Math 342 W lecture 3

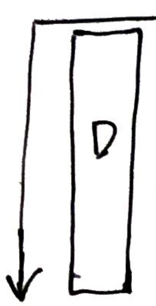
$$y = \underbrace{g(\vec{x})}_{\hat{y} \approx \text{model}} + \underbrace{[(h^*(\vec{x}) - g(\vec{x})] + [(f(\vec{x}) - h^*(\vec{x})] + [t(\vec{z}) - f(\vec{x})]}_{\text{"}\epsilon\text{"} \Leftarrow \text{residual}}$$

$$y = \mathcal{A}(\mathbb{D}, \mathcal{H}) \in \mathcal{H}$$

Lec 2 Recap

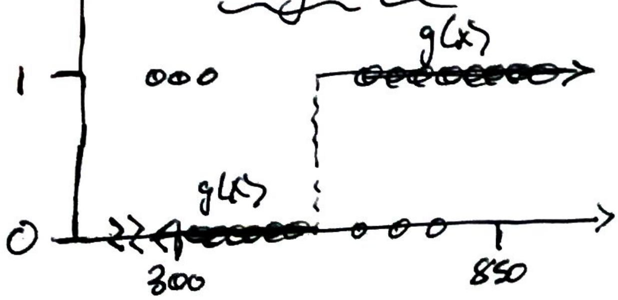
$y = \{0, 1\}^2 \Rightarrow$ Payback Loan

$\mathcal{X} = [\vec{x}_1]$; $\vec{x}_1 \in [300, 850]^2 \Rightarrow$ Credit Scores



B DataFrame $\Rightarrow \left[\mathcal{X} \,\vdots\, \vec{y}\right] = \begin{bmatrix} 810 & 1 \\ 300 & 0 \\ \vdots & \vdots \end{bmatrix} \Rightarrow \mathbb{D}$

Figure 1



$\Rightarrow$ Binary Classification, g returns either a 0 or 1 in the output space

$\Rightarrow \mathcal{H} = \{ \mathbb{1}_{x \geq \theta} : \theta \in \mathcal{X} \}$

--> Indicator Function
1: if cond.
0: if not

Null Model

$g_0 = \mathcal{A}(\vec{y}, \mathcal{H}) = \text{Mode}(\vec{y})$ for classification

↳ no features

Continuing from Figure 1

We are aiming to chose a threshold line, such that we minimize the residuals

$$
\begin{array}{cc|c|c|}
 & & \multicolumn{2}{c}{\hat{y}} \\
 & & 0 & 1 \\ \cline{3-4}
\multirow{2}{*}{y} & 0 & 0 & -1 \\ \cline{3-4}
 & 1 & +1 & 0 \\ \cline{3-4}
\end{array}
\Rightarrow e \Rightarrow \frac{1}{n}\sum_{i=1}^{n} |e_i| = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{y_i \neq \hat{y}_i} \Rightarrow \text{Misclassification Error (ME)}
$$

$\underline{\text{Accuracy}} = 1 - ME$

$\longrightarrow$ Our algorithm

$\mathcal{A}$: Minimize ME

↳ Objective Function

↳ Fitness Function

↳ Target Function

$\mathcal{A}$ returns θ_g s.t ...

$$\theta_g = \underset{\theta \in \mathcal{H}}{\arg\min} \left\{ \sum_{i=1}^{n} \mathbb{1}_{\underbrace{\mathbb{1}_{y_i \geq \theta}}_{\hat{y}_i} \neq y_i} \right\}$$

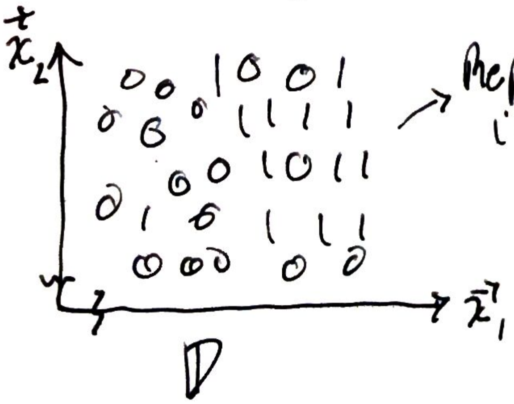
unique values of $\vec{x}_1$

$$\underbrace{\frac{1}{n}\underbrace{\sum_{i=1}^{n} |e_i|}_{\text{SAE: Sum Absolute Error}}}_{\text{MAE: Mean Absolute Error}}$$

MSE: Mean Squared Error $|e_i| \to e_i^2$

NOTE

Now we add a new feature $\vec{x}_2$



$\to$ Representing 3d in 2d

$$\Rightarrow \mathcal{H} = \{ \mathbb{1}_{x_1 \geq \theta_1}, \mathbb{1}_{x_2 \geq \theta_2} : \theta_1, \theta_2 \in \mathcal{X} \}$$