## 1 Causal Transfer Trees

The goal is to create an estimator  $(\hat{\tau}(x))$  of the true individual treatment effect  $(\tau_i)$  that minimizes some loss function in new prediction context, given labelled data in a different prediction context. We will consider the squared loss function:

$$\ell(\hat{\tau}, x_i, \tau_i) = (\tau_i - \hat{\tau}(x_i))^2$$

Which will be equivalent to the formulation of causal trees. We will similarly create discrete partitions in the feature space such that:

$$\mathcal{X} = \bigcup_{i=1}^{P} \Pi_i$$

And define the following estimator:

$$\hat{\tau}(x,\Pi,S(\mathcal{D})) = \bar{y}_i^1 - \bar{y}_i^0 \quad ; x \in \Pi_i \; ; \; i : x \in \Pi_i$$

Where  $S(\cdot)$  consists of a sampling function to sample matched data (y, x) from a domain  $\mathcal{D}$ .  $\Pi$  consists of some particular partitioning of the feature space. As the sample mean is an unbiased estimate of the expectation, this is an unbiased estimate of  $\tau(x, \Pi, \mathcal{D}) := \mathbb{E}[\tau | X \in \Pi_i]$ . We would like to minimize our expected loss:

$$\mathbb{E}\ell(\Pi) = \mathbb{E}_{X,S}[(\tau_i - \hat{\tau}(x_i, \Pi, S(\mathcal{D}_S))^2]$$

Where our expectation is taken over all potentiall samples data in our source domain,  $S(\mathcal{D}_S)$  and all the potential target data points, X. We can expand this loss function:

$$\mathbb{E}\ell(\Pi) = \underbrace{\mathbb{E}_X[\tau_i^2]}_{\text{doesn't depend on }\Pi} -2\mathbb{E}_{X,S}[\tau_i\hat{\tau}(x_i,\Pi,S(\mathcal{D}_S))] + \mathbb{E}_X[\mathbb{E}_S[\hat{\tau}(x_i,\Pi,S(\mathcal{D}_S))^2]]$$

We can expand the middle term with the law of iterated expectations and note that, conditional on the value being within a leaf,  $\tau_i$  and  $\hat{\tau}(x_i, \Pi, S(\mathcal{D}_S))$  become independent:

$$-2\mathbb{E}_{X,S}[\underbrace{\mathbb{E}_{X}[\tau_{i}|x_{i}\in\Pi_{i}]}_{=\tau(x_{i},\Pi,\mathcal{D}_{T})}\underbrace{\mathbb{E}_{S}[\hat{\tau}(x_{i},\Pi,S(\mathcal{D}_{S}))|x_{i}\in\Pi_{i}]}_{=\tau(x_{i},\Pi,\mathcal{D}_{S})}] + \mathbb{E}_{X}[\mathbb{E}_{S}[\hat{\tau}(x_{i},\Pi,S(\mathcal{D}_{S}))^{2}]]$$

$$-2\mathbb{E}_{X}[\tau(x_{i},\Pi,\mathcal{D}_{T})\tau(x_{i},\Pi,\mathcal{D}_{S})] + \mathbb{E}_{X}[\mathbb{E}_{S}[\hat{\tau}(x_{i},\Pi,S(\mathcal{D}_{S}))^{2}]]$$

$$-2\mathbb{E}_{X}[\underbrace{(\tau(x_{i},\Pi,\mathcal{D}_{T})-\tau(x_{i},\Pi,\mathcal{D}_{S})}_{=\delta_{\tau}} + \tau(x_{i},\Pi,\mathcal{D}_{S}))\tau(x_{i},\Pi,\mathcal{D}_{S})] + \mathbb{E}_{X}[\mathbb{E}_{S}[\hat{\tau}(x_{i},\Pi,S(\mathcal{D}_{S}))^{2}]]$$

$$-2\mathbb{E}_{X}[\delta_{\tau}\tau(x_{i},\Pi,\mathcal{D}_{S}) + \tau(x_{i},\Pi,\mathcal{D}_{S})^{2}] + \mathbb{E}_{X}[\mathbb{E}_{S}[\hat{\tau}(x_{i},\Pi,S(\mathcal{D}_{S}))^{2}]]$$

$$-2\mathbb{E}_{X}[\mathbb{E}_{S_{T}}[\tau(x_{i},\Pi,\mathcal{D}_{S_{T}}) - \tau(x_{i},\Pi,\mathcal{D}_{S})]\tau(x_{i},\Pi,\mathcal{D}_{S}) + \tau(x_{i},\Pi,\mathcal{D}_{S})^{2}] + \mathbb{E}_{X}[\mathbb{E}_{S}[\hat{\tau}(x_{i},\Pi,S(\mathcal{D}_{S}))^{2}]]$$

$$-2\mathbb{E}_{X}[\mathbb{E}_{S_{T}}[\tau(x_{i},\Pi,\mathcal{D}_{S_{T}})\tau(x_{i},\Pi,\mathcal{D}_{S})] + \mathbb{E}_{X}[\mathbb{E}_{S}[\hat{\tau}(x_{i},\Pi,S(\mathcal{D}_{S}))^{2}]]$$