Generalized Additive Models.

$$E[Y|X] = \beta_0 + f_1(X_1) + f_2(X_2) + \dots + f_p(X_p)$$

$$\underbrace{E[Y|X]}_{\mu(X)}$$

GLMs: $\mu(X)$

$$g(\mu(x)) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$

identity function

"log linear" $g(\;) = \log(\;\;)$ $\qquad g(a) = \log\left(\frac{a}{1-a}\right)$ "logit"

Tree Based Methods.

CART

$$\hat{f}(x) = \sum_{m=1}^{M} c_m \cdot \mathbb{1}[x \in R_m]$$

$$\boxed{\{c_m, R_m\}_{m=1}^{M}}$$

$$\min_{\{c_m, R_m\}_{m=1}} \sum_{i=1}^{N} \left(y_i - \hat{f}(x_i)\right)^2$$

8

Missing Data

MAR : $\mathbb{P}_\theta(R \mid \text{data}) = \mathbb{P}_\theta\left(R \,\middle|\, \begin{matrix}\text{observed}\\ \text{data}\end{matrix}\right)$ data $= X_{N\times p}, Y$

$R_{N\times p}$

MCAR : $\mathbb{P}_\theta(R \mid \text{data}) = \mathbb{P}(R)$

$$f_0(x) = \underset{\gamma}{\operatorname{argmin}} \left[L(y_1, \gamma) + L(y_2, \gamma) + \cdots + L(y_N, \gamma) \right]$$

$$\frac{\partial L(y_1, \alpha)}{\partial \alpha} = \frac{\partial (y_1 - \alpha)^2}{\partial \alpha} = -2(y_1 - \overset{f_0(x)}{\alpha})$$

1. Bagging & RF
2. MARS
3. SVC

$$
\begin{array}{ccccc}
 & \nearrow & Z^{*1} & \longrightarrow & \hat{f}^{*1}(x) \\
Z & \rightarrow & Z^{*2} & & \vdots \\
 & & \vdots & & \\
 & \searrow & Z^{*B} & & \hat{f}^{*B}(x)
\end{array}
\Bigg\}
\quad
\hat{f}_{bag}(x) = \frac{1}{B}\sum_{b=1}^{B} \hat{f}^{*b}(x)
$$

Var

$$E_\tau\left(\hat{f}_\tau(x) - E_\tau \hat{f}_\tau(x)\right)^2$$

① $W_1 \ldots W_B$ iid RVs $\quad \mathrm{Var}\, V_i = \sigma^2$

$$\mathrm{Var}\left(\frac{1}{B}\sum_{i=1}^{B} W_i\right) = \frac{1}{B^2}\left[V_1 + V_2 + \cdots V_B\right]$$

$$= \frac{B \cdot \sigma^2}{B^2} = \frac{\sigma^2}{B}$$

$$\mathrm{Var}\left(\frac{A+B}{10}\right)$$

$$= \frac{1}{10^2}\left(\mathrm{Var}\, A + \mathrm{Var}\, B\right)$$

(2) $W_1, \dots, W_B$ $\quad \text{Corr}(W_i, W_j) = \rho$

$$\text{Corr}(W_i, W_j) = \frac{\text{Cov}(W_i, W_j)}{\sqrt{\text{Var}(W_i)} \cdot \sqrt{\text{Var}(W_j)}}$$

$$\text{Var}(W_i) = \sigma^2$$

$$\frac{1}{B^2}\text{Var}\left(\sum_{b=1}^{B} W_b\right) = \frac{1}{B^2}\left[\sum_{b=1}^{B} \text{Var}(W_b) + 2\sum_{b_1=1}^{B}\sum_{b_2=1}^{b_1-1} \text{Cov}(W_{b_1}, W_{b_2})\right]$$

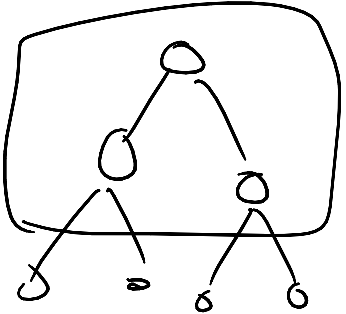
$$= \frac{1}{B^2}\left[B\sigma^2 + 2 \cdot \frac{B(B-1)}{2} \cdot \rho \cdot \sigma^2\right] = \frac{\sigma^2}{B} + \frac{(B-1)}{B} \cdot \rho\sigma^2$$

Random Forest : $m < p$

select random

Interpretation :

$$I_\ell^2 = \sum_{t=1}^{J} \hat{i}_t^2 \, \underbrace{\mathbb{1}[v_t = \ell]}$$



Partial dependence plot

$$\underline{\underline{f(X)}} = f(\underline{\underline{X_S}}, X_C)$$

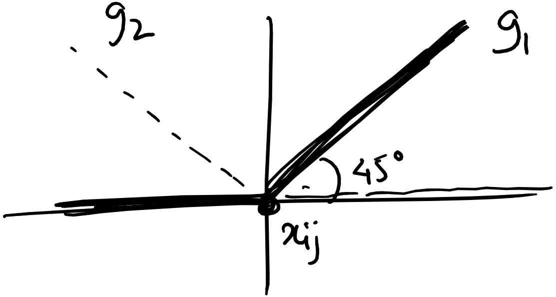
$$f_S(X_S) = E_{X_C}\left[f(X_S, X_C) \right]$$

$$\neq f(X_S, E_{X_C}[X_C])$$

MARS

$$f(X) = \beta_0 + \sum_{m=1}^{M} \beta_m h_m(X)$$

$$X = \left[\; x_{ij} \;\right]_{N \times p}$$



$$\underbrace{\begin{aligned} g_1(X) &= (X_j - x_{ij})_+ \\ g_2(X) &= (x_{ij} - X_j)_+ \end{aligned}}_{\mathcal{C}}$$

Goal: $\beta_0 + \sum^{M} \beta_m h_m(X)$

$\mathcal{C}$

$\mathcal{M}$

$h_0(X) = 1$

$$\hat{\beta}_1 \cdot 1 \cdot \underbrace{(X_j - x_{ij})_+}_{g_1(X)} + \hat{\beta}_2 \cdot 1 \cdot \underbrace{(x_{ij} - X_j)_+}_{g_2(X)}$$

$h_0(X)$

$h_1(X) = (X_2 - x_{72})_+$

$h_2(X) = (x_{72} - X_2)_+$

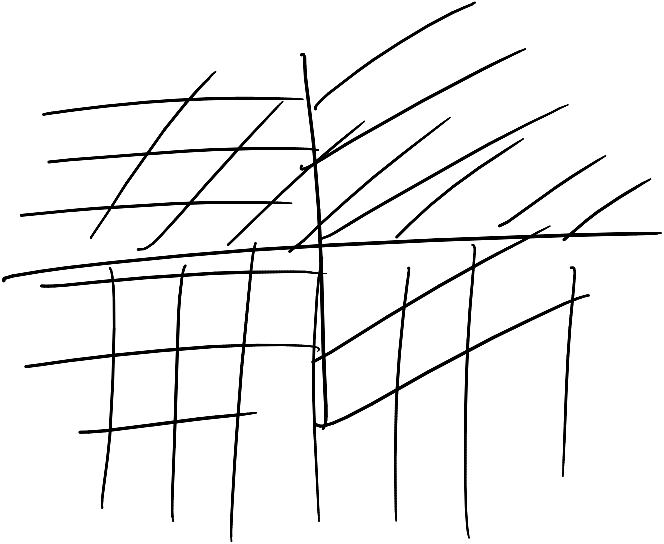
h_0

h_1

h_2

$h_3 = (X_2 - x_{72})_+ (X_5 - x_{25})_+$

h_4 " $\cdot$ X



SVC

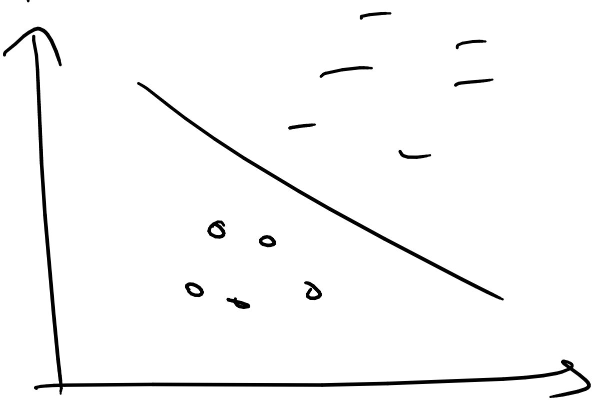
2 Class Classification.

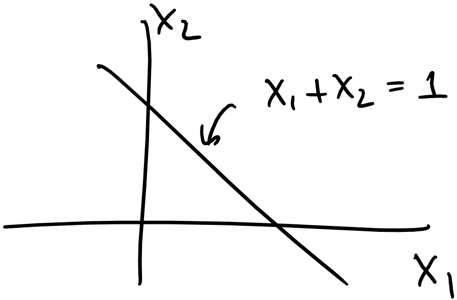
$y_i \in \{-1, 1\}$

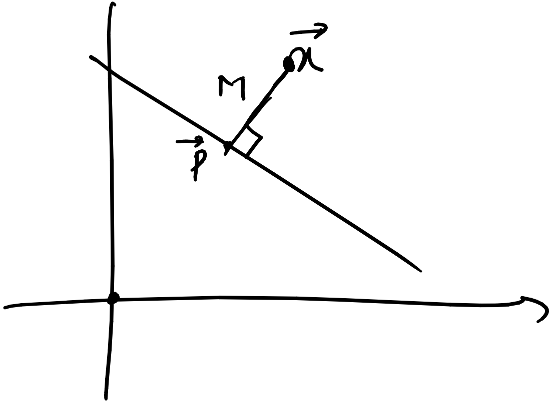
$\{x : f(x) = 0\}$

$\{x : x^T\beta + \beta_0 = 0\}$

Separability







$$x^T\beta + \beta_0 = 0$$

$$\vec{x}\ ,\ \vec{p}$$

$$\vec{p}^{\,T}\beta + \beta_0 = 0$$

$$\vec{x} = \underline{\underline{\vec{p}}} + M \cdot \frac{\vec{\beta}}{\|\beta\|_2}$$

$$\left(\vec{x} - \frac{M\vec{\beta}}{\|\beta\|_2}\right)^T \beta + \beta_0 = 0$$

$$x^T\beta - M\,\|\beta\|_2 + \beta_0 = 0$$

$$\Rightarrow \quad \frac{x^T\beta + \beta_0}{\|\beta\|_2} = M$$

$$\max_{\beta, \beta_0, M} 2M$$

$$y_i \left(\frac{x^T \beta + \beta_0}{\|\beta\|_2} \right) \geq M \quad i = 1 \ldots N.$$