30 / (10+30) = 0.375$
- $P(A|B) = 0.375 * 0.5 / 0.625 = 0.6$
- yani %60 olasılıkla seçilen sade bisküvi (B), T1'dendir (A).