IDALAB

INTELLIGENT DATA ANALYTICS SALZBURG

PARIS
LODRON
UNIVERSITÄT
SALZBURG

Simon Hirländer

Tutorial RL4AA

Direct policy search

- RL as derivative free optimization:

➡ $\text{maximise}_{z \in \mathbb{R}^d} R(z) \Rightarrow \text{maximise}_{p(z)} \mathbb{E}_p[R(z)]$

- Parametrise a distribution $p(z;\theta) \Rightarrow$ maximise$_{p(\theta)}$ $\mathbb{E}_{p(z;\theta)}[R(z)]$

➡ Likelihood trick - estimate the derivative:

$$\nabla_\theta J(\theta) = \int R(z)\,\nabla_\theta p(z;\theta)dz = \int R(z)\frac{\nabla_\theta p(z;\theta)}{p(z;\theta)}p(z\ \theta)dz$$

- $$= \int R(z)\,\nabla_\theta \log p(z;\theta)p(z\ \theta)dz = \mathbb{E}_{p(z;\theta)}[R(z)\,\nabla_\theta \log p(z;\theta)]$$

- Unbiased gradient estimate of J, if sample efficiently from $p(z;\theta)$ and $\log p(z;\theta)$

- High variance

Score function

Direct policy search

- RL as derivative free optimization:
  - ➡ maximise$_{z \in \mathbb{R}^d} R(z) \Rightarrow$ maximise$_{p(z)} \mathbb{E}_p[R(z)]$
  - ➡ Parametrise a distribution $p(z;\theta) \Rightarrow$ maximise$_{p(\theta)} \mathbb{E}_{p(z;\theta)}[R(z)]$
  - ➡ Likelihood trick - estimate the derivative:

$$\nabla_\theta J(\theta) = \int R(z)\,\nabla_\theta p(z;\theta)dz = \int R(z)\frac{\nabla_\theta p(z;\theta)}{p(z;\theta)}p(z\ \theta)dz$$

- $$= \int R(z)\,\nabla_\theta \log p(z;\theta) p(z\ \theta)dz = \mathbb{E}_{p(z;\theta)}[R(z)\overset{\text{Score function}}{\boxed{\nabla_\theta \log p(z;\theta)}}]$$
- Unbiased gradient estimate of J, if sample efficiently from $p(z;\theta)$ and $\log p(z;\theta)$
- High variance

Probabilistic trajectories