Learn conditional $P(Y_h \mid X = x)$

Build graph $\mathcal{G} = (\mathcal{H}, \mathcal{E})$:

$$\{h, h'\} \in \mathcal{E}$$

iff counterfactual stability is not violated for pair given the data and learned marginals

Ψ is greedy approx. of clique cover of $\mathcal{G}$ minimizing

$$\sum_{h,h' \in \mathcal{H}} \underbrace{\mathcal{L}(\mathcal{M}(\Psi), h', h)}_{\text{average empirical loss of inferring } Y_{h'} \text{ given } Y_h} - \underbrace{\mathcal{L}(\mathcal{M}(\mathcal{H}), h', h)}_{\text{model with no similar experts}}$$