

Reward-modulated inference

Matthew Alger and Buck Shlegeris

Unsupervised learning

minimise:

NLL (negative log likelihood) of observations
+ complexity of model

Reinforcement learning

minimise:
negative reward
+ complexity of model

Reward-modulated inference

minimise:

$\lambda * \text{NLL of observations}$
 $+ (1 - \lambda) * \text{reward}$
 $+ \text{complexity of model}$

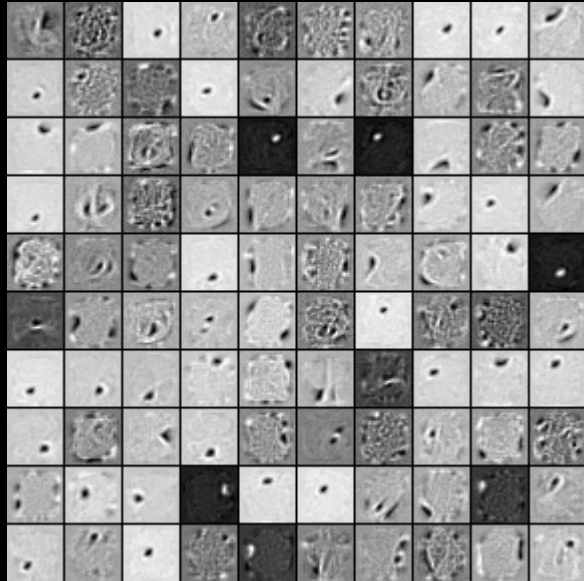
Logistic regression



minimise loss with model:

$$P = s(Wx + b)$$

Denoising Autoencoders



Noisy input →
Hidden layer →
Denoised output

Plan

- Implement RMI for typical RL tasks
- Determine RMI effectiveness