$$F(\underline{x}) = a_0 + \sum_{m=1}^{M} a_m h_m(\underline{x})$$

$h_m(\underline{x}) \in \mathcal{F}$ "base learner"

$$\{a_m, h_m(\underline{x})\} = \underset{\{b_m,\, g_m(\underline{x}) \in \mathcal{F}\}}{\arg\min} \frac{1}{N} \sum_{i=1}^{N} L\left(y_i, b_0 + \sum_{m=1}^{M} b_m g_m(\underline{x}_i)\right)$$

$L(y, F)$ = loss function

Very difficult:

<u>Greedy forward stagewise procedure</u>

"Boosting"

"base learner"

$$F(\underline{x}) = a_0 + \sum_{m=1}^{M} a_m h_m(\underline{x})$$

$h_m(\underline{x}) \in \mathcal{F}$ using $L(y, F)$

$$F_0(\underline{x}) = \underset{a}{\operatorname{argmin}} \sum_{i=1}^{N} L(y_i, a)$$

For $m = 1$ to M do {

$$(a_m, h_m(\underline{x})) = \underset{a, h(\underline{x}) \in \mathcal{F}}{\operatorname{argmin}} \frac{1}{N} \sum_{i=1}^{N} L(y_i, F_{m-1}(\underline{x}_i) + a \cdot h(\underline{x}))$$

$$F_m(\underline{x}) = F_{m-1}(\underline{x}) + a_m h_m(\underline{x})$$

}

$$F(\underline{x}) = F_M(\underline{x}) = \sum_{m=1}^{M} a_m h_m(\underline{x})$$

Fundamental boosting algorithm

Forward stagewise regression

$h_m(\underline{x}) \in \mathcal{F}$ "base learner"

Basic step: ave. over training data

minimize $\hat{E}\, L(y, F_{m-1}(\underline{x}) + a\, h(\underline{x}))$
WRT a, $h(\underline{x}) \in \mathcal{H}$

$\hat{E}$ = average over training data

WLG: $\hat{E}\, h(\underline{x}) = 0$; $\hat{E}\, h^2(\underline{x}) = 1$

$$\hat{E}\, L(y, F_{m-1}(\underline{x}) + a\, h(\underline{x}))$$

$$\simeq \hat{E}\left[L(y, F_{m-1}(\underline{x})) + \frac{\partial L}{\partial F}\Big|_{F_{m-1}(x)} a\, h(\underline{x}) \right]$$

$$(a_m, h_m(\underline{x})) = \underset{a,\, h(\underline{x}) \in \mathcal{H}}{\operatorname{argmin}}\ \hat{E}\left[\frac{\partial L}{\partial F}\Big|_{F_{m-1}(x)} a\, h(\underline{x}) \right]$$

let $\ell(y, F_{m-1}(\underline{x})) = \frac{\partial L}{\partial F}\Big|_{F_{m-1}(x)}$

$$(a_m, h_m(\underline{x})) = \underset{a,\, h(\underline{x}) \in \mathcal{H}}{\operatorname{argmin}}\ \hat{E}\left[\ell(y, F_{m-1}(\underline{x}))\, a\, h(\underline{x}) \right]$$

note $\tilde{y} = -\ell(y, F_{m-1}(\underline{x}))$

$$(a_m, h_m(\underline{x})) = \underset{a,\ h(\underline{x}) \in \mathcal{F}}{\arg\max}\ \hat{E}[\tilde{y}\, a\, h(\underline{x})]$$

for $a > 0$ (WLG)

$$h_m(\underline{x}) = \underset{h(\underline{x}) \in \mathcal{F}}{\arg\max}\ \hat{E}[\tilde{y} \cdot h(\underline{x})]$$

$$= \underset{h(\underline{x}) \in \mathcal{F}}{\arg\max}\ \mathrm{corr}[\tilde{y}, h(\underline{x})]$$

$$= \underset{\substack{h(\underline{x}) \in \mathcal{F} \\ \beta}}{\arg\min}\ \hat{E}[\tilde{y} - \beta\, h(\underline{x})]^2$$

$$a_m = \underset{a}{\arg\min}\ \hat{E}\, L(y, F_{m-1}(\underline{x}) + a\, h_m(\underline{x}))$$