E6666 Nov 14

Discussion 13

Today:
- Boosting
- EM for SSL with GMM

Notes:
- No discussion on Thanksgiving (Nov 21)
- No office hours Monday after Thanksgiving (Nov 26)

1) Boosting

Random forest: multiple trees generated indep.

vs

Boosting: each tree depends on previous ones

$$f_m(x) = f_{m-1}(x) + \beta_m \phi(x; \gamma_m)$$

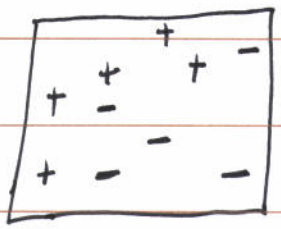
a) Classification - Adaboost

$$L_m(\phi) = \sum_{i=1}^{N} w_{i,m} e^{-\beta \hat{y}_i \phi(x_i)}$$

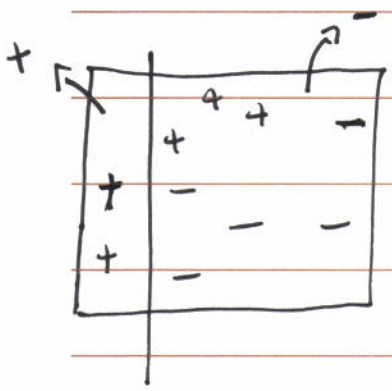
$$w_{i,m} = e^{-\hat{y}_i f_{m-1}(x_i)}$$

$\hat{y}_i \neq f_{m-1}(x_i) \Rightarrow w_{i,m}$ high $\Rightarrow e^{-\beta \hat{y}_i \phi(x_i)}$ get more weight

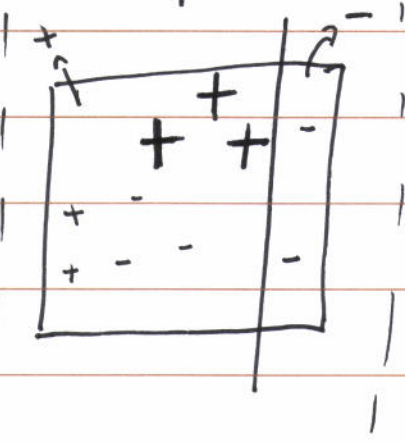
Example:



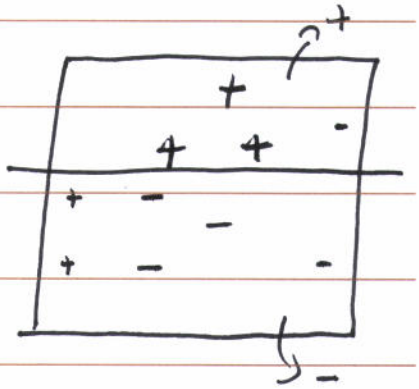
1st step

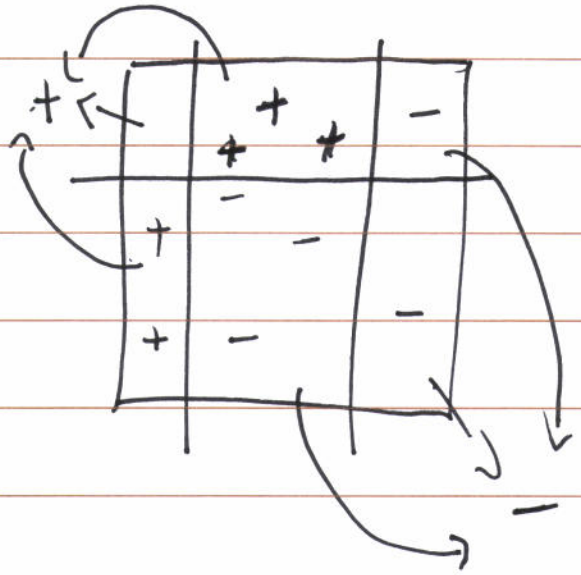


2nd step



3rd





b) Regression

ϕ = stump 

L2 boosting: $L(y_i, f_{m-1}(x_i) + \beta\phi(x_i, \gamma))$

$= (r_{i,m} - \phi(x_i, \gamma))^2$

$\downarrow$

$y_i - f_{m-1}(x_i)$, $\beta = 1$

~~Matlab~~ Matlab simulation

2. EM algorithm: Matlab simulation