Ensemble Methods

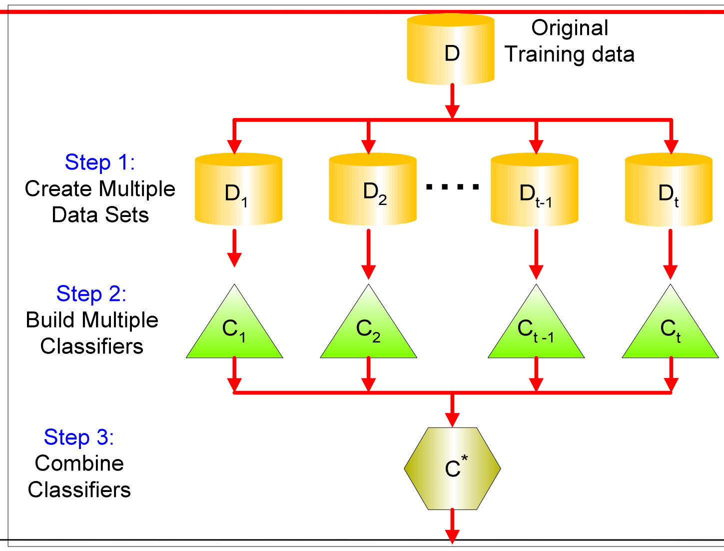
nanopoulos@ismll.de

Ensemble Methods

Construct a set of classifiers from the training data

Predict class label of previously unseen records by aggregating predictions made by multiple classifiers

General Idea



Why does it work?

n independent classifier (2 class problem)

Each has error rate *e*

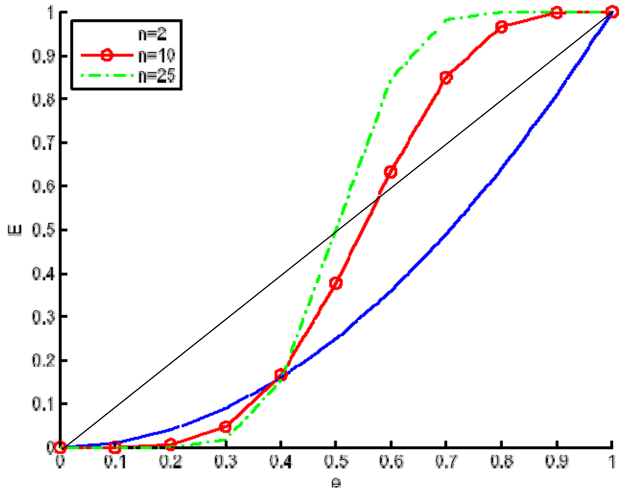
Assign the majority-vote class

Error when more than *n*/2 are eroneous

Expected error E:

$$E = \sum_{i=\lfloor \frac{n}{2} \rfloor + 1}^{n} \binom{n}{i} e^i (1-e)^{n-i}$$

Plot of expected error



n=10, e=0.2, E =0.0064

Examples of Ensemble Methods

How to generate an ensemble of classifiers?

- Bagging
- Boosting

Bagging

Sampling with replacement

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

Build classifier on each bootstrap sample

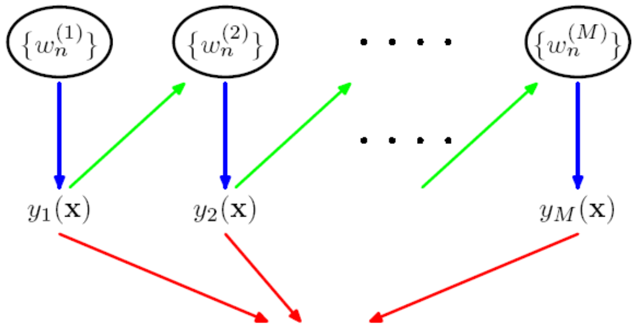
Each sample has probability $1-(1 - 1/n)^n$ of being selected

Boosting

An iterative procedure to adaptively change distribution of training data by focusing more on previously misclassified records

- Initially, all N records are assigned equal weights
- Unlike bagging, weights may change at the end of boosting round

Schematic illustration of boosting



$$Y_M(\mathbf{x}) = \text{sign}\left(\sum_m^M \alpha_m y_m(\mathbf{x})\right)$$

AdaBoost

1. Initialize the data weighting coefficients $\{w_n\}$ by setting $w_n^{(1)} = 1/N$ for $n = 1, \ldots, N$.
2. For $m = 1, \ldots, M$:

   (a) Fit a classifier $y_m(\mathbf{x})$ to the training data by minimizing the weighted error function

$$J_m = \sum_{n=1}^{N} w_n^{(m)} I(y_m(\mathbf{x}_n) \neq t_n) \qquad (14.15)$$

   where $I(y_m(\mathbf{x}_n) \neq t_n)$ is the indicator function and equals 1 when $y_m(\mathbf{x}_n) \neq t_n$ and 0 otherwise.

AdaBoost

(b) Evaluate the quantities

$$\epsilon_m = \frac{\sum_{n=1}^{N} w_n^{(m)} I(y_m(\mathbf{x}_n) \neq t_n)}{\sum_{n=1}^{N} w_n^{(m)}} \quad (14.16)$$

and then use these to evaluate

$$\alpha_m = \ln \left\{ \frac{1 - \epsilon_m}{\epsilon_m} \right\}. \quad (14.17)$$

(c) Update the data weighting coefficients

$$w_n^{(m+1)} = w_n^{(m)} \exp \{ \alpha_m I(y_m(\mathbf{x}_n) \neq t_n) \} \quad (14.18)$$

3. Make predictions using the final model, which is given by

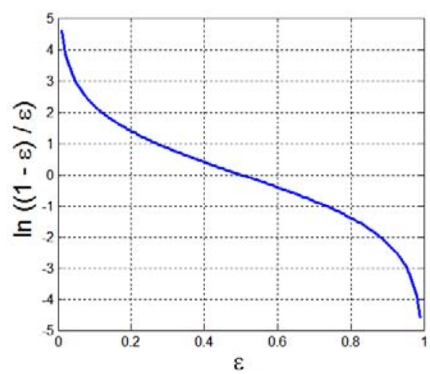
$$Y_M(\mathbf{x}) = \text{sign} \left(\sum_{m=1}^{M} \alpha_m y_m(\mathbf{x}) \right). \quad (14.19)$$

11

Importance of each classifier

- **Importance of a classifier:**

$$\alpha_i = \frac{1}{2} \ln \left(\frac{1 - \varepsilon_i}{\varepsilon_i} \right)$$



Example AdaBoost