Lecture 22 announcements

- HW 12 will be posted.
- Tomorrow's discussion has been pre-recorded
  - will be played back tomorrow at usual time and place
  - archive is available online as usual
- Graded midterm exams - handed back today
- Complete midterm exam solution will be posted
  - Problems 2 and 3 were already given in Discussion 10

Lecture 22 outline

- Boosting and Adaboost (part 2)

BOOSTING (pt. 2)

WANT TO FIND:

$$\hat{f}(\underline{x}) = \underset{f(\underline{x})}{\operatorname{argmin}} \sum_{i=1}^{N} L_{exp}\left[\tilde{y}_i, f(\underline{x}_i)\right]$$

$$\rightarrow \underset{f_0, \beta_m, \underline{\gamma}_m \, \forall m=1,2,\ldots,M.}{\operatorname{argmin}} \sum_{i=1}^{N} L_{exp}\left[\tilde{y}_i, f_0 + \sum_{m=1}^{M} \beta_m \cdot \phi(\underline{x}_i, \underline{\gamma}_m)\right]$$

EASIER TO MIN. EACH TERM, SEQUENTIALLY:

AT m^{th} ITERATION:

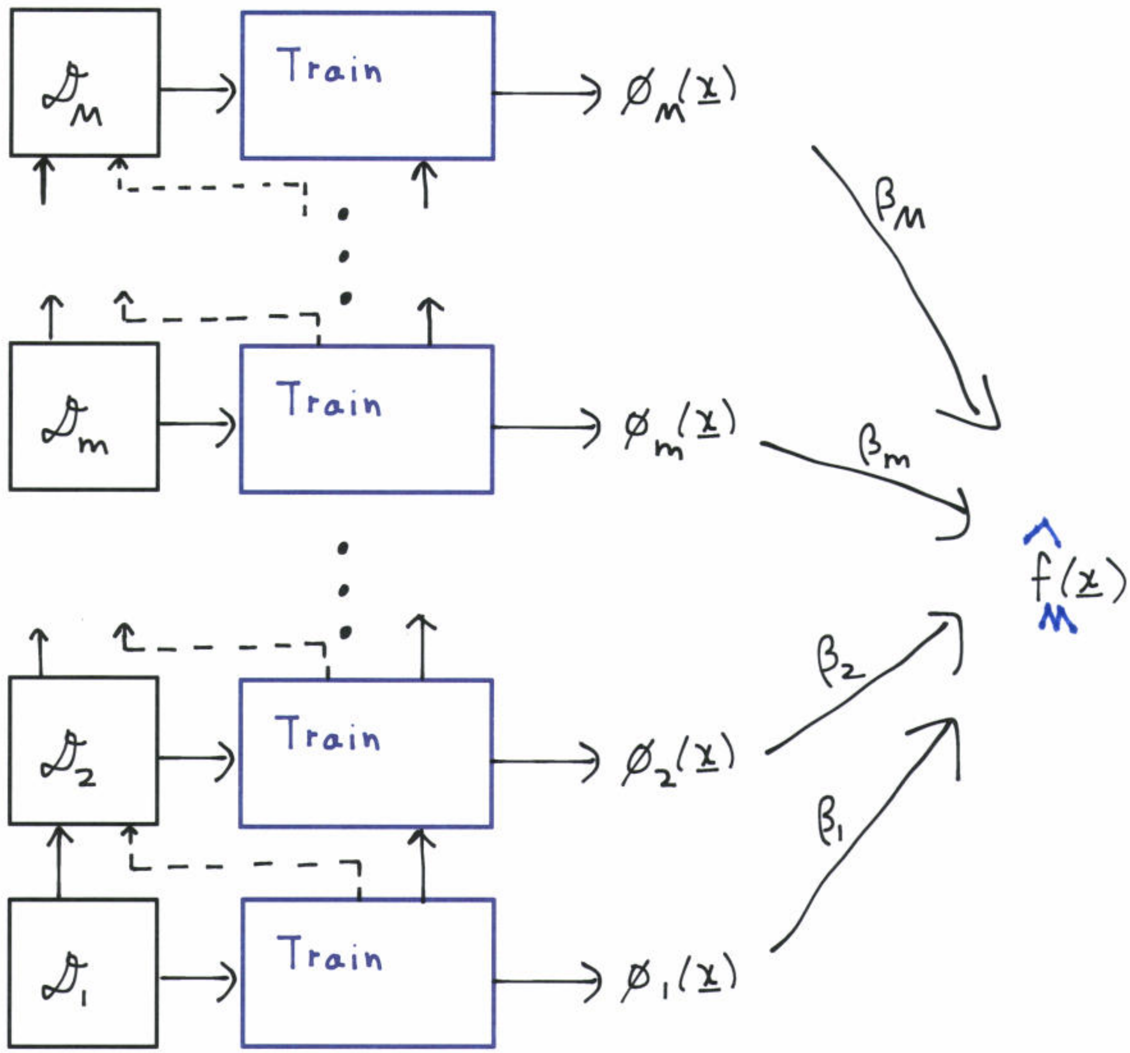
$$\underset{\beta_m, \underline{\gamma}_m}{\operatorname{argmin}} \sum_{i=1}^{N} L_{exp}\left[\tilde{y}_i, \hat{f}_{m-1}(\underline{x}_i) + \beta_m \phi(\underline{x}_i, \underline{\gamma}_m)\right]$$

FORWARD STAGEWISE ADDITIVE MODELING

1. INITIALIZE $f_0 = 0$.
2. FOR $m = 1$ TO M:
   (i) FIND:
   $$(\beta_m, \underline{\gamma}_m) = \underset{\beta_m', \underline{\gamma}_m'}{\operatorname{argmin}} \sum_{i=1}^{N} L\left[\tilde{y}_i, \hat{f}_{m-1}(\underline{x}_i) + \beta_m' \phi(\underline{x}_i, \underline{\gamma}_m')\right]$$
   (ii) $\hat{f}_m(\underline{x}) = \hat{f}_{m-1}(\underline{x}) + \beta_m \phi_m(\underline{x}, \underline{\gamma}_m)$
3. FINAL CLASSIFIER IS:

$$\hat{y}(\underline{x}) = \text{sign}\left\{\hat{f}_M(\underline{x})\right\}$$

L COULD BE L_{exp} OR OTHER LOSS FCN.



$\mathcal{D}_1 = \mathcal{D}_{Tr}$

$$\hat{y}(\underline{x}) = \text{sign}\left\{\hat{f}_M(\underline{x})\right\} = \text{sign}\left\{\sum_{m=1}^{M} \beta_m \phi_m(\underline{x})\right\}$$

ADABOOST

→ USE $L = L_{exp}$

FROM ABOVE:

2.(i) FIND $(\beta_m, \underline{\gamma}_m)$

$$= \arg\min_{\beta_m', \underline{\gamma}_m'} \underbrace{\sum_{i=1}^{N} L_{exp}\left[\tilde{y}_i, \hat{f}_{m-1}(\underline{x}_i) + \beta_m' \phi(\underline{x}_i, \underline{\gamma}_m')\right]}_{L_m}$$

(5) $$L_m = \sum_i \underbrace{\exp\left\{-\tilde{y}_i \left[\hat{f}_{m-1}(\underline{x}_i)\right.\right.}_{w_{i,m}} \left.\left. + \beta_m' \phi(\underline{x}_i, \underline{\gamma}_m')\right]\right\}$$

$w_{i,m}$ (const. of $\beta_m, \underline{\gamma}_m$)

(6) LET $w_{i,m} \triangleq \exp\left\{-\tilde{y}_i \hat{f}_{m-1}(\underline{x}_i)\right\}$

(7) $$L_m = \sum_{i=1}^{N} w_{i,m} \underbrace{\exp\left[-\tilde{y}_i \beta_m' \phi(\underline{x}_j, \underline{\gamma}_m')\right]}$$

sum over all data pts.

weight on i^{th} data pt., at m^{th} iteration.

behaves like regular exp. loss

$\beta_m' > 0.$

L_m CAN BE MINIMIZED ALGEBRAICALLY.
[see Murphy].

CAN RE-ARRANGE L_m ABOVE (see Murphy), to get EQNS. FOR ADABOOST ALGORITHM.

$\Rightarrow$

(*) $\phi(\underline{x}_i, \underline{\gamma}_m)$ IS CHOSEN TO MINIMIZE A SUM OF WEIGHTS OF MISCLASSIFIED DATA POINTS,

(**) $\beta_m = \frac{1}{2} \log \frac{1 - err_m}{err_m}$, with:

(***) $$err_m = \frac{\sum_{i=1}^{N} w_{i,m} \mathbb{I}\left[\tilde{y}_i \neq \phi(x_i, \gamma_m)\right]}{\sum_{i=1}^{N} w_{i,m}}$$

= SAMPLE-WEIGHTED ERROR RATE (AT m^{th} ITERATION)

ALGORITHM: ADABOOST. M1

1. INITIALIZE $w_i = \frac{1}{N} \quad \forall i$.
2. FOR $m = 1$ TO M:

   (i) TRAIN CLASSIFIER $\phi_m(\underline{x})$ [1-NODE CART] ON WEIGHTED DATASET $\mathcal{D}_m$ (WEIGHTS w_i) PER (*).

   (ii) COMPUTE [FROM (***)]:

$$err_m = \frac{\sum_{i=1}^{N} w_i \, \mathbb{I}\left[\tilde{y}_i \neq \phi_m(\underline{x}_i)\right]}{\sum_{i=1}^{N} w_i}$$

   (iii) COMPUTE

$$\alpha_m = \log\left[\frac{1 - err_m}{err_m}\right] = 2\beta_m \text{ in } (**).$$

   (iv) UPDATE

$$w_i \leftarrow w_i \exp\left[\alpha_m \mathbb{I}\left(\tilde{y}_i \neq \phi_m(\underline{x}_i)\right)\right] \quad \forall i.$$

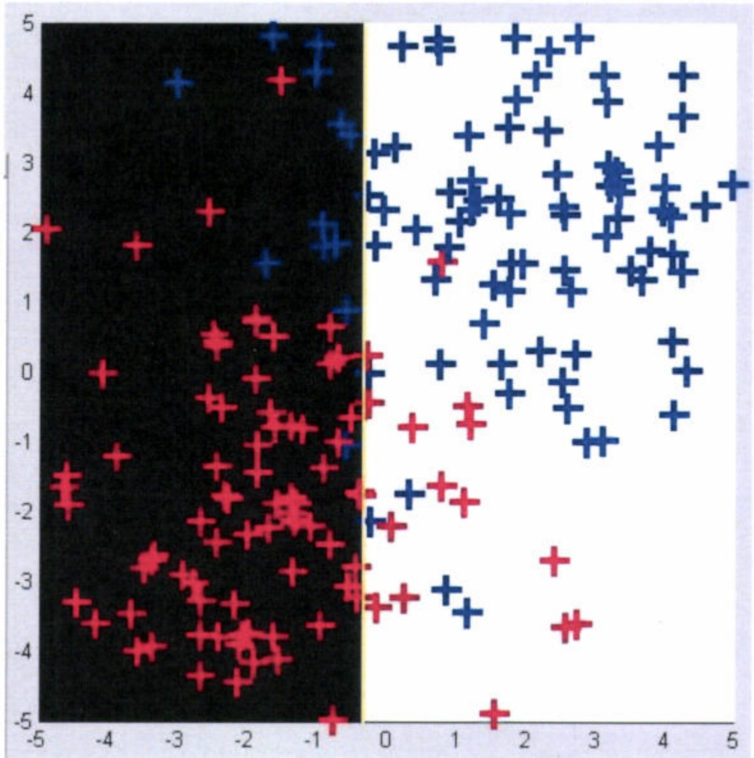
3. RETURN $\hat{f}(\underline{x}) = \sum_{m=1}^{M} \alpha_m \phi_m(\underline{x})$

   AND $\hat{y}(\underline{x}) = \text{sign}\{\hat{f}(\underline{x})\}$.

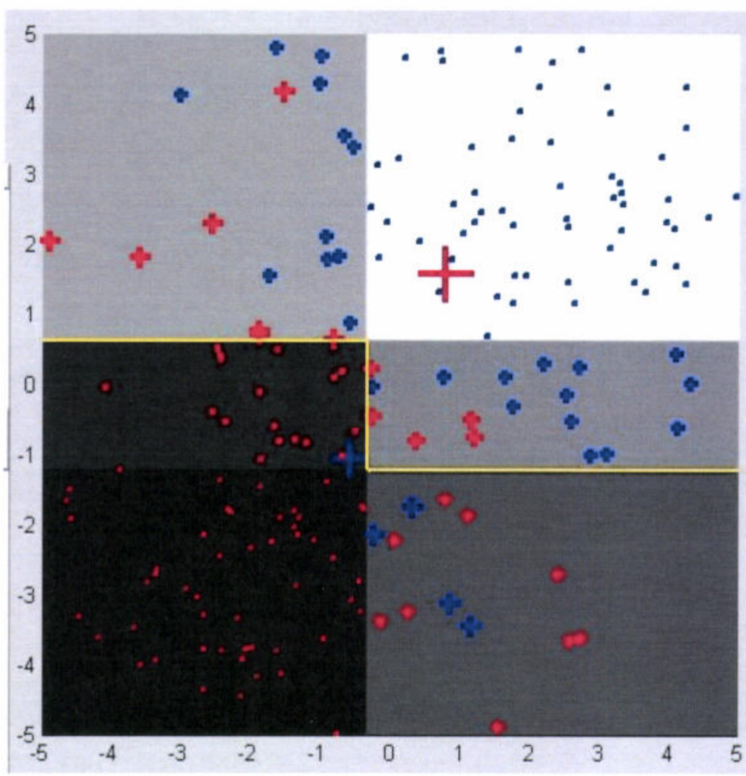
NOTE: ERROR IN MURPHY, ALG. 16.2 (Adaboost.M1)

7 RETURN $\text{sgn}\{f(\mathbf{x})\} = \text{sgn}\left[\sum_{m=1}^{M} \alpha_m \phi_m(\mathbf{x})\right]$

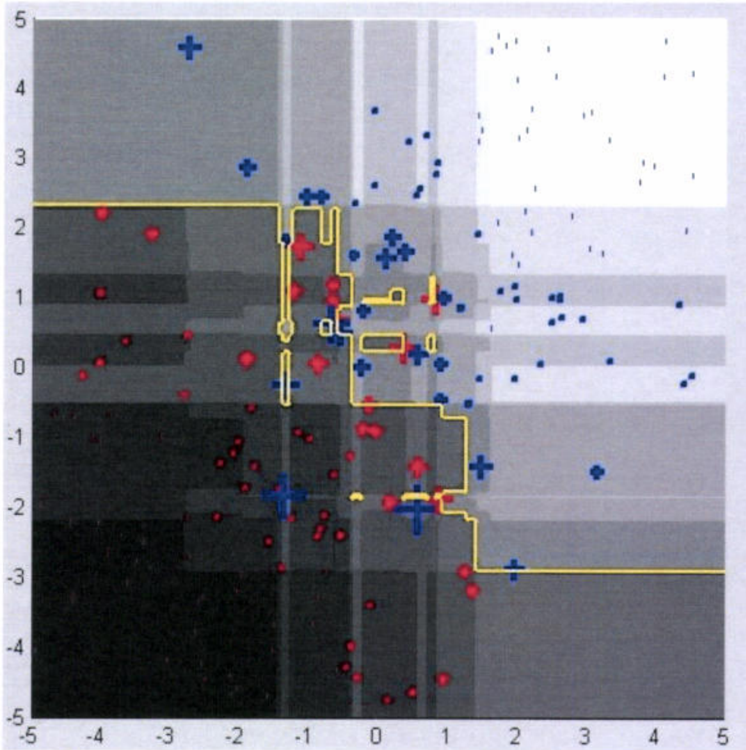
Murphy Fig. 16.10 :



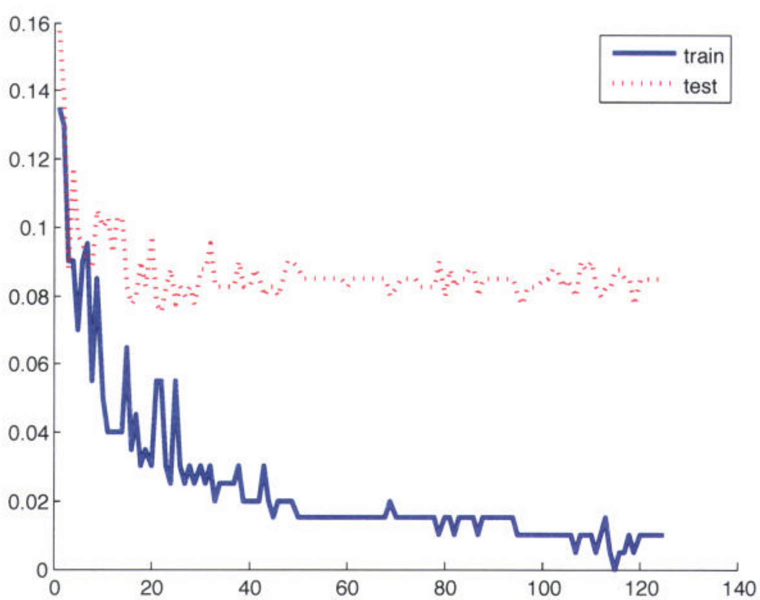
m=1



m=3



m = 120



Murphy Fig. 16.8.

also - [Hastie Fig. 15.1]