Formulating Detection and Breaking of Parameter Symmetries

Kenny Chiu

November 26, 2019

1 Background

[TODO context]

We represent probabilistic models using factor graphs (θ, F) of variables $\theta = (\theta_1, ..., \theta_N)$ and factors $F = (F_1, ..., F_k)$. Let Θ denote the space of variable values. $\theta \in \Theta$ contains all parameters, latent variables, and observations (which are fixed). F represents all functions, operations, constraints, priors and likelihoods that the posterior distribution may factorize into. Given data, the unnormalized posterior distribution can then be expressed as

$$\prod_k F_k(\theta)$$

where factor F_k may not necessarily depend on all of θ. In the context of symmetries, priors are not of interest as it is assumed that the fixed parameters of the prior are chosen by the user. The remainder of this report will refer to non-prior factors when considering F_k.

A symmetry $\sigma : \Theta \to \Theta$ is a measurable function with a measurable inverse that satisfies

$$\prod_k F_k(\theta) \propto \prod_k F_k(\sigma(\theta))$$

and where if θ_i is some observed variable, then the symmetry keeps θ_i fixed. In other words, a symmetry is a transformation on the variables that preserves the product likelihood up to a scaling constant.

A local symmetry is a symmetry σ that satisfies

$$F_k(\theta) \propto F_k\left(\sigma(\theta)\right)$$

for all non-prior factors F_k. In contrast to a symmetry, a local symmetry is a transformation on the variables that preserves the likelihood at each factor up to a scaling constant.

1.1 Equivalence class

Let A be a set. For all $a, b, c \in A$, an equivalence relation Ξ on A is a binary relation that satisfies the following three properties:

1. **reflexivity**: $a \; \Xi \; a$
2. **symmetry**: if $a \; \Xi \; b$ then $b \; \Xi \; a$
3. **transitivity**: if $a \; \Xi \; b$ and $b \; \Xi \; c$ then $a \; \Xi \; c$

An equivalence relation Ξ partitions A into sets called equivalence classes. For $a, b \in A$, if $a \; \Xi \; b$ then a and b belong to the same equivalence class. We denote the equivalence class of a as $[a]$. The set of all equivalence classes in A with respect to Ξ is called the quotient set of A by Ξ, denoted A/Ξ.

A section is a function $f : A/\Xi \to A$ that maps an equivalence class to one of its members. The member $f([a])$ is called the representative of $[a]$ with respect to f.

2 Formulation based on equivalence classes

Let (θ, F) be a factor graph. Let Ξ be an equivalence relation such that for $\theta, \theta_* \in \Theta$, we have $\theta \,\Xi\, \theta_*$ if and only if there exists a local symmetry σ such that $\sigma(\theta) = \theta_*$.

We show that Ξ is a proper equivalence relation. Consider $\theta_1, \theta_2, \theta_3 \in \Theta$.

1. **reflexivity**: let σ be the identity map. Then trivially, $F_k(\theta_1) = F_k(\sigma(\theta_1))$.

2. **symmetry**: suppose $\theta_1 \,\Xi\, \theta_2$. Then there exists some symmetry σ such that $\sigma(\theta_1) = \theta_2$ and $F_k(\theta_1) \propto F_k(\theta_2)$. By definition of symmety, σ^{-1} exists and $\sigma^{-1}(\theta_2) = \theta_1$. Then

$$F_k(\theta_2) \propto F_k(\theta_1) = F_k(\sigma^{-1}(\theta_2))$$

   and hence $\theta_2 \,\Xi\, \theta_1$.

3. **transitivity**: suppose $\theta_1 \,\Xi\, \theta_2$ and $\theta_2 \,\Xi\, \theta_3$. Then there exists symmetries σ_1, σ_2 such that

$$\begin{aligned} \sigma_1(\theta_1) &= \theta_2 & \qquad F_k(\theta_1) &\propto F_k(\theta_2) \\ \sigma_2(\theta_2) &= \theta_3 & \qquad F_k(\theta_2) &\propto F_k(\theta_3) \end{aligned}$$

   Then $\sigma_2(\sigma_1(\theta_1)) = \theta_3$ and

$$F_k(\theta_1) \propto F_k(\theta_2) \propto F_k(\theta_3) = F_k(\sigma_2(\sigma_1(\theta_1)))$$

   and hence $\theta_1 \,\Xi\, \theta_3$.

Notice that Ξ is dependent on the factor graph (θ, F). It is mathematically more convenient to consider factor graphs equipped with a Ξ that corresponds to a specific type of symmetry. We denote this as $\mathcal{F} = (\theta, F, \Xi)$.

We now have all the definitions needed to formulate symmetry detection and symmetry breaking in terms of equivalence classes.

- A symmetry detector $\Delta_{\mathcal{F}} : \Theta \to \Theta/\Xi$ is a measurable function that maps θ to $[\theta]$ with respect to Ξ.
- A symmetry breaker $\phi_{\mathcal{F}} : \Theta/\Xi \to \Theta$ is a section that maps $[\theta]$ to a representative $\theta_* \in [\theta]$.
- An automatic symmetry breaker $\phi_{\mathcal{F}} \circ \Delta_{\mathcal{F}} = \Phi_{\mathcal{F}} : \Theta \to \Theta$ is the composition of a symmetry breaker and its corresponding symmetry detector. It maps θ to a representative θ_* of its equivalence class.

The dependence of Δ, ϕ, and Φ on a factor graph $\mathcal{F}$ is made clear by its subscript. When considering two functions of the same type, e.g. $\Phi_{\mathcal{F}_1}$ and $\Phi_{\mathcal{F}_2}$, we will view them as being defined on the same factor graph (θ, F) but with Ξ_1, and Ξ_2 corresponding to different types of symmetries.

This formulation is most useful in the case where the main concern is the nonidentifiability of the parameters in the presence of symmetries. If inference has been made on a posterior distribution but the question remains whether the inferred value is the "correct" one, it may instead be more interpretable to work with the symmetry broken posterior distribution

$$\prod_k F_k(\Phi_{\mathcal{F}_M} \circ ... \circ \Phi_{\mathcal{F}_1}(\theta))$$

where $\Phi_{\mathcal{F}_1}, ..., \Phi_{\mathcal{F}_M}$ correspond to M different types of symmetries that are to be broken. The corresponding $\phi_{\mathcal{F}_1}, ..., \phi_{\mathcal{F}_M}$ are chosen to select representatives from the respective equivalence classes based on some desired criteria.

We show an example to demonstrate this idea.

2.1 Scaling symmetries

A scaling symmetry [NMT13]

3 Open questions and research directions

A Exercises

References

[NMT13] R. Nishihara, T. Minka, and D. Tarlow. *Detecting Parameter Symmetries in Probabilistic Models.* 2013. arXiv: 1312.5386 [stat.ML].