McInerney, James, Benjamin Lacker, Samantha Hansen, Karl Higley, Hugues Bouchard, Alois Gruson, and Rishabh Mehrotra. "Explore, exploit, and explain: personalizing explainable recommendations with bandits." In Proceedings of the 12th ACM Conference on Recommender Systems, pp. 31-39. ACM, 2018.

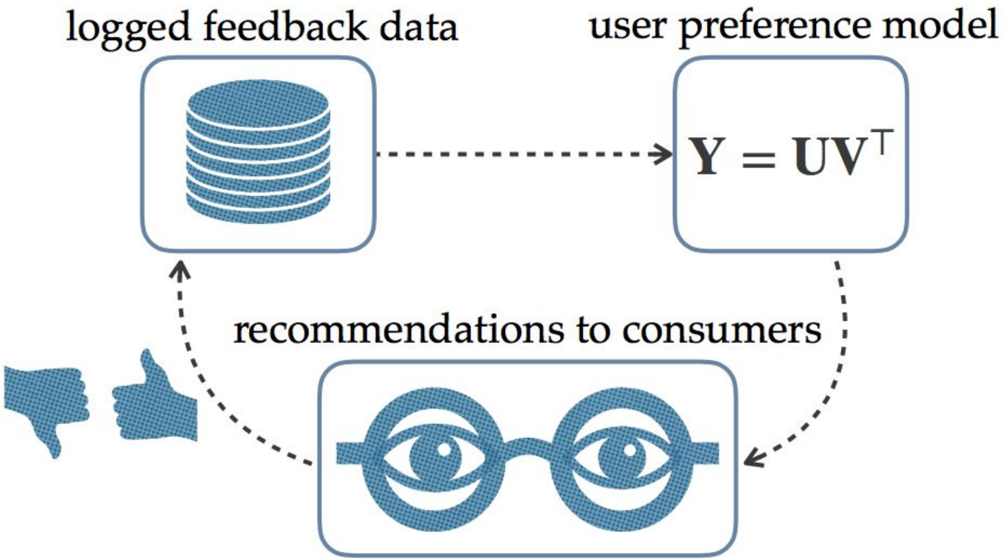
Slides

- Spotify
- Factorization Machine
- Epsilon-greedy
- IPS(inverse propensity score)

Factorization Machine

$$r^{(2)}(j, e, x) = \sigma(\theta_{\text{global}} + \theta^{\top} x' + \sum_{a=1}^{D} \sum_{b>a}^{D} v_a{}^{\top} v_b x'_a x'_b) \qquad (3)$$

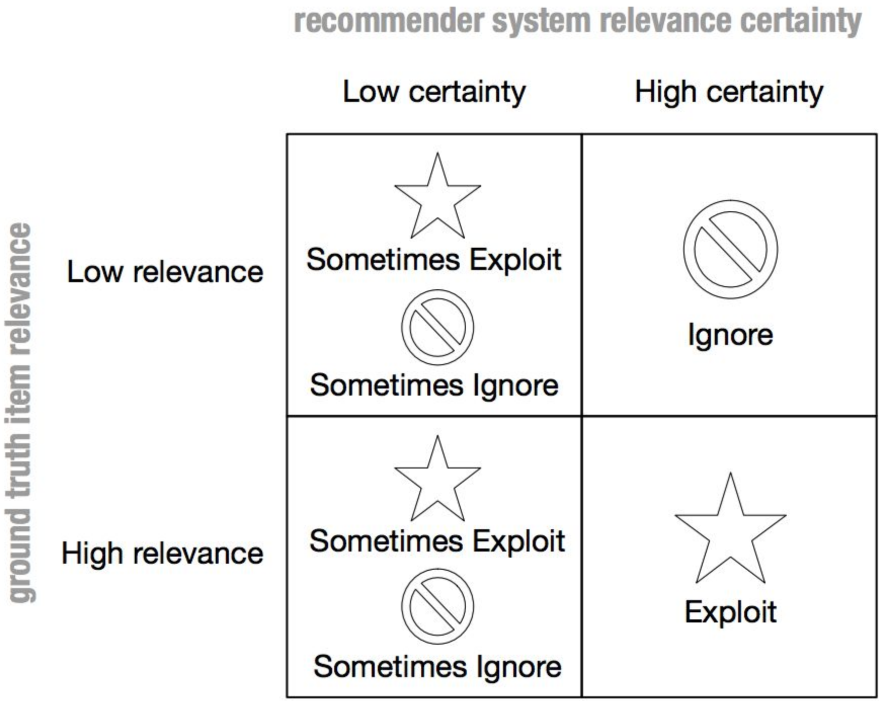
Collaborative filtering perpetuates the Pareto principle



"How Algorithmic Confounding in Recommendation Systems Increases Homogeneity and Decreases Utility" (Chaney et al. 2017)

"Modeling User Exposure in Recommendation" (Liang et al. 2016)

Standard collaborative filtering methods are limited because they can only exploit or ignore



Exploration-Exploitation Policy & Propensity Scoring

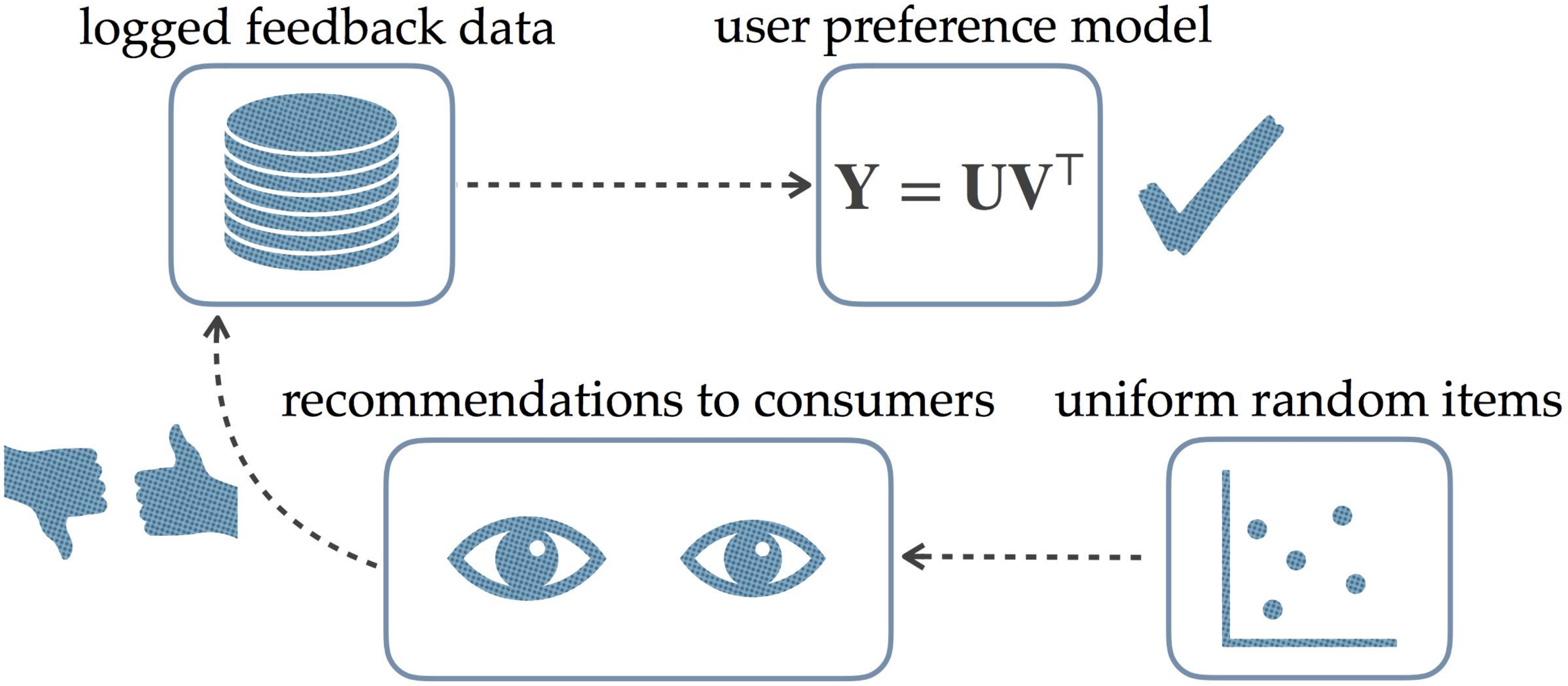
$$\pi_c^{\text{item}}(j \mid x, e) = \begin{cases} (1-\epsilon) + \frac{\epsilon}{|f(e,u)|}, & \text{if } j = j^*, j \in f(e,x) \\ \frac{\epsilon}{|f(e,u)|}, & \text{if } j \neq j^*, j \in f(e,x) \\ 0, & \text{otherwise.} \end{cases}$$

$$\text{where } j^* = \arg_{j_1} \max r(j_1, e, x) \qquad (7)$$

$$\pi_c^{\text{expl.}}(e \mid x, j) = \begin{cases} (1-\epsilon) + \frac{\epsilon}{|\mathcal{E}|}, & \text{if } e = e^*, j \in f(e,x) \\ \frac{\epsilon}{|\mathcal{E}|}, & \text{if } e \neq e^*, j \in f(e,x) \\ 0, & \text{otherwise.} \end{cases}$$

$$\text{where } e^* = \arg_{e_1} \max r(j, e_1, x). \qquad (8)$$

Let’s restart from the basic ideal of randomized controlled trials



Off-Policy Training

$$\hat{\theta}, \hat{v} = \arg_{\theta, v} \max \mathbb{E}_{A \sim \text{Uniform}(\cdot)} [\mathbb{E}_{X,R} [\log p_{\theta,v}(R|A, X)]] \quad (5)$$

$$\approx \arg_{\theta, v} \max \frac{1}{N} \sum_{n=1}^{N} \frac{\text{Uniform}(a_n)}{\pi_c(a_n)} \log p_{\theta,v}(r_n|a_n, x_n). \quad (6)$$

IPS(inverse propensity score)

Off-Policy Training

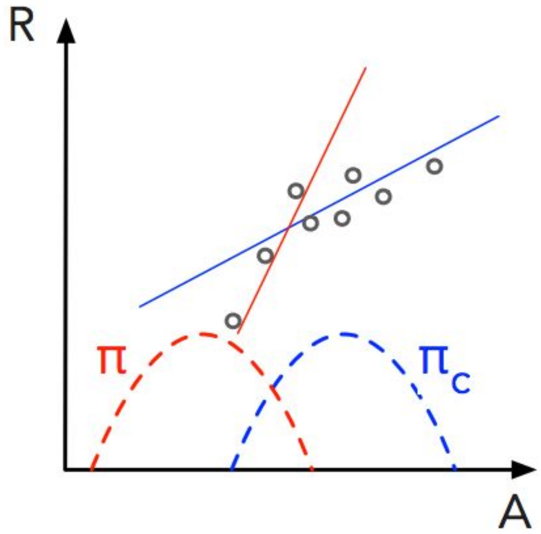
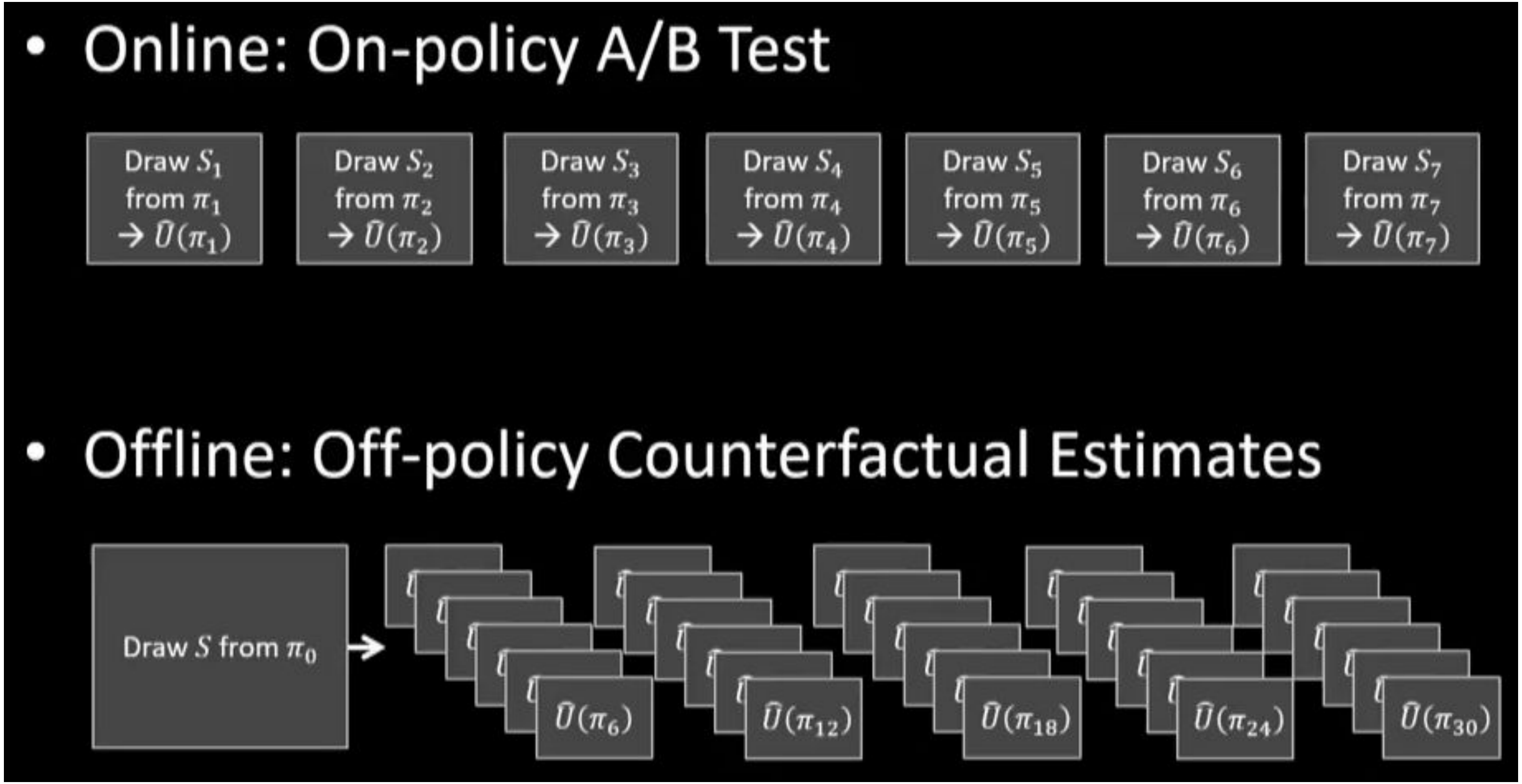


Figure 3: Off-policy training fits a reward function that best fits the input points with respect to the target policy π using input points generated by the collection policy π_c.

Causal Inference Recommendation Papers

- Schnabel, Tobias, Adith Swaminathan, Ashudeep Singh, Navin Chandak, and Thorsten Joachims. "Recommendations as treatments: Debiasing learning and evaluation." arXiv preprint arXiv:1602.05352 (2016).
    - Propensity-Scored Matrix Factorization (= IPS + MF)
- Liang, Dawen, Laurent Charlin, and David M. Blei. "Causal Inference for Recommendation." (2016).
    - IPS + MF
- Bonner, Stephen, and Flavian Vasile. "Causal embeddings for recommendation." In Proceedings of the 12th ACM Conference on Recommender Systems, pp. 104-112. ACM, 2018.
    - Criteo, MF + Counterfactural Risk Minimization(CRM)
- Gilotte, Alexandre, Clément Calauzènes, Thomas Nedelec, Alexandre Abraham, and Simon Dollé. "Offline A/B testing for Recommender Systems." In Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, pp. 198-206. ACM, 2018.
    - Criteo, IPS
- Joachims, Thorsten, Adith Swaminathan, and Maarten de Rijke. "Deep learning with logged bandit feedback." (2018).
    - Unbiased estimate of risk => IPS
    - Partian Information => Variacne Control
    - Propensity Overfitting => SNIPS(self-normalized IPS estimator)

Evaluating Online Metrics Offline

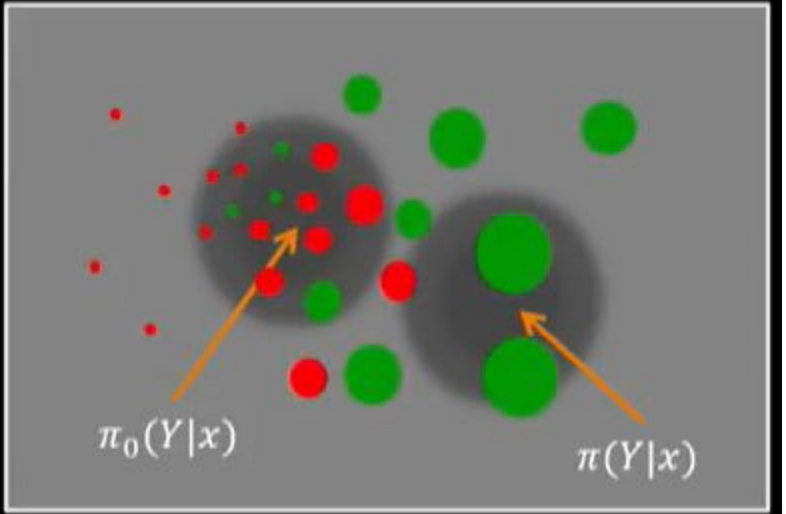


Off-Policy Risk Evaluation

Given $S = \big((x_1, y_1, \delta_1), \ldots, (x_n, y_n, \delta_n)\big)$ collected under π_0,

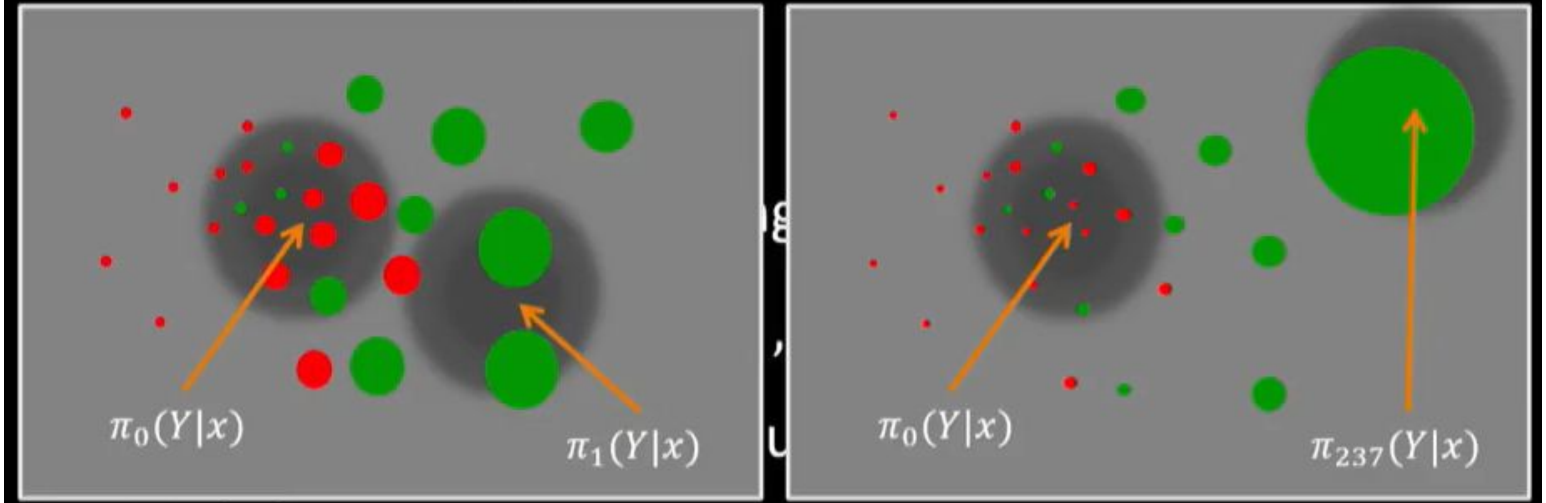
$$\hat{R}(\pi) = \frac{1}{n}\sum_{i=1}^{n} \delta_i \frac{\pi(y_i|x_i)}{\pi_0(y_i|x_i)}$$

Propensity p_i



→ Unbiased estimate of risk, if propensity nonzero everywhere (where it matters).

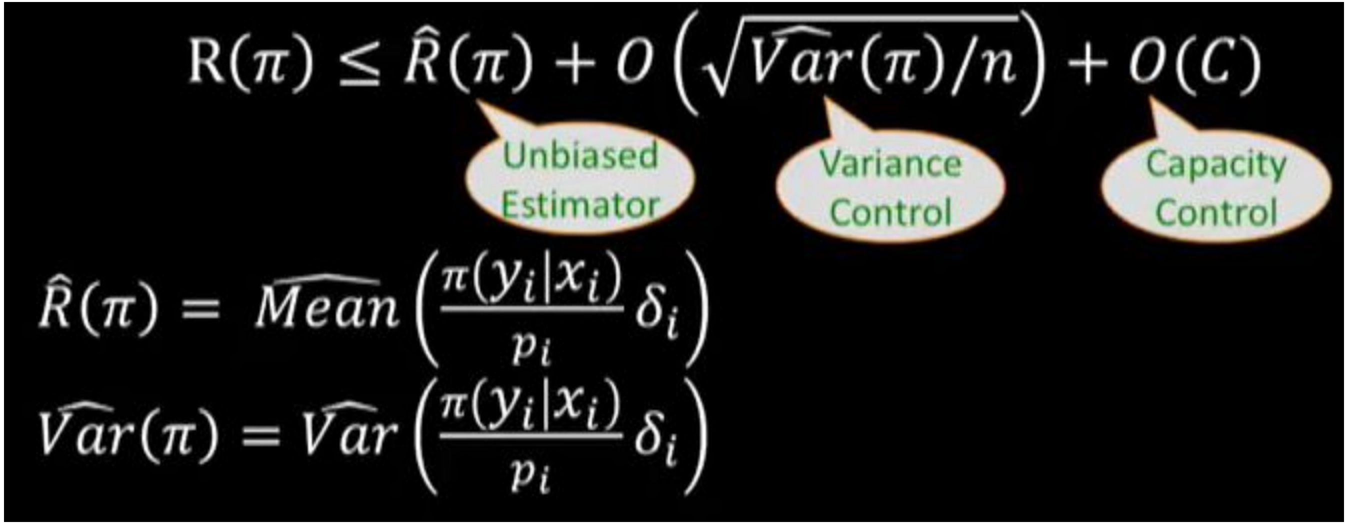
Partial Information Empirical Risk Minimization



- Training

$$\hat{\pi} := \operatorname{argmin}_{\pi \in H} \sum_{i}^{n} \frac{\pi(y_i|x_i)}{p_i} \delta_i$$

Variance Control



Problem: Propensity Overfitting

- Example
  - Training sample with losses:

X

0	1	1	1	1	1
1	1	1	0	1	1
1	0	1	1	1	1
1	1	1	1	1	0
1	1	0	1	1	1
1	1	1	1	1	0
1	1	1	0	1	1

Y

  - Which $\pi(y|x)$ minimize IPS?

$$R(\pi) = \min_{\pi \in H} \frac{1}{n} \sum_{i}^{n} \frac{\pi(y_i|x_i)}{p_i} \delta_i$$

EQUIVARIANT COUNTERFACTUAL RISK MINIMIZATION

$$\hat{R}_{SNIPS}(\pi_w) = \frac{\frac{1}{n}\sum_{i=1}^{n} \delta_i \frac{\pi_w(y_i|x_i)}{\pi_0(y_i|x_i)}}{\frac{1}{n}\sum_{i=1}^{n} \frac{\pi_w(y_i|x_i)}{\pi_0(y_i|x_i)}}.$$

TRAINING ALGORITHM

$$w = \underset{w \in \Re^N}{\operatorname{argmin}} \left[\hat{R}^{SNIPS}(w) + \lambda_1 \sqrt{\widehat{Var}\left(\hat{R}^{SNIPS}(w)\right)} + \lambda_2 \left|\left|w\right|\right|^2 \right]$$

More about causal inference

- Ryan Tibshirani's lecture notes on causal inference
  - http://www.stat.cmu.edu/~larry/=sml/Causation.pdf
- Hernán, Miguel A. and James M. Robins. 2012. Causal Inference. Forthcoming, Cambridge University Press.
  - https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/