

Size norm model ideas

Caleb Weinreb

May 2023

1 Modeling framework

Suppose we have pose data $\{y_t\}_{t=1}^T$ and $\{y'_t\}_{t=1}^{T'}$ for a pair of animals, where $y_t, y'_t \in \mathbb{R}^{KD}$, represent the positions of K keypoints in D dimensions. For now we will assume that the keypoints are in egocentric coordinates, meaning each animal is always centered and pointing in the same direction. Let's also assume the data have been rescaled to account for gross size differences, so that all remaining differences are subtleties of body shape. Our goal is to define a canonical (low-dimensional) pose space and learn how the data from each animal map onto it. Put in terms of generative modeling, we wish to explain each animal's observed pose y_t as the (noisy) realization of some latent pose state $z_t \in \mathbb{R}^M$, where the space of latent states is shared between animals. This can be formalized as follows:

$$y_t = F(z_t) + \xi \text{ where } \xi \sim \mathcal{N}(0, \Sigma) \text{ and } z_t \sim P_z \quad (1)$$

$$y'_t = F'(z'_t) + \xi' \text{ where } \xi' \sim \mathcal{N}(0, \Sigma') \text{ and } z'_t \sim P'_z \quad (2)$$

where F, F' are respective functions mapping from the latent space to each animals pose space, and P_z, P'_z are distributions over latent states. Ideally, we want F and F' to capture morphological differences between the two animals, and P_z, P'_z to capture differences in the frequency of behaviors. We also want to make sure that F and F' are as similar as possible, i.e. to avoid arbitrary rotations of the latent space for one animal compared to the other. Here are some ideas for how to achieve these goals:

- **Shared Gaussian pose distribution:** We could start with a simple shared distribution over the latent space $P_z = P'_z = \mathcal{N}(0, I_M)$.
- **Gaussian mixture pose distribution:** Another option is to model P_z and P'_z as Gaussian mixtures, where the mixture components are shared between animals but the mixture weights are allowed to vary. This would naturally capture differences in the frequency of behaviors between animals.
- **Affine pose mappings:** We probably want to assume that F and F' are affine, i.e. that $F(z) = Az + b$ for some $A \in \mathbb{R}^{KD \times M}$ and $b \in \mathbb{R}^{KD}$.
- **Ensuring similar mappings:** There are a few ways to make sure that F and F' are similar. One is to model their parameters as additive perturbations of a common mapping, e.g. $F = (A + \Delta A, b + \Delta b)$, $F' = (A + \Delta A', b + \Delta b')$ where A, b can vary broadly but $\Delta A, \Delta b, \Delta A', \Delta b'$ have a tight prior. Another option is to perturb the common mapping multiplicatively.

2 Optimizing the mixture of linear Gaussians

We assume the following generative model:

$$z_t^i \sim \text{Cat}(\pi^i) \quad \pi^i \sim \text{Dir}(\alpha) \quad (3)$$

$$x_t^i \sim \mathcal{N}(m_{z_t^i}, Q_{z_t^i}) \quad m_n, Q_n \sim \text{Normal-Inverse-Wishart} \quad (4)$$

$$y_t^i \sim \mathcal{N}(F^i x_t^i + d^i, I_{KD}) \quad d^i, F^i \sim \text{Matrix-Normal} \quad (5)$$

where i indexes animals, t indexes frames, $x_t^i, y_t^i \in \mathbb{R}^{KD}$, $z_t^i \in \{1, \dots, N\}$, and $\sigma^2 \in \mathbb{R}^+$ is fixed. The parameters can be optimized using variation mean field EM, in which the posterior over latent variables is approximated by the product of factors $q(z_t^i), q(x_t^i)$. During the M-step, we find the parameters $\theta = (m, Q, d, F, \pi)$ that maximize $\mathbb{E}_{q(x, z)} \log P(y, x, z | \theta)$, where $q(x, z) = \prod_{t, i} q(z_t^i) q(x_t^i)$. During the E-step, we iteratively updating the factors $q(z_t^i), q(x_t^i)$ using coordinate ascent, with one or more passes through the data per iteration. The updates are given by:

$$\log q^*(z_t^i) = \mathbb{E}_{q(x_t^i)} \log P(y_t^i, x_t^i, z_t^i | \theta) + \text{const.} \quad (6)$$

$$= \int_{x_t^i} \mathcal{N}(x_t^i | \mu, \Sigma) \log \mathcal{N}(x_t^i | m_{z_t^i}, Q_{z_t^i}) + \log \pi_{z_t^i} + \text{const.} \quad (7)$$

$$= \mathcal{N}(\mu | m_{z_t^i}, Q_{z_t^i}) - \text{tr}[Q_{z_t^i}^{-1} \Sigma] + \log \pi_{z_t^i} + \text{const.} \quad (8)$$

$(x_t^i - m_{z_t^i=k})^T Q_{z_t^i=k}^{-1} (x_t^i - m_{z_t^i=k})$ under $q(x_t^i)$ is $(\mu_{x_t^i} - m_{z_t^i=k})^T Q_{z_t^i=k}^{-1} (\mu_{x_t^i} - m_{z_t^i=k}) + \text{Tr}[Q_{z_t^i=k}^{-1} \Sigma_{x_t^i}]$, where $\text{Tr}[\cdot]$