Option GAN

Current graphical model

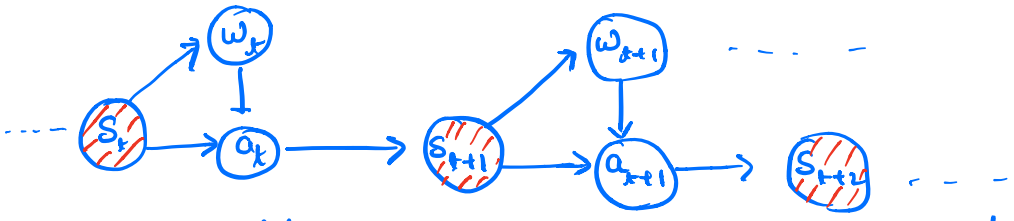

Fig 1. Nodes annotated in Red are observed, others are not.

Using the above model we can write,

$$\pi_\theta(a_t|s_t) = \sum_{\omega_t} p_\theta(\omega_t|s_t)\, p_\phi(a_t|s_t,\omega_t) \qquad \text{— (1)}$$

The above model is a simple mixture of experts model where the ω_t can be considered as the learned gating mechanism i.e. which expert to choose.

We will now show that if we extend the above graphical model by adding directed connections between the latent variable, it can no longer be modelled as a simple mixture of experts model.

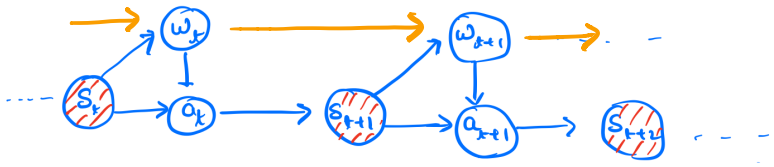

Fig 2. The same graphical model as above but with connections between latent variables as well. → represents new directed connections.

Graphical model in Fig 2 can be written as,

$$\sum_{\omega_t} p_\theta(a_t|s_t,\omega_t)\, p_\phi(\omega_t|s_t,\omega_{t-1})$$

$$= \sum_{\omega_t} p_\theta(a_t|s_t,\omega_t,\omega_{t-1})\, p_\phi(\omega_t|s_t,\omega_{t-1})$$

$$(\because a_t \perp\!\!\!\perp \omega_{t-1} \mid \omega_t)$$

$$= \sum_{\omega_t} p(a_t, \omega_t \mid s_t, \omega_{t-1}) \quad \text{(using chain rule)}$$

$$= p(a_t \mid s_t, \omega_{t-1})$$

$$\Rightarrow a_t \not\perp\!\!\!\perp \omega_{t-1} \mid s_t$$

Thus, in this case the action is not independent of ω_{t-1} given s_t. Hence, a simple mixture of models will not suffice for this more general case.