

# Reinforcement Learning

# Vahvistettu oppiminen

- “Oikea tekoäly”
- Agentti oppii aidossa ympäristössä erehdyksen ja yrityksen avulla.
- Jos agentti toimii oikein se saa palkinnon jos väärin niin rangaistuksen
- Ajan myötä botti oppii toimimaan niin että se maksimoi palkinnon eri tilanteissa

# Policy network

- Unsupervised neuroverkko(FCL tai konvoluutio)
- Aloitetaan täysin sattumanvaraisella neuroverkolla
- Pong-pelissä Input on kuva pelistä output mahdollinen action (ylös, alas, pysy paikallaan)
- Policy gradient yritetään karsia huonot actionit pois

- Policy gradient on ongelmallinen monimutkaisimmissa tilanteissa
- Tekoälyn on opittava itse mikä toiminta johtaa positiiviseen lopputulokseen  
-> vaikea treenattava
- Credit Assignment Problem

- Reward shaping
- Jaetaan palkitseminen pienempiin tavoitteisiin
- Ongelmana että joka ympäristö tarvitsee oman reward function
- Agentti voi huijata