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:
λ * NLL of observations
+ (1 - λ) * reward
+ complexity of model
Logistic regression

minimise loss with model:
\[ P = s(Wx + b) \]

Image: MNIST dataset example, from CS365. http://cse.iitk.ac.in/users/cs365/2013/hw1.html
Denoising Autoencoders

Noisy input → Hidden layer → Denoised output

Plan

- Implement RMI for typical RL tasks
- Determine RMI effectiveness