

# Resampling for classifier design

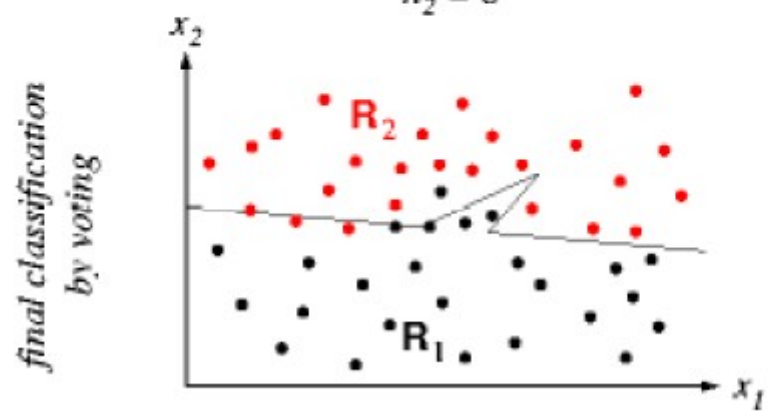
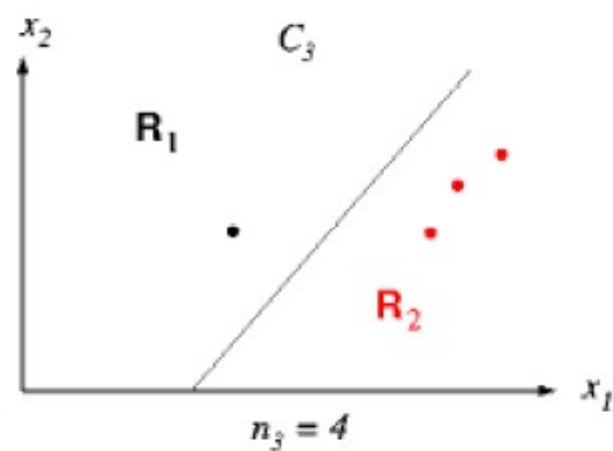
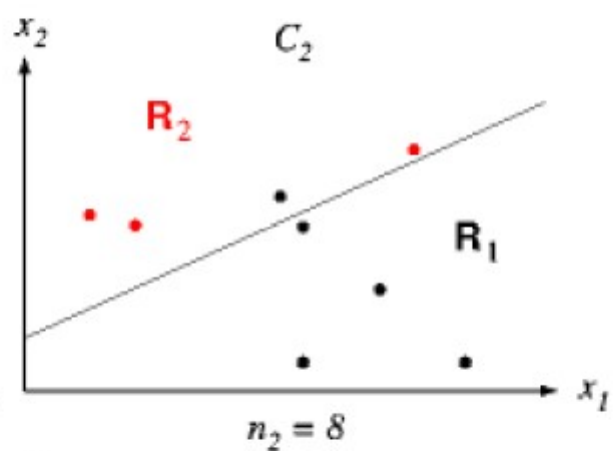
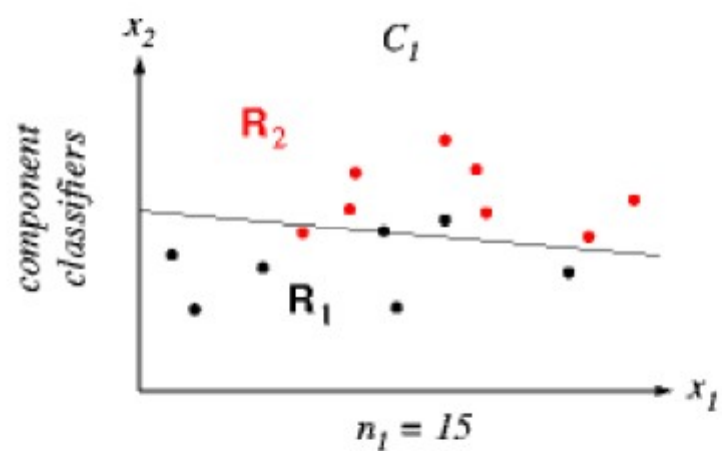
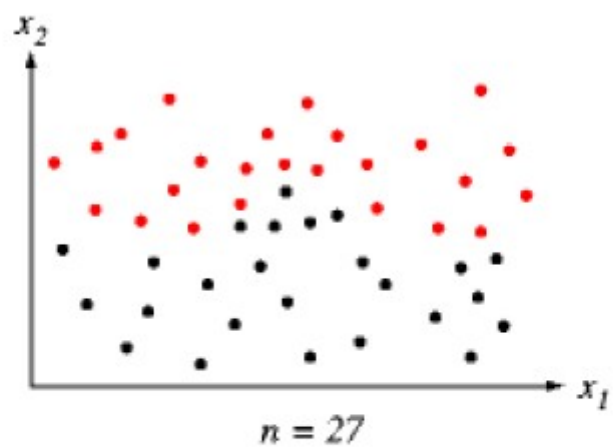
- Bagging (bootstrap aggregation)
- Boosting
  - AdaBoost
- Learning with queries
- Arcing (adaptive reweighting and combining), learning with queries, bias and variance

# Bagging

- Genera multiples training sets.
- Cada clasificador usa un training set.
- La clasificación se efectúa por votación simple.
- Problema: es inestable, pequeños cambios en el data set produce grandes cambios en los resultados.

# Boosting

- Se toma un clasificador mejor que la media para un dataset reducido (pero no muy bueno).
- Se van añadiendo clasificadores aumentando el dataset con lo que conseguimos una mejora.
- La decision se toma tambien por voto simple.



# AdaBoost

- Permite añadir clasificadores debiles asignandoles pesos.
- Podran usarse o no según el trainingset.

```
1 begin initialize  $\mathcal{D} = \{\mathbf{x}^1, y_1, \mathbf{x}^2, y_2, \dots, \mathbf{x}^n, y_n\}, k_{max}, W_1(i) = 1/n, i = 1, \dots, n$   
2      $k \leftarrow 0$   
3     do  $k \leftarrow k + 1$   
4         Train weak learner  $C_k$  using  $\mathcal{D}$  sampled according to distribution  $W_k(i)$   
5          $E_k \leftarrow$  Training error of  $C_k$  measured on  $\mathcal{D}$  using  $W_k(i)$   
6          $\alpha_k \leftarrow \frac{1}{2} \ln[(1 - E_k)/E_k]$   
7          $W_{k+1}(i) \leftarrow \frac{W_k(i)}{Z_k} \times \begin{cases} e^{-\alpha_k} & \text{if } h_k(\mathbf{x}^i) = y_i \text{ (correctly classified)} \\ e^{\alpha_k} & \text{if } h_k(\mathbf{x}^i) \neq y_i \text{ (incorrectly classified)} \end{cases}$   
8     until  $k = k_{max}$   
9     return  $C_k$  and  $\alpha_k$  for  $k = 1$  to  $k_{max}$  (ensemble of classifiers with weights)  
10 end
```

# AdaBoost

- La clasificación será (suma de pesos):

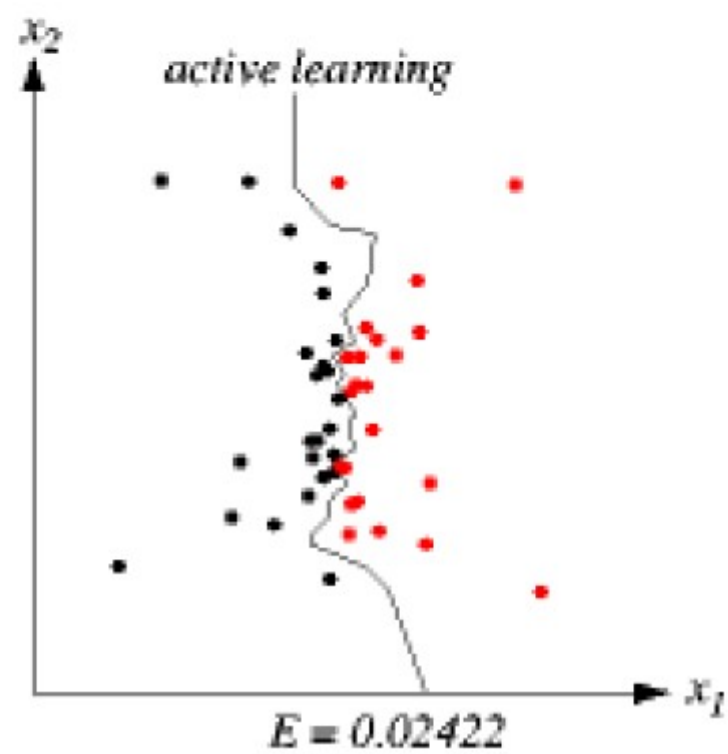
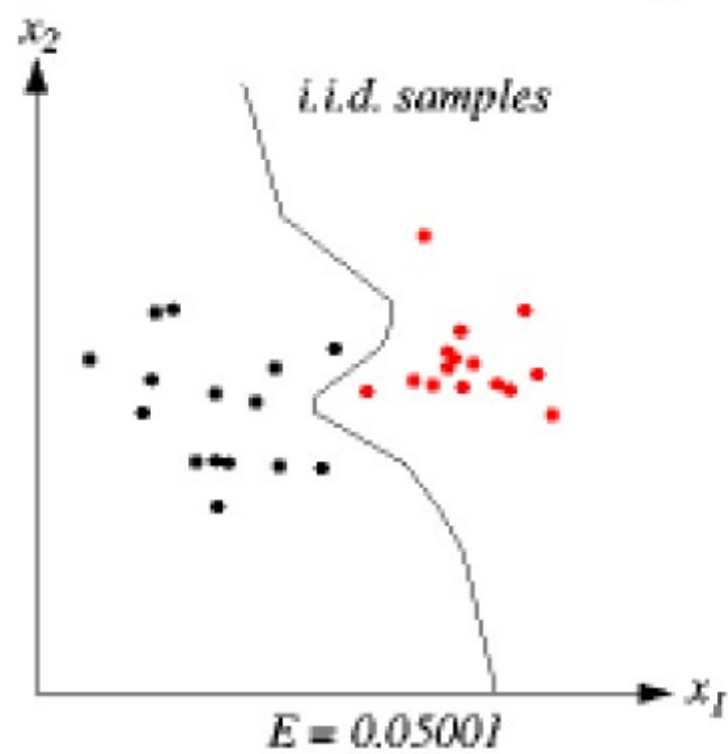
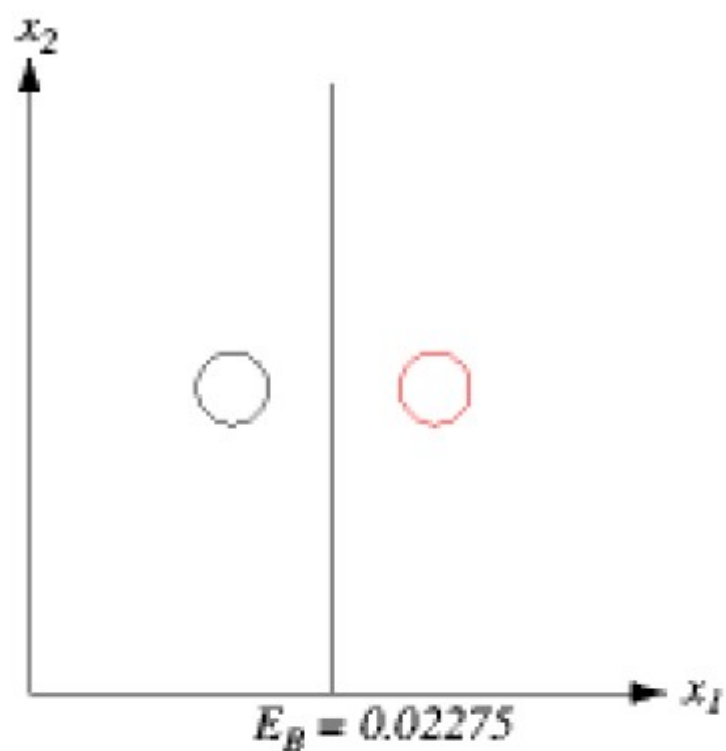
$$g(\mathbf{x}) = \left[ \sum_{k=1}^{k_{max}} \alpha_k h_k(\mathbf{x}) \right]$$

- El error se reduce según:

$$\begin{aligned} E &= \prod_{k=1}^{k_{max}} \left[ 2\sqrt{E_k(1-E_k)} \right] = \prod_{k=1}^{k_{max}} \sqrt{1-4G_k^2} \\ &\leq \exp \left( -2 \sum_{k=1}^{k_{max}} G_k^2 \right) \end{aligned}$$

# Learning with queries

- Cuando no tenemos etiquetas usamos listas de patrones que creemos mas significativos y se lo enseñamos a un oraculo.
- Se toman clasificadores basandose en los patrones hasta dar con el mejor
- Otra forma es crear ensembles.
- Los clasificadores que dan los resultados mas discordantes se consideraran los mejores.





# Arcing, Learning with queries, bias and variance

- Se supone que el remuestreo debe dar peores resultados que sin el, pero no es así.
- Estos metodos son para partes del espacio.
- Cuando el numero de clasificadores es muy grande es dificil sin usar estos metodos.
- Permiten ajustar los clasificadores reduciendo el Bias y la varianza.