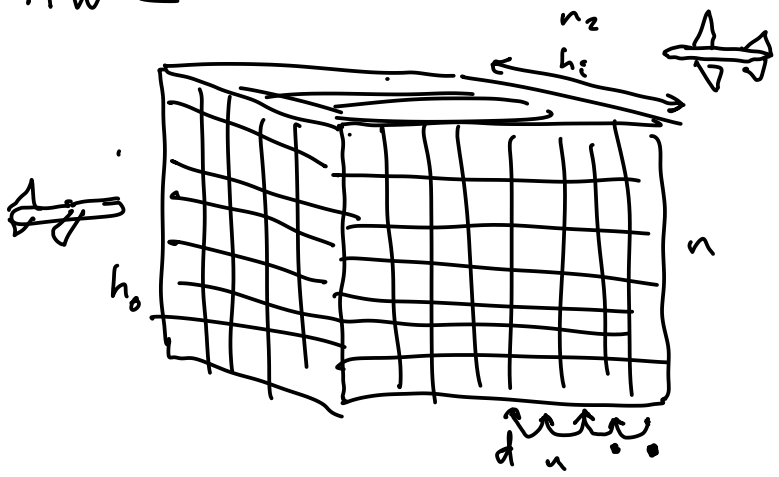
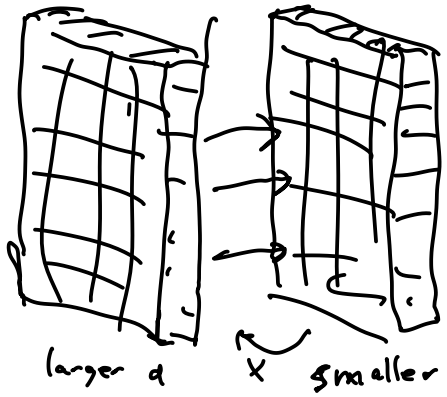
HW2



$|V| = n_1 \times n_2 \times n_3$

$|S| \times |S|$

$\frac{|S|}{\#d} \times \frac{|S|}{\#d}$



$V_{new} = V$

Gause - Seidel Value Iteration

for s

for a

$Q[a] = R[a][s] + \gamma[a][s,:] \cdot V$

$V[s] =$ maximum

same Vector

Reinforcement Learning

Model Uncertainty

Optimal Control

Known S, A, T, R

Aleatory Uncertainty

No Learning

Explicit

$T(s'|s,a)$

Machine Learning

$\hat{S}, \hat{A}, \hat{T}, \hat{R}$

Epistemic Uncertainty

Learning

Generative

$G(s, a, w)$

Environment Interaction

$s', r = \text{step}(\text{env}, a)$

mutable rep. of state

Exploration

- Episodic
- Continuing Task

Return $= \sum_{t=0}^{\infty} r_t$

Algorithms with Guarantees ⟷ Tricks

Classical RL

Deep RL

Theorems ⟷ Opinions

Smaller Problems ⟷ Larger Problems

Grid World ⟷ Automonous Driving Cartoon ⟷ Star Craft

Ben Recht

Alex Irpan

Sergei Levine

2 reasons people use RL

1. Don't have a good model
2. Problem is so big nothing else can handle

---

3 Challenges

1. Balancing Exploration + Exploitation ← Ordering Food
2. Credit Assignment
3. Generalization

Secretary Problem

$\frac{1}{10}$ $\lim_{n \to \infty}$

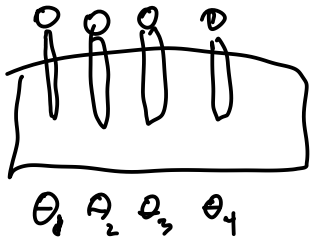
~~Finding a Spouse~~

Finding a House reject $\frac{1}{e}$ 35% e^x

pick next that is better than all previous

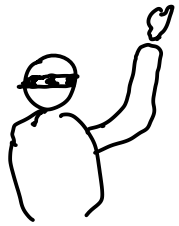
$\frac{1}{e}$ chance of picking best

k-armed bandit



$p(1)=\theta$

Bernoulli: (θ)
k-armed bandit

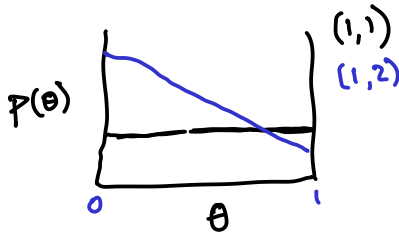


Estimate θ $\quad \hat{\theta} = \frac{w}{w+l}$ Max likelihood

Beta distribution : Dist <u>over</u> Bernoulli Dist

↖ conjugate

Beta $(w+1, l+1)$
pseudo-counts



Conjugate Priors

Normal

Dirichlet Distribution : Disribution over <u>Categorical</u>

n-1 parameters