1 Proposal scratch notes

- $\mathcal{D}$, $\mathcal{I}$: (unknown) disjoint sets of indices that describe the groups of transformations that are relevant and irrelevant to the output, respectively. $\mathcal{D} \cup \mathcal{I} = \{1, \ldots, m\}$.
- U_Y, $U_{\mathcal{I}}$, $U_{\mathcal{D}}$, $\widetilde{U}_{\mathcal{I}}$: independent latent variables that influence the value of the variable(s) that they point to.
- $X^{(\text{hid})}$: some unknown canonical form of the observed input X. It is assumed that given $U_{\mathcal{D}}$ and $U_{\mathcal{I}}$, X was obtained from an ordered sequence of transformations on the canonical form, i.e.,

$$X = T_{U_{\mathcal{D}}, U_{\mathcal{I}}} \circ X^{(\text{hid})}$$

where transformations

$$T_{U_{\mathcal{D}}, U_{\mathcal{I}}} = T_{\mathcal{I}}^{(1)} \circ T_{\mathcal{D}}^{(1)} \circ T_{\mathcal{I}}^{(2)} \circ \ldots$$

make up the overgroup $\mathcal{G}_{\mathcal{D} \cup \mathcal{I}}$, $T_{\mathcal{D}}^{(j)}$ is a transformation in group $\mathcal{G}_j$ from the overgroup $\mathcal{G}_{\mathcal{D}} = \langle \cup_{j \in \mathcal{D}} \mathcal{G}_j \rangle$, and $T_{\mathcal{I}}^{(i)} \in \mathcal{G}_i \subset \mathcal{G}_{\mathcal{I}} = \langle \cup_{i \in \mathcal{I}} \mathcal{G}_i \rangle$. Note that $\mathcal{G}_{\mathcal{I}}$ is also assumed to be a normal subgroup of $\mathcal{G}_{\mathcal{D} \cup \mathcal{I}}$.

- Y: observed output assumed to be generated by

$$Y = h(X^{(\text{hid})}, U_{\mathcal{D}}, U_Y)$$

where h is a deterministic function.

- $X_{U_{\mathcal{I}} \leftarrow \widetilde{U}_{\mathcal{I}}}$: counterfactual variable to X where $U_{\mathcal{I}}$ has been replaced by $\widetilde{U}_{\mathcal{I}}$, i.e.,

$$X_{U_{\mathcal{I}} \leftarrow \widetilde{U}_{\mathcal{I}}} = T_{U_{\mathcal{D}}, \widetilde{U}_{\mathcal{I}}} \circ X^{(\text{hid})} .$$

- Want CG-invariant representation

$$\Gamma(X) = \Gamma(T_{U_{\mathcal{D}}, U_{\mathcal{I}}} \circ X^{(\text{hid})}) = \Gamma(T_{U_{\mathcal{D}}, \widetilde{U}_{\mathcal{I}}} \circ X^{(\text{hid})}) = \Gamma(X_{U_{\mathcal{I}} \leftarrow \widetilde{U}_{\mathcal{I}}})$$

When $\mathcal{G}_{\mathcal{I}}$ is a normal subgroup of $\mathcal{G}_{\mathcal{D} \cup \mathcal{I}}$, G-invariant representation

$$\Gamma(X) = \Gamma(T_{\mathcal{I}} \circ X)$$

for all $T_{\mathcal{I}} \in \mathcal{G}_{\mathcal{I}}$ is sufficient.

- Because groups are finite linear automorphisms, each transformation T is just a linear function and so Reynolds operator can be applied directly to the group actions

$$\bar{T} = \frac{1}{|\mathcal{G}|} \sum_{T \in \mathcal{G}} T$$

For continuous linear groups, orbit-average over a Haar measure λ

$$\bar{T} = \int_{\mathcal{G}} T \lambda(T)$$

TODO: need to estimate the operator. Assume uniform Haar and sample?

2 Kernel hypothesis test notes

- Let X be a r.v. on domain $\mathcal{X}$. A kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ induces a RHKS $\mathcal{H}$ of functions $f : \mathcal{X} \to \mathbb{R}$ where for $f \in \mathcal{H}$ and $x \in \mathcal{X}$,
$$\langle f, k(x, \cdot)\rangle = f(x)$$
(reproducing property). $k(x, \cdot) = \phi_x$ can also be considered an implicit feature map ($\phi_x : \mathcal{X} \to \mathbb{R}$) where
$$\langle \phi_x, \phi_{x'}\rangle = k(x, x')$$
is a measure of similarity.

- Kernel methods work on inner products of feature maps of observations in the RKHS associated with kernel. Inner products may be computed without explicitly computing the high-dimensional feature map ("kernel trick").

- Main challenge of designing kernel-based hypothesis tests is deriving large-sample distribution of test statistic under null.

- Gram matrix should be positive semidefinite. Satisfised if kernel is symmetric and positive semidefinite.

- Let X be a r.v. with distribution $\mathbb{P}$. Mean element
$$\mu_{\mathbb{P}} = \mathbb{E}_{\mathbb{P}}[\phi_X]$$
associated with X is unique element of RKHS $\mathcal{H}$ s.t. for all $f \in \mathcal{H}$,
$$\langle \mu_{\mathbb{P}}, f\rangle = \mathbb{E}_{\mathbb{P}}[f(X)]$$
Covariance operator $\Sigma_{\mathbb{P}} : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ associated with X is unique operator s.t. for all $f, g \in \mathcal{H}$,
$$\langle f, \Sigma_{\mathbb{P}} g\rangle = \mathrm{Cov}(f(X), g(X)) = \mathbb{E}_{\mathbb{P}}[f(X), g(X)] - \langle \mu_{\mathbb{P}}, f\rangle\langle \mu_{\mathbb{P}}, g\rangle$$
Empirical estimates of inner products that lead to estimates of element/operator are available.

- Kernel is characteristic if mean embedding $\mu : \mathbb{P} \to \mathcal{H}$ is injective. Each distribution can be uniquely represented in the RKHS and all statistica features of distributions are preserved (TODO) by a characteristic kernel.

- If $\dim(\mathcal{H}) = \infty$, $\mu_{\mathbb{P}}$ has more significance than in classical statistics.

- Detecting (conditional?) invariances in single training environment via hypothesis testing. Explicitly testing for invariances? Invariant testing