Bagging:

$$Z \longrightarrow Z^{*1}, \ldots\ldots, Z^{*B} \qquad Z = (X, Y)$$

$N \times p \qquad N \times 1$

$$f^{*1}(X) \ \ldots\ldots \ f^{*B}(X)$$

$$\hat{f}_{bagging}(X) = \frac{1}{B}\sum_{i=1}^{B} f^{*i}(X)$$

$$f^{baseline}(X)$$

$W_1, \ldots, W_B$ $\underline{i}$.i.d μ, σ^2

$P(W_1, W_2) = P(W_1) \cdot P(W_2)$

$$\mathrm{Var}\left(\frac{1}{B} \sum_{i=1}^{B} W_i\right) = \frac{1}{B^2} \sum_{i=1}^{B} \mathrm{Var}(W_i) = \frac{\sigma^2}{B}$$

$$\mathrm{Var}(aV) = a^2 \mathrm{Var}(V)$$

$W_1, \ldots, W_B$ id $\underline{\mathrm{Corr}}(W_i, W_j) = \rho$

$$\frac{1}{B^2} \underbrace{\mathrm{Var}\left(\sum_{i=1}^{B} W_i\right)} = \frac{1}{B^2}\left[\sum_{i=1}^{B} \mathrm{Var}(W_i) + 2 \underbrace{\sum_{i=1}^{B} \sum_{j=1}^{i-1} \underline{\mathrm{Cov}}(W_i, W_j)}\right]$$

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

$$= \frac{\sigma^2}{B} + \underbrace{\left(\frac{B-1}{B}\right)}_{\approx 1} \cdot \rho\,\sigma^2$$

RF

$$\left[\; \underbrace{f^{*1}} \quad \cdots\cdots \quad \underbrace{f^{*B}} \;\right]$$

$\boxed{m < p}$ at each split

OOB : out of bag samples.

$\underline{\underline{x_i, y_i}}$

Interpretation:

Variable importance scores:

X_ℓ

$$\underline{\underline{I_\ell}} = \sum_{t=1}^{J} \hat{i}_t \, \underbrace{\mathbb{1}[v(t) = \underline{\ell}]}$$

Partial dependence plots:

$$f(X_s, X_c)$$

$$\overbrace{\text{age} \quad \text{edu}}^{X_s} \;\Big|\; \overbrace{\text{income} \quad \cdots}^{X_c}$$

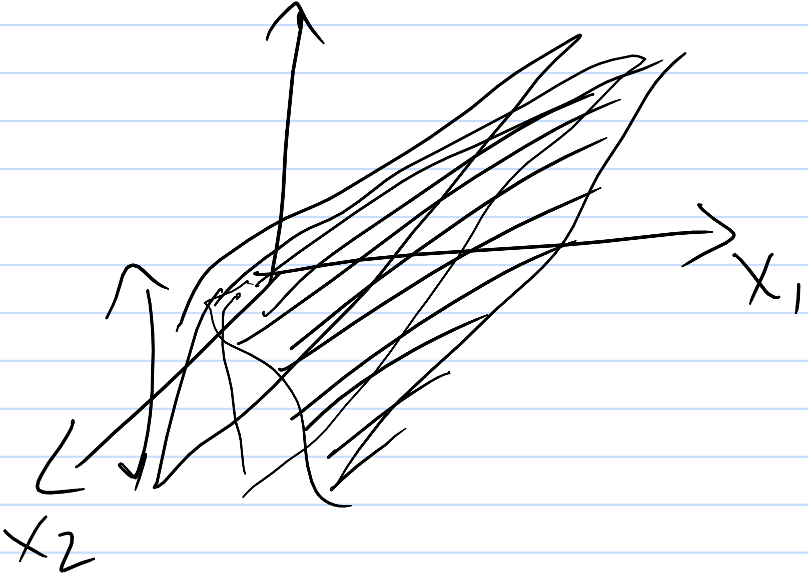
Partial dependences

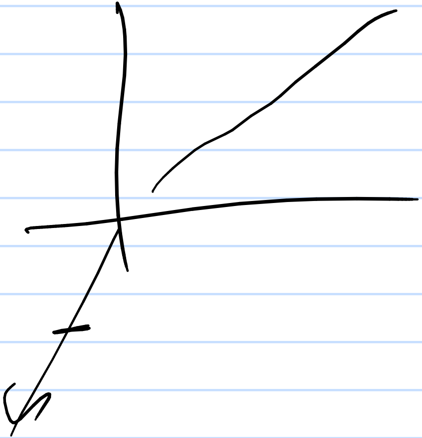
$f(X)$ on $\{\text{age}, \text{edu}\} \equiv X_s$

$$f_s(X_s) = E_{X_c}[f(X)]$$

$$\approx \frac{1}{N} \sum_{i=1}^{N} f(X_s, x_{ic})$$

$$f\left(X_s, \frac{1}{N}\sum \text{income}, \frac{1}{N}\sum 0\right)$$


x_1
x_2



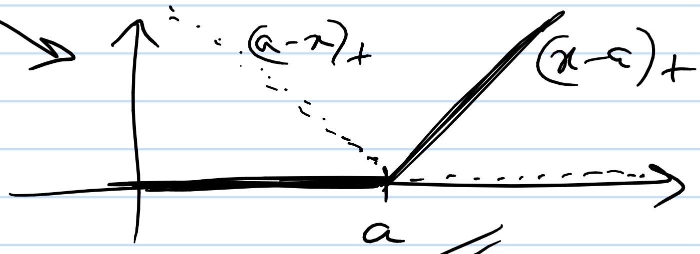
2 MARS Regression

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

$\mathcal{C}$ g functions
base

$$(x-a)_+, \quad (a-x)_+$$

$$\max(0, x-a), \quad \max(0, a-x)$$



X

$N \times p$

$$\left\{ \begin{array}{ll} (X_1 - x_{11})_+ & (X_j - x_{ij})_+ \\ (X_1 - x_{21})_+ & \\ \vdots & \end{array} \right\}$$

$N \times p \times 2$

$\mathcal{C}$ = {MARS function $(x-a)_+, (a-x)_+$}

$\beta_1 \quad \beta_2 = RSS_1$

$\beta_1 (x-a')_+ \; \beta_2' (a'-x)_+ = RSS_2$

$$\boxed{\mathcal{M} = \{ \underline{h_m}(X) \}}$$

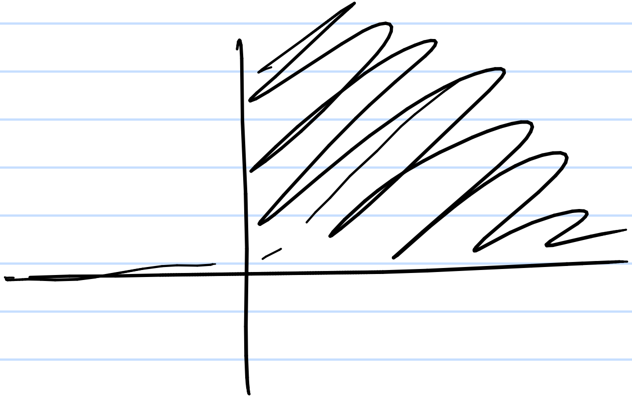
1. add pair of functions such that RSS is minimized.

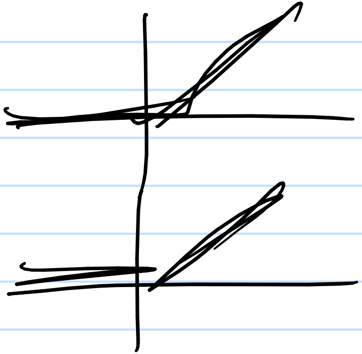
eg: $\underbrace{(X_2 - x_{72})_+}_{h_1}, \underbrace{(x_{72} - X_2)_+}_{h_2}$

2. $\underbrace{(X_2 - x_{72})_+ \cdot (X_1 - x_{51})_+}_{h_3} \quad \underbrace{(X_2 - x_{72})_+ \cdot (x_{51} - X_1)_+}_{h_4}$

$(X_1 - a_1)_+ \; (a_2 - X_5)_+ \; \cdots \; (a_{10} - X_{11})_+$

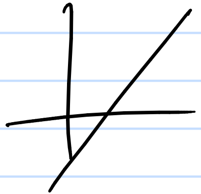
$E[Y \mid X = x]$





SVM

linear 2 class classification. $\quad Y \in \{-1, 1\}$



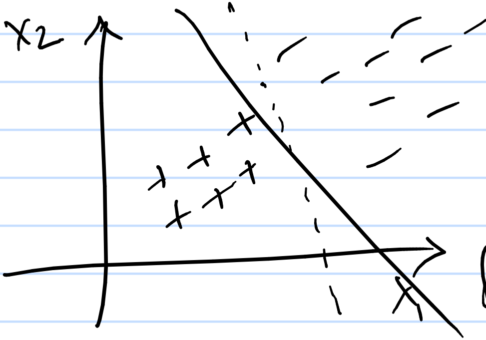
$$f(x) = x^T\beta + \beta_0$$

$$\{x : f(x) = \beta^T x + \beta_0 = 0\} \text{ : decision boundary}$$

it is a hyperplane

sign$(f(x))$

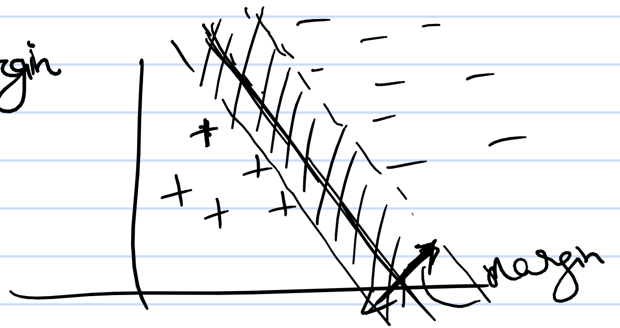
1. seperable data
2. Relax $\uparrow$



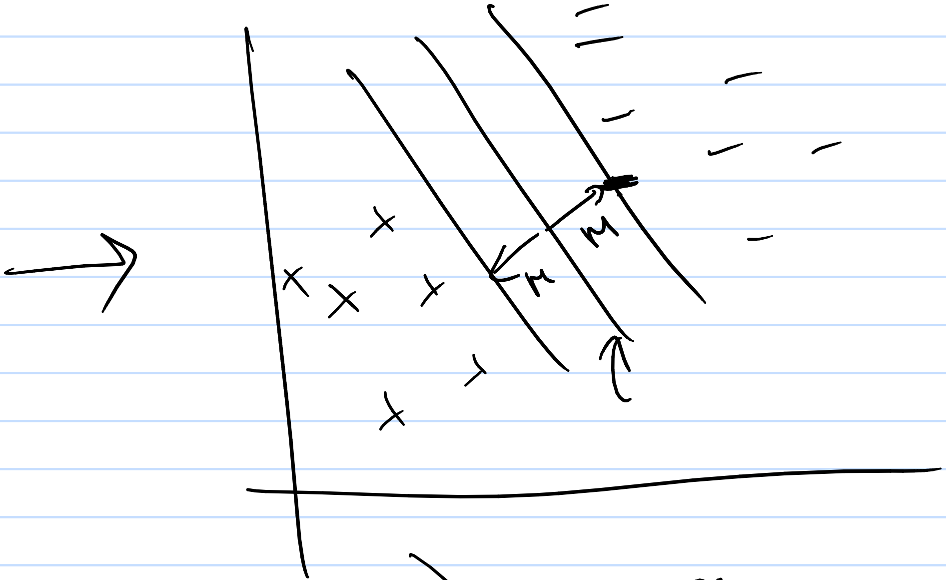
$$\beta_1 x_1 + \beta_2 x_2 + \beta_0 = 0$$

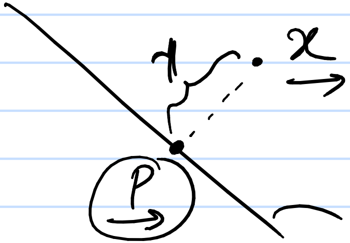
Which hyperplane to choose

Geometric margin



" get hyperplane that has the maximum margin "



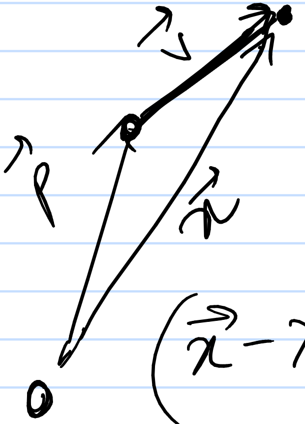


$$\beta^T Z + \beta_0 = 0$$

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

$$\text{vector } v = \gamma \cdot \frac{\vec{\beta}}{\|\beta\|_2}$$

$$\vec{p} = \vec{x} - \gamma \frac{\vec{\beta}}{\|\beta\|_2}$$

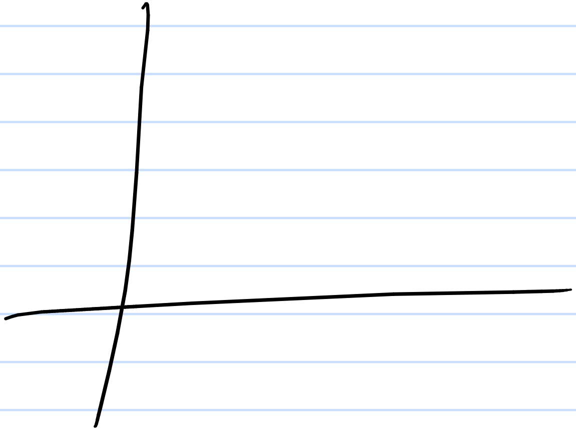


$$\left(\vec{x} - \gamma \frac{\vec{\beta}}{\|\beta\|_2}\right)^T \beta + \beta_0 = 0$$

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

$$\left|\frac{x^T\beta+\beta_0}{\|\beta\|_2}\right| = \gamma$$

$$\boxed{y_i\left(\frac{x^T\beta+\beta_0}{\|\beta\|_2}\right) \geq \gamma}$$



Opt 1 : $\displaystyle\max_{\beta,\beta_0,\gamma} \gamma$

st $\displaystyle y_i\left(\frac{x^T\beta+\beta_0}{\|\beta\|_2}\right) \geq \gamma$

Scale: $\gamma = \frac{1}{\|\beta\|_2}$

Opt 2: $\max_{\beta, \beta_0} \frac{1}{\|\beta\|_2}$

$y_i (x_i^T \beta + \beta_0) \geq 1 \quad i=1,\dots N$

$\|\beta\|_2^2$

$\sqrt{\sum_j \beta_j^2} = \|\beta\|_2$

Opt 3: $\min_{\beta, \beta_0} \|\beta\|_2^2 = \sum_j \beta_j^2$

subject to $y_i(x_i^T\beta + \beta_0) \geq 1 \quad i=1,\dots N$

Extend to non-seperable data:

$$\min_{\beta, \beta_0} \|\beta\|_2^2 + \overset{10^5}{\underset{0}{C}} \cdot \sum_{i=1}^{N} \xi_i$$

$$y_i \underbrace{(x_i^T\beta + \beta_0)}_{f(x)} \geq 1 - \xi_i \quad i=1 \dots N$$

$$\xi_i \geq 0 \quad i=1 \dots N$$

$$\|\beta\|_r := \sqrt[r]{\sum_{j=1}^{p} \beta_j^r}$$

$r=2$

$$\left\| \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix} \right\|_2 = \sqrt{1^2 + 2^2 + 3^2} = \sqrt{14}$$

$$\frac{x_i^T \beta + \beta_0}{\|\beta\|_2}$$

$$\| \quad \|_2 = 6 \qquad \| \quad \|_\infty = 3$$