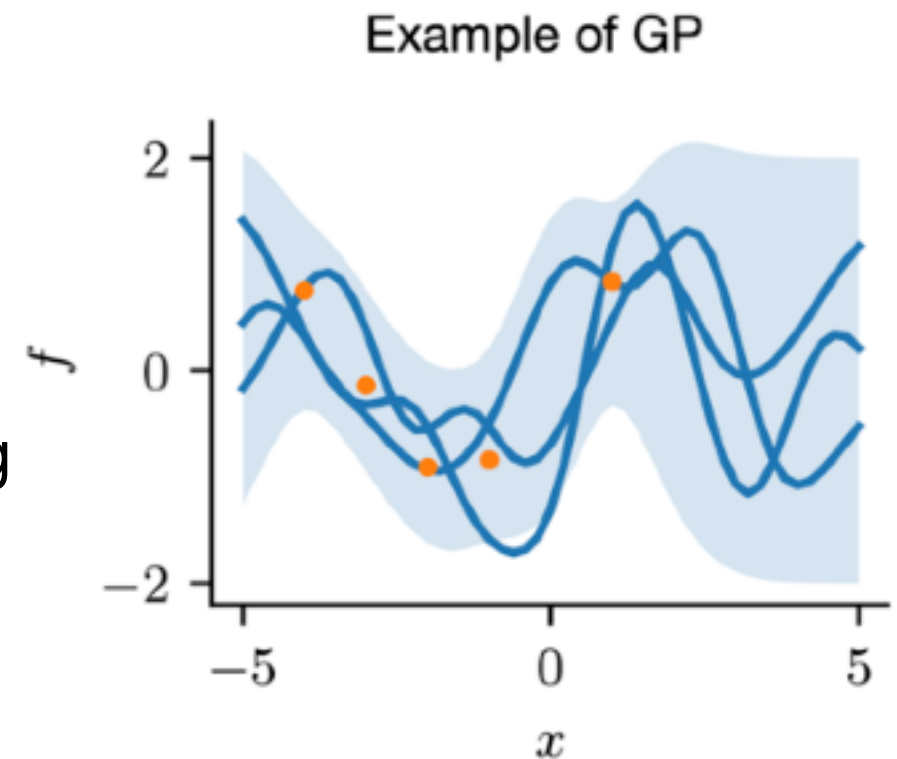


We don't know the model

- Learn the model from data:
 - ➔ Aleatoric uncertainties
 - ➔ Epistemic uncertainties - minimise model bias
- Gaussian processes (GPs) are used assuming $\mathbf{s}_{t+1} = \mathbf{f}(\mathbf{s}_t, \mathbf{a}_t, \omega_t)$ and $\omega_t \sim \mathcal{N}(0, \sigma)$
- Include if needed the emitted reward
- Use RBF Kernel - allow for analytical propagation of uncertainties
- Standard GPs training: evidence maximization



Uncertainty propagation

- Moment matching for deterministic propagation of the mean $\mu(s_t)$ and the covariance $\Sigma(s_t)$ of the distribution of dynamics-reward model
- The immediate performance measure is:
$$\mathbb{E}[r(s_t, a_t)] = \int r(s_t, a_t) \mathcal{N}(s_t | \mu_t, \Sigma_t) ds_t$$
- If reward not emitted - formulated as polynomial function