Tuesday, May 5

Last class!

"Quiz" for specific feedback

↳ topics

↳ course components
(readings, activities, reviews,
self-grades, whiteboard lectures)

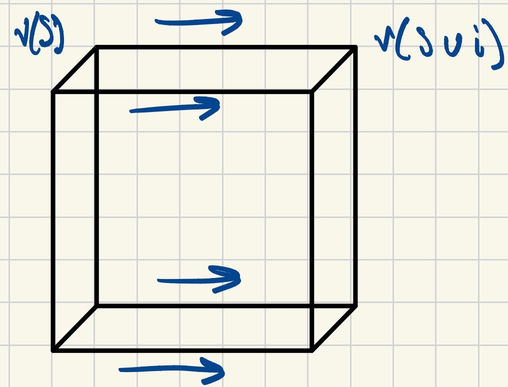
Student experience surveys
for more general feedback

Senior projects / talks due Thursday midnight

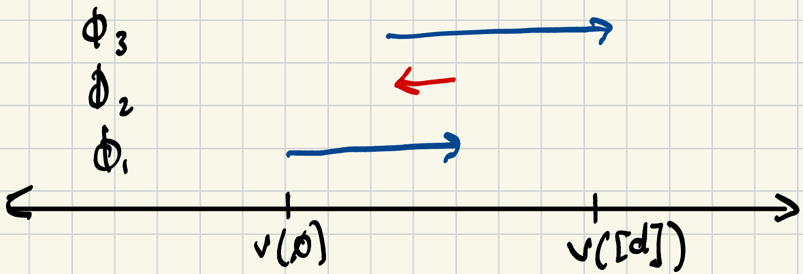
Presentations Thursday, May 14 2-5pm

Shapley Value

$$\phi_i = \sum_{S \subseteq [d] \setminus \{i\}} [v(S \cup i) - v(S)] \mu_S \qquad \text{where} \quad \mu_S = \frac{1}{d \binom{d-1}{|S|}}$$



Efficiency: $\sum_{i=1}^{d} \phi_i = v([d]) - v(\emptyset)$



Monte Carlo Estimator

Sample m sets S wp p_S

$$\tilde{\phi}_i^{MC} = \frac{1}{m} \sum_{j=1}^{m} \left[v(S_j \cup i) - v(S_j) \right] \frac{\mu_S}{p_S} \overset{\text{choose } \mu_S = p_S}{=} \frac{1}{m} \sum_{j=1}^{m} v(S_j \cup i) - v(S_j)$$

$$\mathbb{E}[\tilde{\phi}_i^{MC}] = \phi_i$$

$$\text{Var}(\tilde{\phi}_i^{MC}) = \frac{1}{m} \sum_{S \subseteq [d] \setminus \{i\}} (v(S \cup i) - v(S))^2 \mu_S$$

Idea: Instead of using $v(S \cup i)$, $v(S)$ only for i,
let's reuse for all $\tilde{\phi}_j$

Maximum Sample Reuse

$$\phi_i = \sum_{S \subseteq [d] \setminus i} [v(S \cup i) - v(S)]\, \mu_S$$

rewrite so $v(S)$ appears by itself in each term

$=$

$=$

$\tilde{\phi}_i^{MSR} =$

$Var(\tilde{\phi}_i^{MSR})$ depends on ...

Linear Regression Perspective

$\boxed{v(\emptyset) = 0 \quad \text{WLOG}}$

$$\phi = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_d \end{bmatrix} = \underset{\beta \in \mathbb{R}^d : \langle \beta, 1 \rangle = v([d])}{\arg\min} \sum_{S : 0 < |S| < d} w_S \left(v(S) - \sum_{i \in S} \beta_i \right)^2$$

$\nwarrow$ ensure efficiency

where $w_S = \dfrac{1}{\binom{d}{|S|} |S| (d - |S|)}$

$$A' \in \mathbb{R}^{2^d - 2 \times d}, \quad b' \in \mathbb{R}^{2^d - 2}$$

For $\emptyset \subset S \subset [d]$, $\quad [A']_{S,i} = \mathbb{1}[i \in S] \sqrt{w_S}, \quad [b']_S = v(S) \sqrt{w_S}$

$$\phi = \underset{\beta \in \mathbb{R}^d : \langle \beta, 1 \rangle = v([d])}{\arg\min} \| A'\beta - b' \|_2^2$$

$\swarrow$ weird form...

$$\langle 1, 1\tfrac{v(d)}{d} \rangle = v(d)$$

$$\phi = \operatorname*{argmin}_{\beta \in \mathbb{R}^d : \langle \beta, 1 \rangle = v([d])} \| A'\beta - b' \|_2^2$$

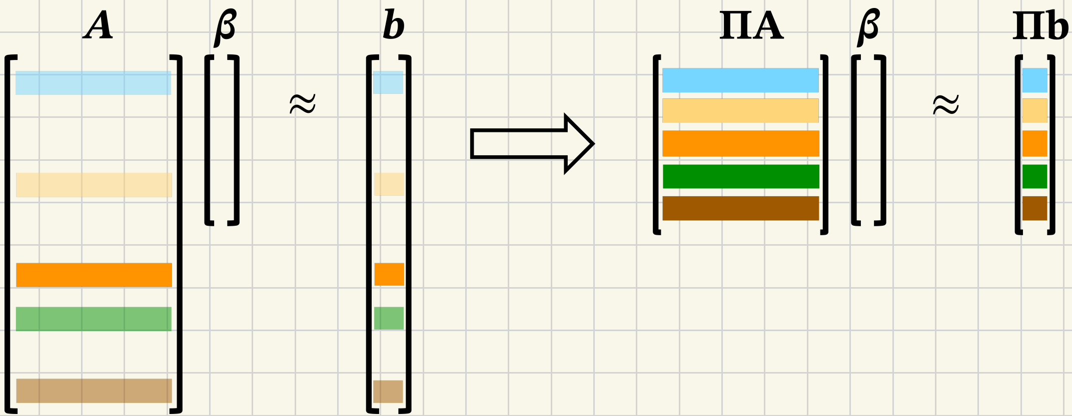
$$= \operatorname*{arg\,min}_{\beta_\perp \in \mathbb{R}^d : \langle \beta_\perp, 1 \rangle = 0} \| A'(\beta_\perp + 1\tfrac{v([d])}{d}) - b' \|_2^2 + 1\frac{v([d])}{d}$$

$$= \operatorname*{argmin}_{\beta \in \mathbb{R}^d} \| \underbrace{A'P}_{A}\beta + \underbrace{A'1\tfrac{v(d)}{d} - b'}_{-b} \|_2^2 + 1\frac{v(d)}{d}$$

where $P = I - \frac{1}{d}11^T$ so
$$\langle 1, P\beta \rangle = 1^T P \beta$$
$$= 1^T(\beta - \tfrac{1}{d}11^T\beta)$$
$$= 1^T\beta - \tfrac{d}{d}1^T\beta = 0$$

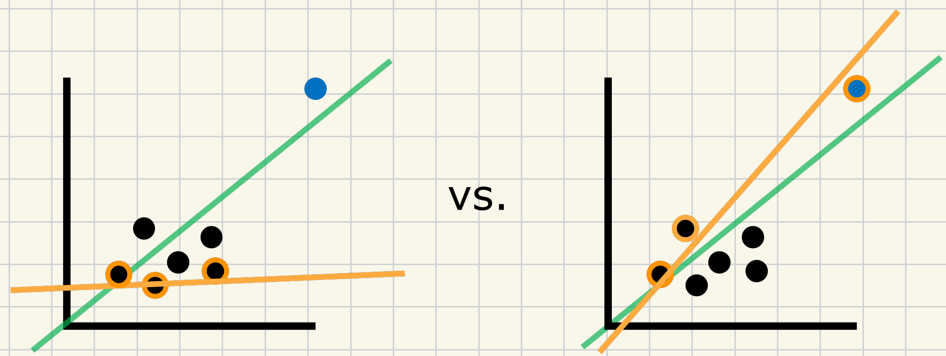
$$= \operatorname*{argmin}_{\beta \in \mathbb{R}^d} \| A\beta - b \|_2^2 + 1\frac{v(d)}{d}$$

Leverage Score Sampling



$$\ell_s = A_s^T (A^T A)^+ A_s$$

$$= \frac{1}{d \binom{d-1}{|S|}}$$



Regression Adjustment

1. Fit surrogate $\tilde{v} \approx v$
2. Return $\phi_i(\tilde{v}) + \tilde{\phi}_i(v - \tilde{v})$

Exercise: $\phi_i(v) = \phi_i(\tilde{v}) + \phi_i(v - \tilde{v})$

Use linear function, polynomial function, trees, Gaussian Processes where surrogate is expressive, and we can efficiently extract its Shapley values