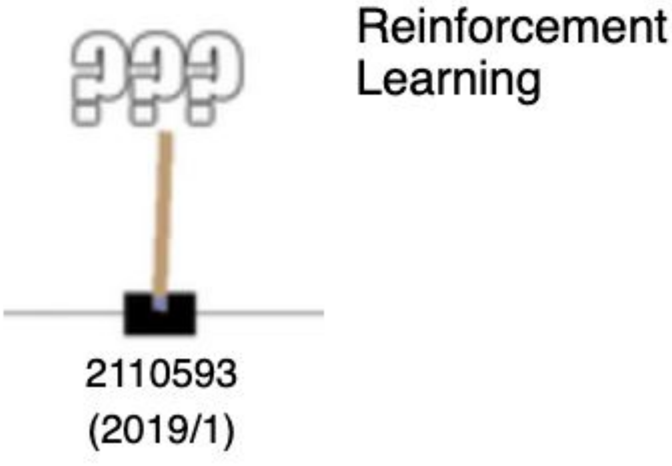
Reinforcement Learning

Introduction

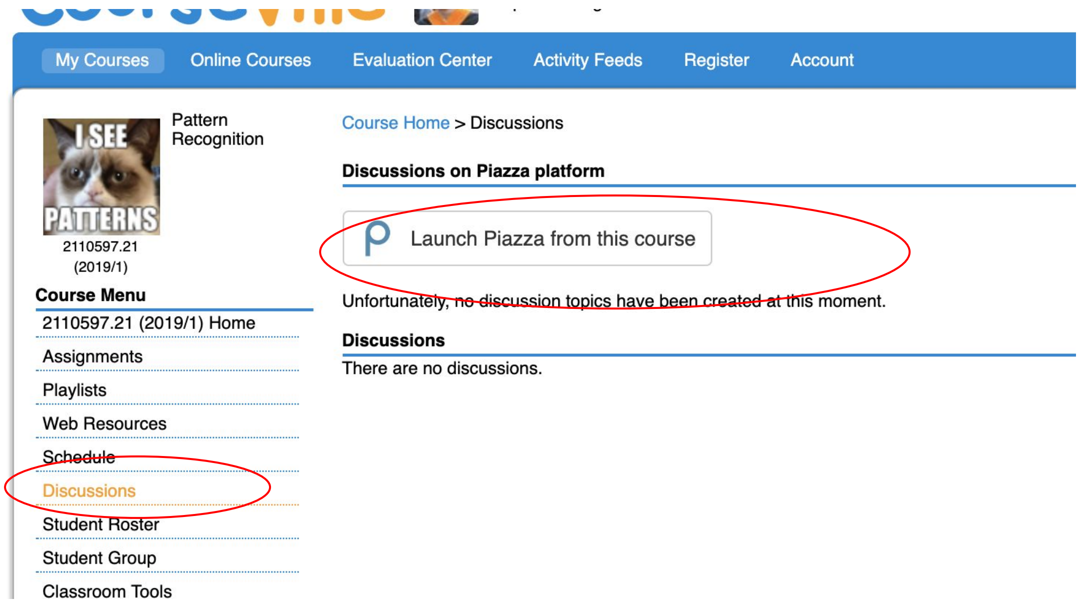
Mycourseville & github

2110593 (2019/1)

https://github.com/ekapolc/RL_course_2019

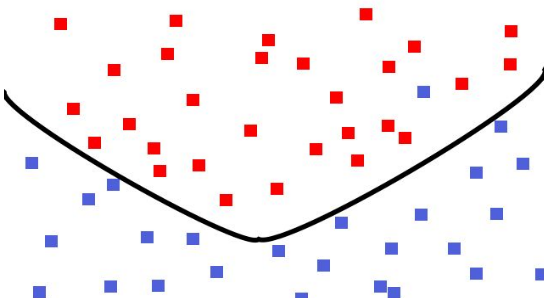


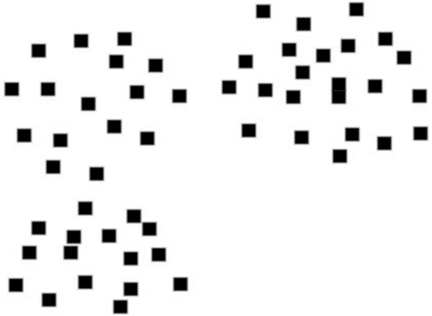
Piazza


My Courses
Online Courses
Evaluation Center
Activity Feeds
Register
Account
I SEE PATTERNS
Pattern Recognition
2110597.21 (2019/1)
Course Menu
2110597.21 (2019/1) Home
Assignments
Playlists
Web Resources
Schedule
Discussions
Student Roster
Student Group
Classroom Tools
Course Home > Discussions
Discussions on Piazza platform
Launch Piazza from this course
Unfortunately, no discussion topics have been created at this moment.
Discussions
There are no discussions.

Syllabus

3 Modes of Learning

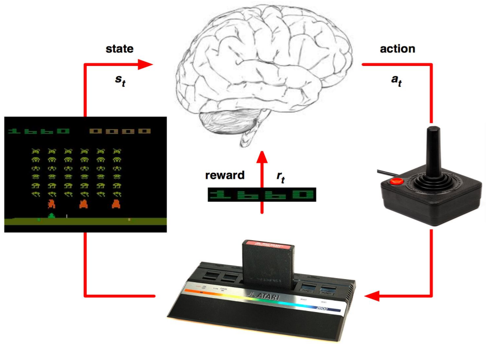




Supervised Learning

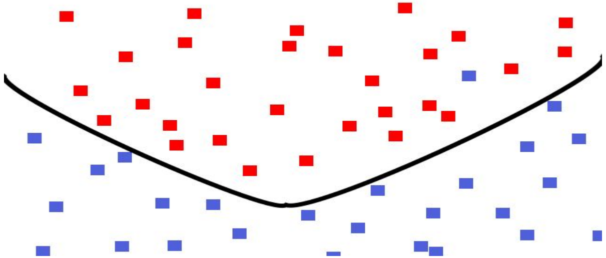
Reinforcement Learning

Unsupervised Learning



3 Modes of Learning

Supervised Learning

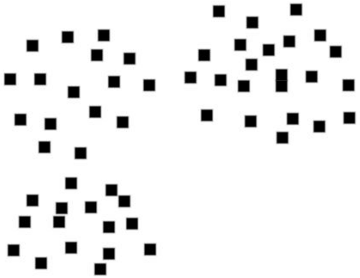


- Observe:
  - $(x_1, y_1), (x_2, y_2), \ldots$
- Objective:
  - Input an unseen x_{new}
  - What is y_{new}?

3 Modes of Learning

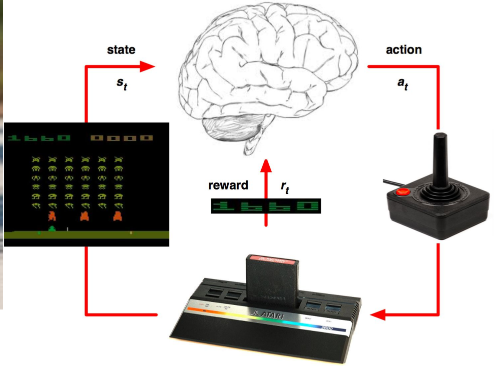
Unsupervised Learning

- Observe:
  - $x_1, x_2, x_3, x_4, \ldots$
- Objective:
  - What is P(x) ?
  - What is a *good* representation of x ?
  - What can we learn from P(x) ?



3 Modes of Learning

Reinforcement Learning (RL)



- Observe:
  - The states (x_1, x_2, x_3, …)
  - The reward (r_1, r_2, r_3, …)
- Can also take actions
  - a_1 , a_2 , a_3 , …
- What are the best actions?
  - Such that we will receive highest accumulative rewards

What is RL?

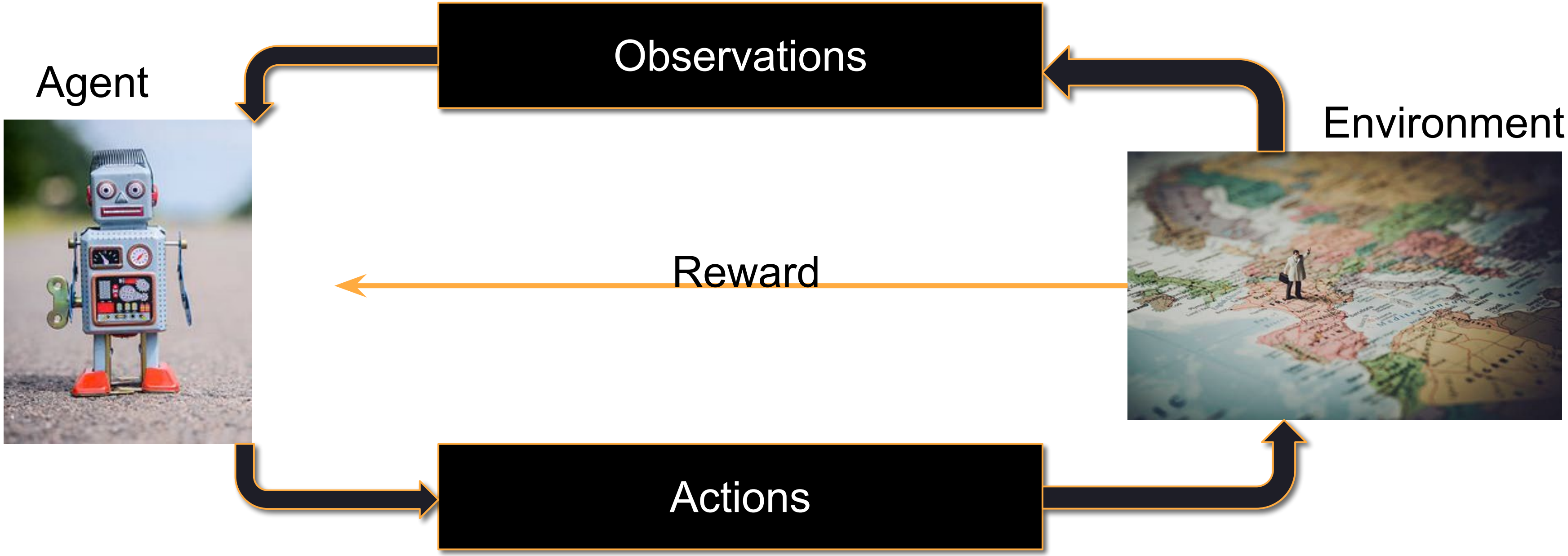
1) A problem
2) A community working on 1)
3) Methods produced by 2) which can be applicable to other problems

Benjamin Van Roy

Professor at Stanford University; Research Lead at DeepMind, Mountain View

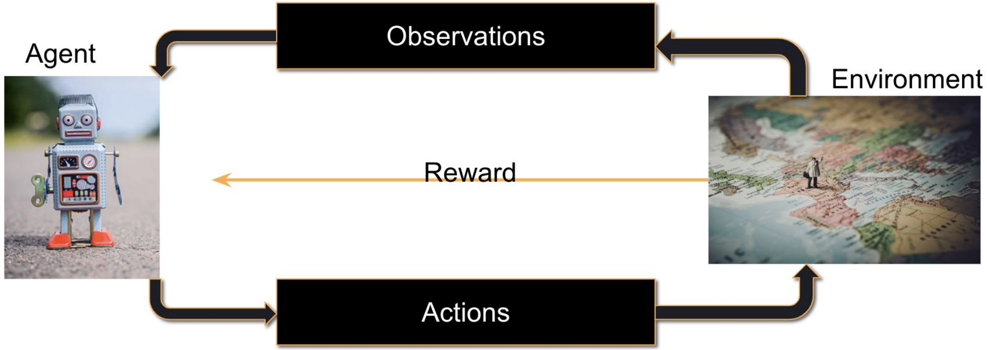
Topic: Reinforcement Learning

RL problem



Difference between RL and other modes of learning

- Sequential decisions
- You have a goal vs
  You have means to get there
- No concept of “training set” and “test set”
- “Passive” vs “Active” learning



RL(DeepMind)'s goal

Even a 4-1 victory for Lee Se-dol would represent a major achievement for DeepMind, a British company acquired by Google for a reported $400 million in 2014. The unit's ultimate mission is no less than to "solve intelligence," with potential uses ranging from healthcare to robotics, but attaining the long-sought computer science dream of a world-beating Go program would catapult DeepMind to the forefront of AI research.

https://www.theverge.com/2016/3/8/11178462/google-deepmind-go-challenge-ai-vs-lee-sedol

Now time for videos

https://www.youtube.com/watch?v=wiOopO9jTZw

2017

Now time for videos

https://blog.openai.com/openai-five/
https://youtu.be/eHipy_j29Xw

2018

Now time for videos

https://deepmind.com/blog/article/alphastar-mastering-real-time-strategy-game-starcraft-ii

Now time for videos

https://ai.facebook.com/blog/pluribus-first-ai-to-beat-pros-in-6-player-poker/

https://www.youtube.com/watch?v=u90TbxK7VEA

RL and Artificial General Intelligence

Yann LeCun

March 14, 2016 ·

Follow

Statement from a Slashdot post about the AlphaGo victory: "We know now that we don't need any big new breakthroughs to get to true AI"

That is completely, utterly, ridiculously wrong.

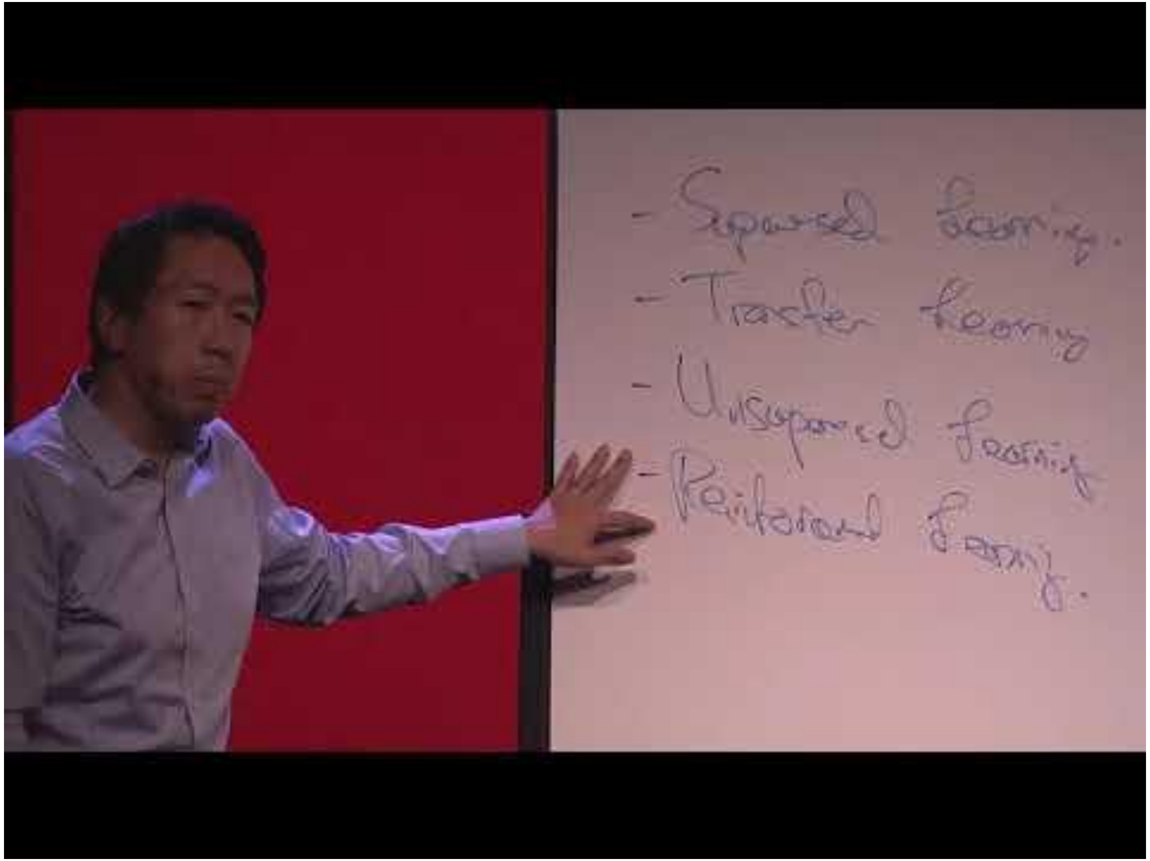
As I've said in previous statements: most of human and animal learning is unsupervised learning. If intelligence was a cake, unsupervised learning would be the cake, supervised learning would be the icing on the cake, and reinforcement learning would be the cherry on the cake. We know how to make the icing and the cherry, but we don't know how to make the cake.

We need to solve the unsupervised learning problem before we can even think of getting to true AI. And that's just an obstacle we know about. What about all the ones we don't know about?

#deeplearning #AI #AlphaGo

RL and $$$

“The excitement and PR hype behind reinforcement learning is a bit disproportionate relative to the economic value it’s creating today”
- Andrew Ng



https://www.youtube.com/watch?v=NQK4ZY_gwKI&feature=youtu.be&t=1m16s

https://craft.co/deepmind/metrics

DeepMind Income Statement

Annual

GBP	FY, 2016	FY, 2017
Revenue	40.3m	54.4m
Revenue growth, %		*35%*
General and administrative expense	163.8m	333.9m
Operating expense total	163.8m	333.9m
Depreciation and amortization	863.5k	1.9m
EBIT	(123.5m)	(279.4m)
EBIT margin, %	*(307%)*	*(513%)*
Interest expense	2.0m	2.6m
Pre tax profit	(126.6m)	(281.9m)
Income tax expense	32.6m	(20.3m)
Net Income	(93.9m)	(302.2m)

RL use cases

Go, chess, starcraft, dota, poker

Finance (https://www.jpmorgan.com/global/LOXM)...

Robotics...but...

https://research.googleblog.com/2016/03/deep-learning-for-robots-learning-from.html

https://towardsdatascience.com/applications-of-reinforcement-learning-in-real-world-1a94955bcd12
https://www.oreilly.com/ideas/practical-applications-of-reinforcement-learning-in-industry

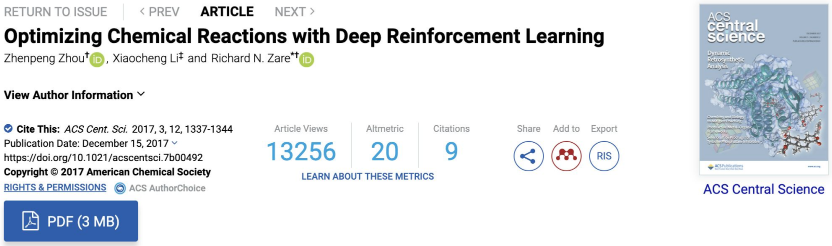
RL use cases

Data center and resource management (https://people.csail.mit.edu/alizadeh/papers/deeprm-hotnets16.pdf System configuration http://ranger.uta.edu/~jrao/papers/ICDCS09.pdf DRAM controller https://ieeexplore.ieee.org/abstract/document/4556714/ Recommender (Bandits) https://people.cs.umass.edu/~pthomas/papers/Barto2017.pdf)

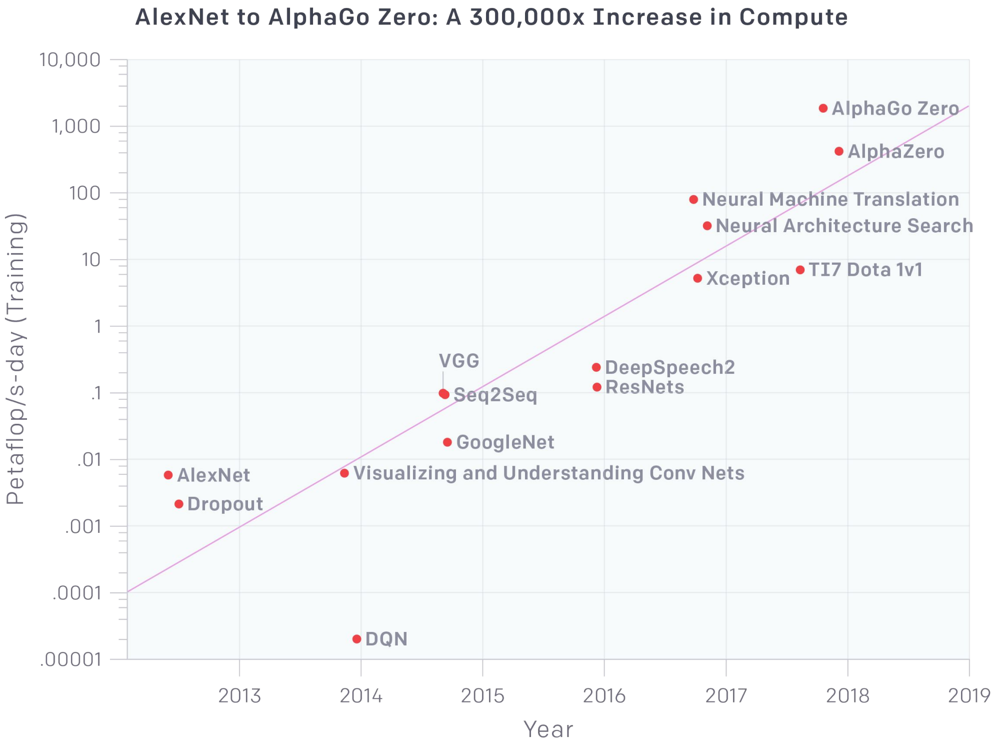
Ad bidding (https://arxiv.org/abs/1701.02490)

Chemistry (https://pubs.acs.org/doi/full/10.1021/acscentsci.7b00492)

Some other tasks that use algorithms from RL to help perform model training (autoML, REINFORCE)



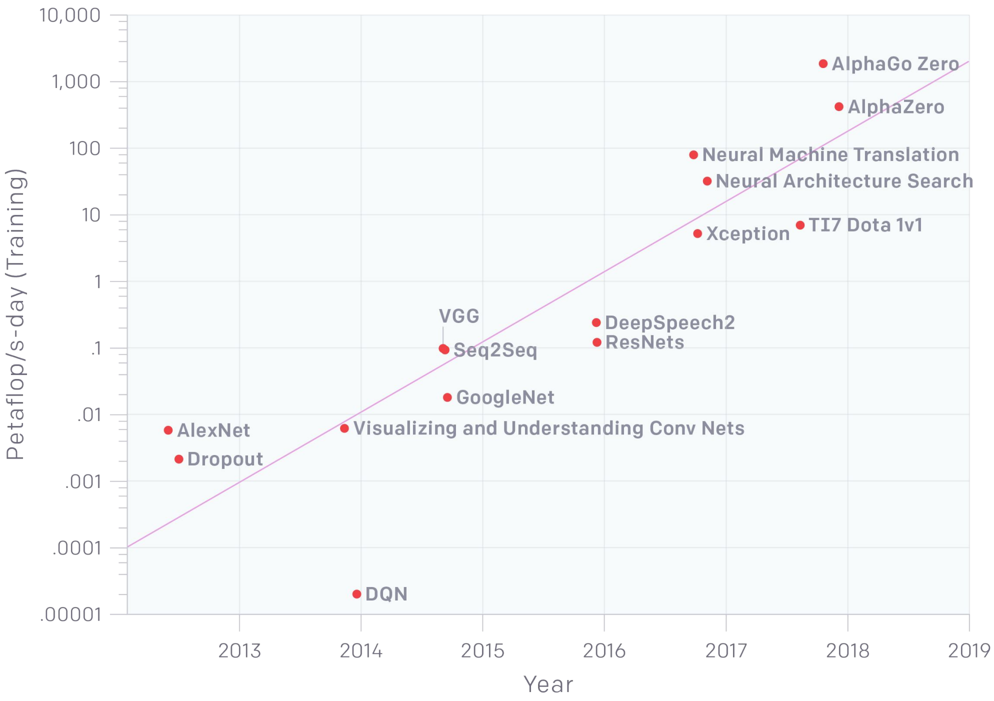
Problems with (current) RL

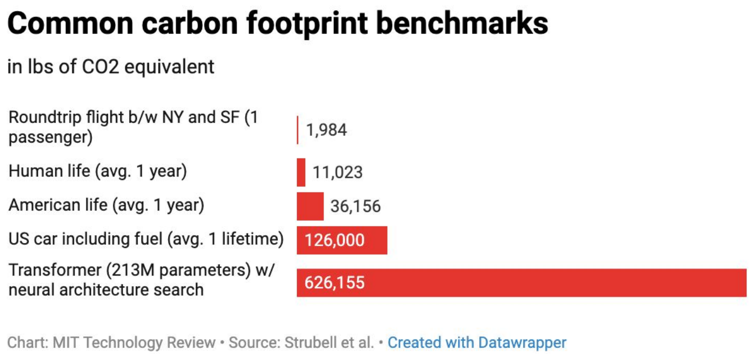


https://blog.openai.com/ai-and-compute/

Problems with (current) RL

AlexNet to AlphaGo Zero: A 300,000x Increase in Compute





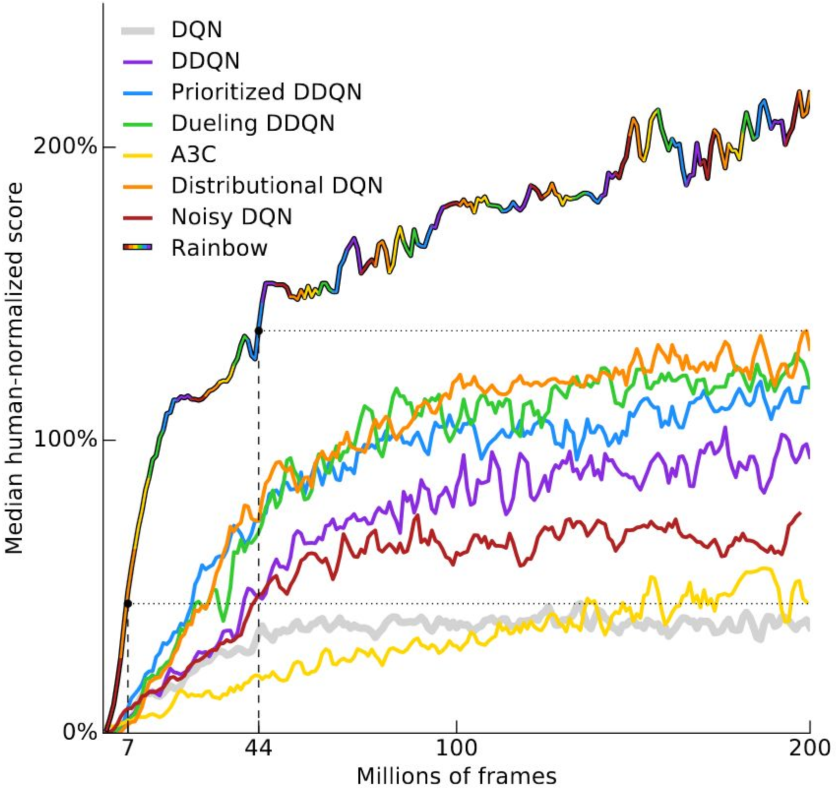
https://www.technologyreview.com/s/613630/training-a-single-ai-model-can-emit-as-much-carbon-as-five-cars-in-their-lifetimes/

https://blog.openai.com/ai-and-compute/

Problems with (current) RL

Data inefficient

- Many use case can be better solved with supervised learning (efficiency and accuracy)



https://arxiv.org/abs/1710.02298

Problems with (current) RL

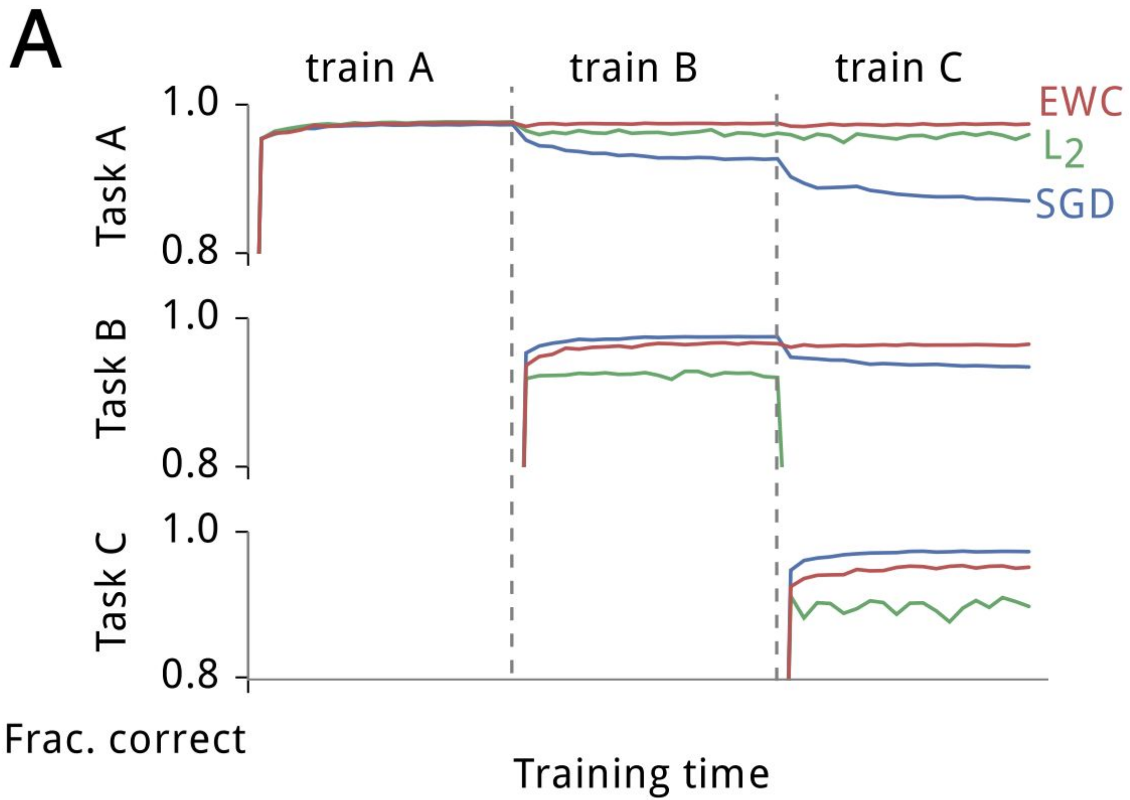
Rewards engineering is hard

- Sparse rewards

https://www.youtube.com/watch?v=tlOIHko8ySg

Problems with (current) RL

Catastrophic forgetting

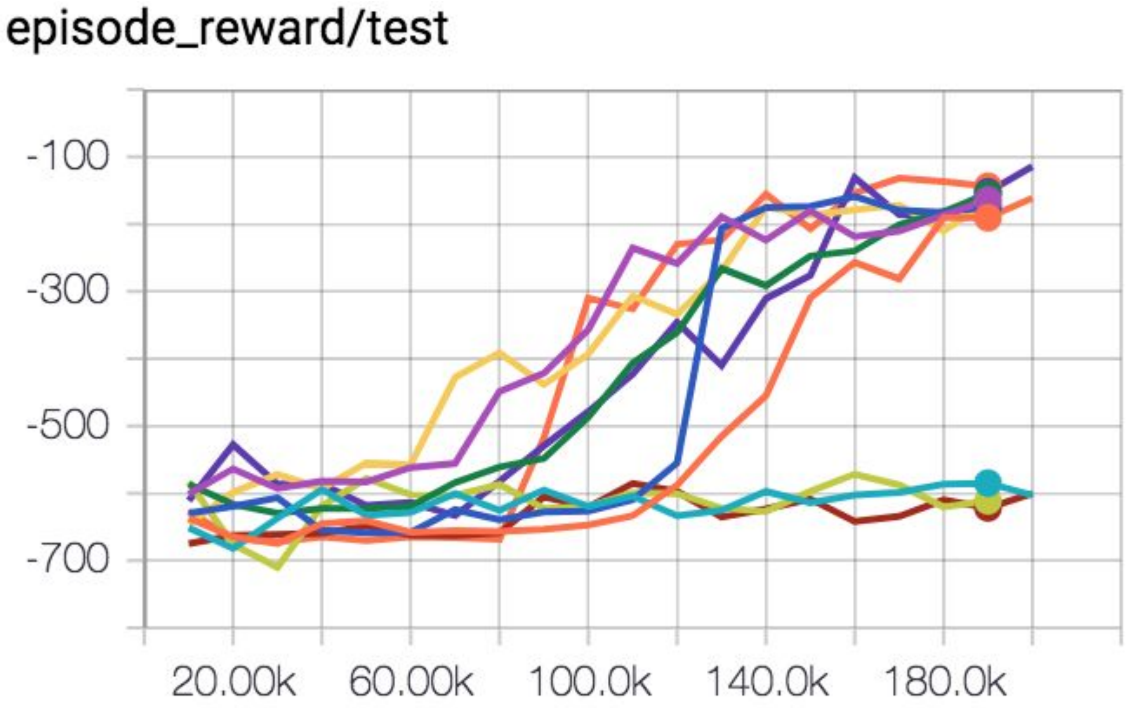


http://www.pnas.org/content/early/2017/03/13/1611835114.full.pdf?with-ds=yes

Problems with (current) RL

Randomness

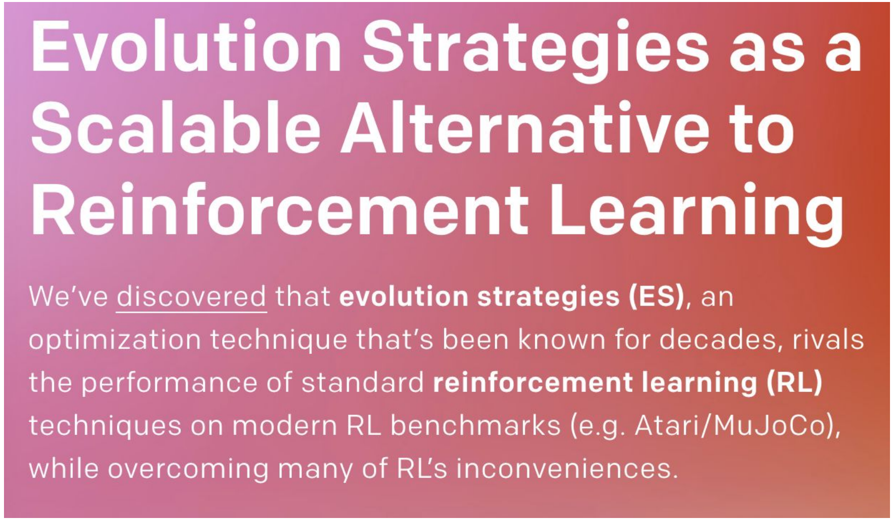
- Random initialization
- Random exploration
- Random environment



https://www.alexirpan.com/2018/02/14/rl-hard.html

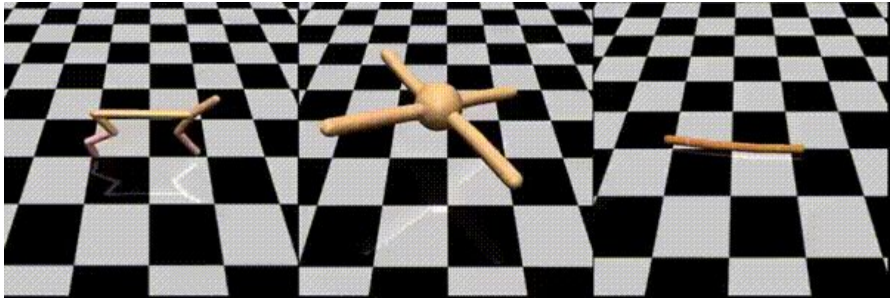
Problems with (current) RL

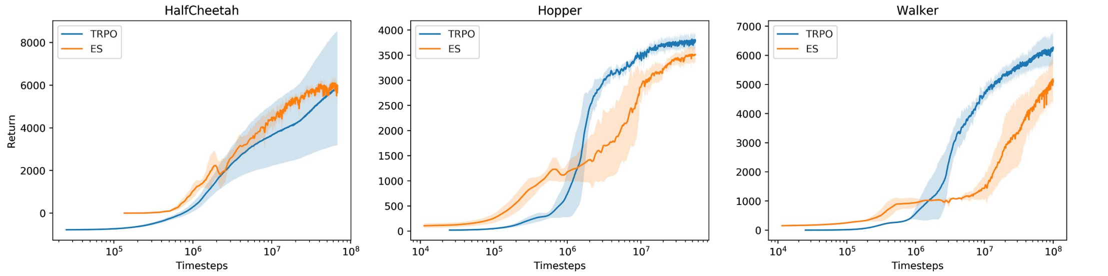
Sometimes evolution strategies might be better



Evolution Strategies as a Scalable Alternative to Reinforcement Learning

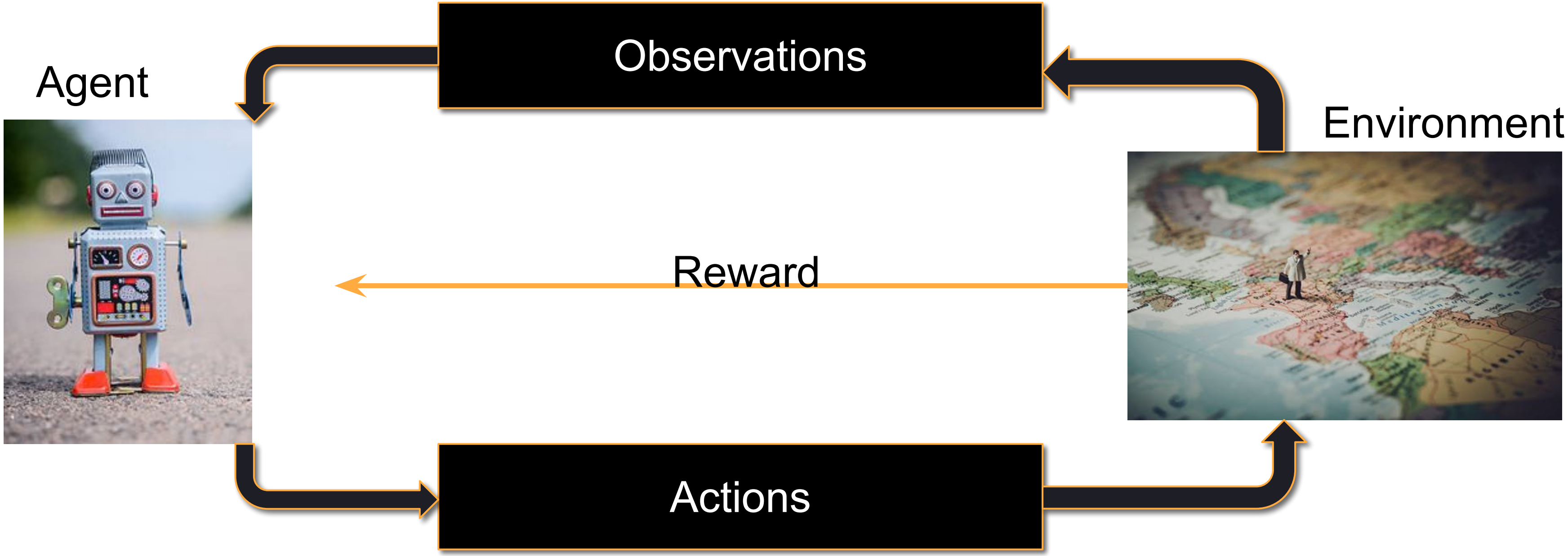
We've discovered that **evolution strategies (ES)**, an optimization technique that's been known for decades, rivals the performance of standard **reinforcement learning (RL)** techniques on modern RL benchmarks (e.g. Atari/MuJoCo), while overcoming many of RL's inconveniences.





https://openai.com/blog/evolution-strategies/

RL problem



Imitation learning

Learn through experts actions

Becomes a supervised learning problem

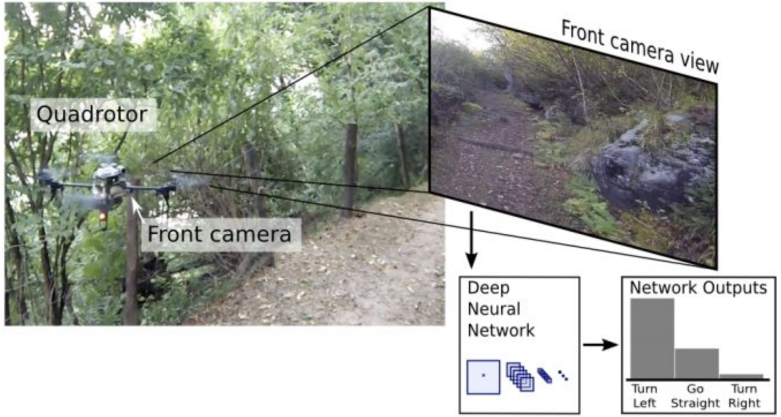


Fig. 1: Our quadrotor acquires the trail images from a forward-looking camera; a Deep Neural Network classifies the images to determine which action will keep the robot on the trail.

Fig. 4: *Left:* stylized top view of the acquisition setup; *Right:* our hiker during an acquisition, equipped with the three head-mounted cameras.

http://rpg.ifi.uzh.ch/docs/RAL16_Giusti.pdf

Imitation learning

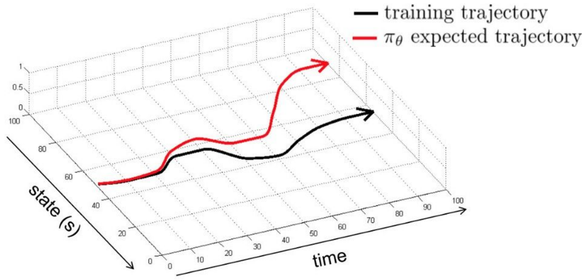
Learn through experts actions

Becomes a supervised learning problem

What if the agent goes into regions where we don't have expert's supervision?

Needs some kind of compensation of difference behavior

Can be used in conjunction with RL



http://rail.eecs.berkeley.edu/deeprlcourse-fa17/f17docs/lecture_2_behavior_cloning.pdf

Review of probabilities

Notation

Expectation (multivariate)

- Correlation - correlation vs causation

Variance

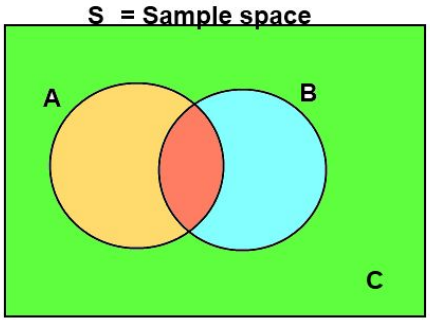
Sampling from a distribution (gaussian, uniform, softmax)

Estimation (MLE)

Conditional probability

- P(A|B) probability of A given B has occurred

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$



Different notations P(A| B = b)

https://www.wyzant.com/resources/lessons/math/statistics_and_probability/probability/further_concepts_in_probability

Independence

- Two events are independent (statistically independent or stochastically independent) if the occurrence of one does not affect the probability of occurrence of the other.

$$\mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(B) \Leftrightarrow \mathrm{P}(B) = \mathrm{P}(B \mid A)$$

Bayes' Rule (Bayes's theorem or Bayes' law)

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)}$$

Usefulness: We can find P(A|B) from P(B|A) and vice versa

Expected value

- Expected value

$$\boxed{E[x] = \int_{-\infty}^{\infty} x p(x) dx}$$

$$\boxed{E[g(x)] = \int_{-\infty}^{\infty} g(x) p(x) dx}$$

- Variance (σ^2) (Standard Deviation = σ)

$$Var[x] = E[(x - E[x])^2] = \sigma^2 = \int_{-\infty}^{\infty} (x - E[x])^2 p(x) dx$$

$$E[(x - E[x])^2] = E[x^2] - (E[x])^2$$

Expected value and Variance properties

- $E[a] = a$; a is a constant.
- $E[aX+b] = aE[X]+b$
- $E[X+Y] = E[X]+E[Y]$
- $Var[a] = 0$
- $Var[aX+b] = a^2 Var[X]$

Conditional Expected Value

$$E[x \mid A] = \int_{-\infty}^{\infty} x p(x \mid A) dx$$

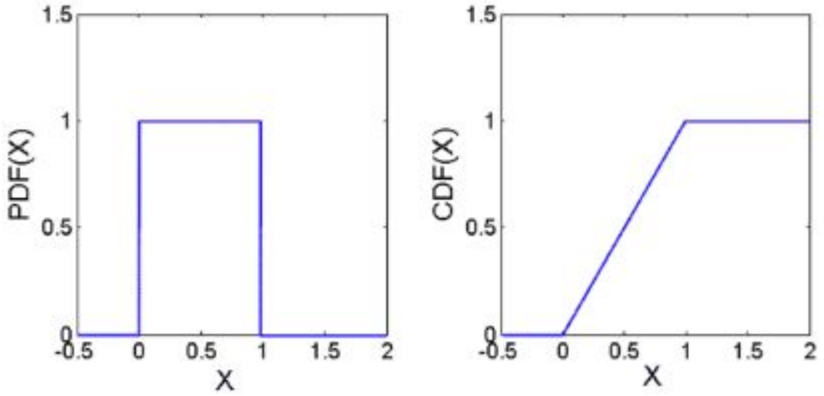
$$E[g(x) \mid A] = \int_{-\infty}^{\infty} g(x) p(x \mid A) dx$$

Cumulative Distribution Functions CDFs

- Probability that the RV is less than a certain amount

$$F_X(x_0) = P(X \leq x_0) = \int_{-\inf}^{x_0} p(x)dx$$

- CDF is the integral of PDF. Differentiating CDF wrt x gives the PDF



Useful for sampling

Joint distributions

- If we want to monitor how two events are jointly occurring, we consider the joint distribution $p_{X,Y}(x,y)$
- $p_{X,Y}(x,y) = p_X(x)p_Y(y)$ if x and y are independent

$$P(A) = \iint_A p_{XY}(x,y)dxdy$$

$$p_X(x) = \int_{-\infty}^{\infty} p_{XY}(x,y)dy$$

$$p_Y(y) = \int_{-\infty}^{\infty} p_{XY}(x,y)dx$$

Can we take expectation from a joint distribution?
What about a conditional expectation?

Expectation of multivariate distributions

$$E[g(X_1, X_2)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x_1, x_2) p_{X_1, X_2}(x_1, x_2) dx_1 dx_2$$

$$E[g(X_1)h(X_2)] = E[g(X_1)]E[h(X_2)]$$

If X_1 and X_2 independent

Sum of Random variables

- Z = Y + X
- What is the pdf of Z? Where Y and X continuous RVs

$$p_{X+Y}(z) = (p_X * p_Y)(z) = (p_Y * p_X)(z)$$

Central Limit Theorem (CLT)

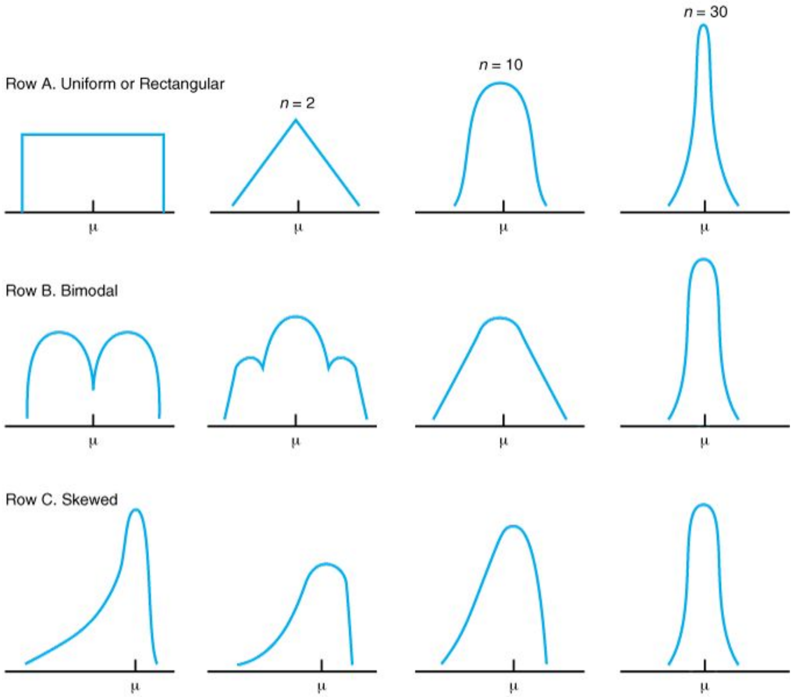
- Suppose $X_1, X_2, \ldots$ is a sequence of iid (independent and identically distributed) RVs. As n approaches infinity the sum of the sequence converge in distribution to a Normal distribution

$$\sqrt{n}\left(\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - \mu\right) \xrightarrow{d} N\left(0, \sigma^2\right)$$

- Other variants of CLT exists, without the independence or identically distributed assumption

CLT implications

- A sum of RVs tends to become Normally distributed very quickly



Gaussian distribution (normal distribution)

- X is normal (Gaussian): $X \sim N(\mu,\sigma^2)$

$$\boxed{p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}}$$

$$E[x] = \mu$$

$$Var[x] = \sigma^2$$

- X is Standard normal (Standard Gaussian):

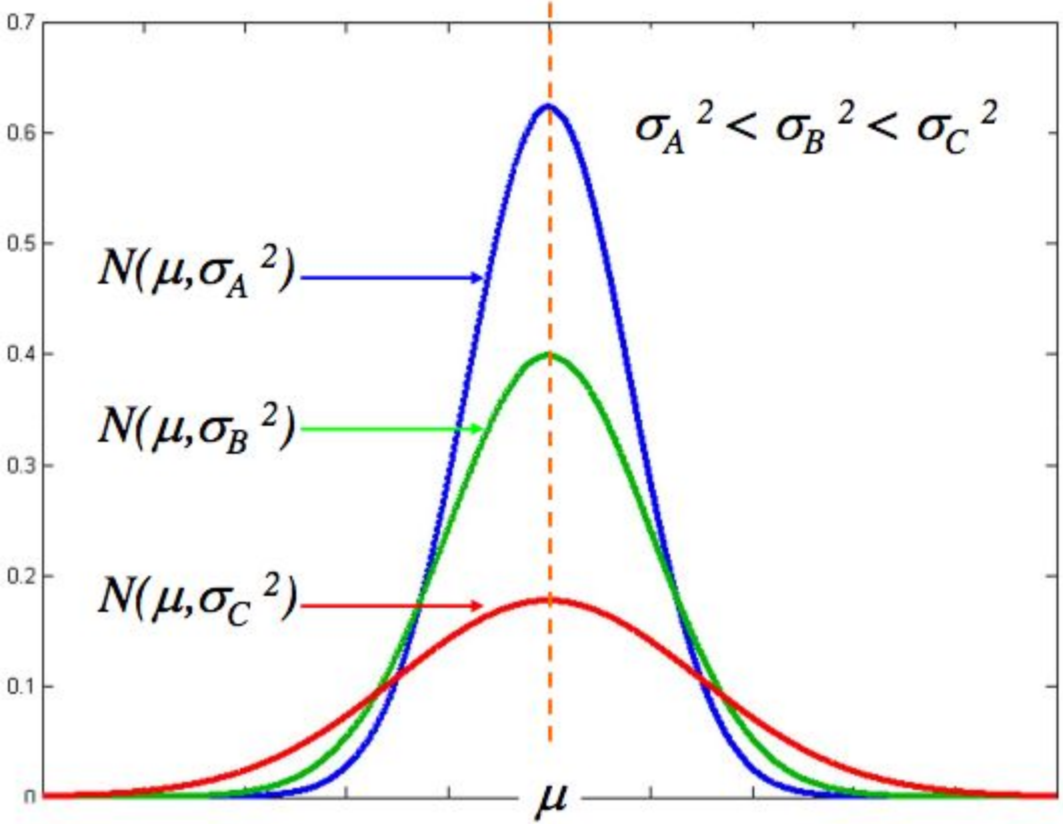
$X \sim N(0,1)$ when $\mu=0,\ \sigma^2=1$

$$\boxed{p(x) = \frac{1}{\sqrt{2\pi}} e^{-(x-\mu)^2/2}}$$

$$E[x] = 0$$

$$Var[x] = 1$$

Gaussian pdf



Linear transformation of Gaussian RV

- Normality is preserved by linear transformation. Calculation involving the normal variable is usually done in terms of standard normal.
- Let $Y = aX + b$,

  if $X \sim N(\mu, \sigma^2) \rightarrow Y \sim N(a\mu + b, a^2\sigma^2)$

- Let $Z = (X - \mu)/\sigma$,

  if $X \sim N(\mu, \sigma^2) \rightarrow Z \sim N(0, 1)$: Standard Normal

Can you prove this?

Summation of 2 Gaussian RVs

- X mean m_1 variance σ_1^2
- Y mean m_2 variance σ_2^2
- X and Y are independent

- X+Y is normally distributed with mean m_1+m_2 variance $\sigma_1^2+\sigma_2^2$

Covariance of multivariate distributions

- $cov(X_1,X_2) = E[(X_1 - m_1)(X_2 - m_2)]$

- $cov(X_1,X_2) = E[(X_1)(X_2)] - m_1 m_2$

- Covariance with itself is just the Variance
- Correlation

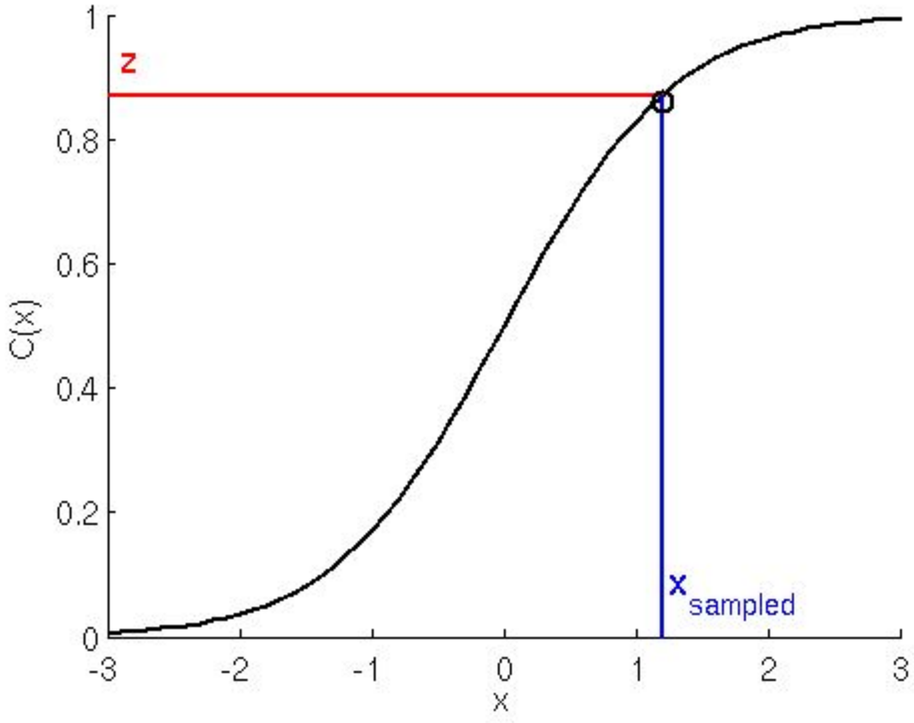
$$\rho = \frac{cov(X_1, X_2)}{\sqrt{V(X_1)V(X_2)}}$$

Covariance matrix

- Given a set of RVs, X_1 X_2 … X_n
- The covariance matrix is a matrix which has the covariance of the i and j RV in position (i,j)

$$\Sigma = \begin{bmatrix} \mathrm{E}[(X_1-\mu_1)(X_1-\mu_1)] & \mathrm{E}[(X_1-\mu_1)(X_2-\mu_2)] & \cdots & \mathrm{E}[(X_1-\mu_1)(X_n-\mu_n)] \\ \mathrm{E}[(X_2-\mu_2)(X_1-\mu_1)] & \mathrm{E}[(X_2-\mu_2)(X_2-\mu_2)] & \cdots & \mathrm{E}[(X_2-\mu_2)(X_n-\mu_n)] \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{E}[(X_n-\mu_n)(X_1-\mu_1)] & \mathrm{E}[(X_n-\mu_n)(X_2-\mu_2)] & \cdots & \mathrm{E}[(X_n-\mu_n)(X_n-\mu_n)] \end{bmatrix}.$$

Sampling using the inverse of the CDF



Draw from a uniform random generator

Look at the inverse of the CDF for the new x

Estimation

Bias and variance of estimators

Estimation

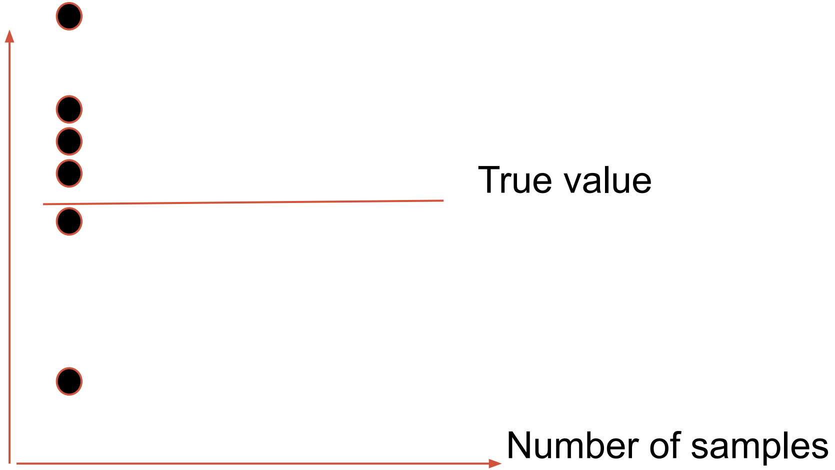
Given observations $x_1,x_2,x_3,\ldots,x_n$

Find the sample mean and sample variance

Estimator bias

$E[\hat{x} - E[X]]$

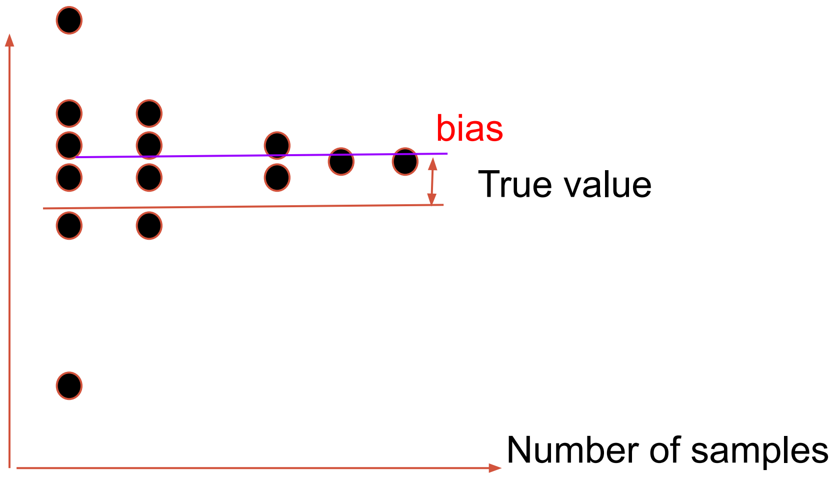
True value



Estimator bias

$E[\hat{x} - E[X]]$

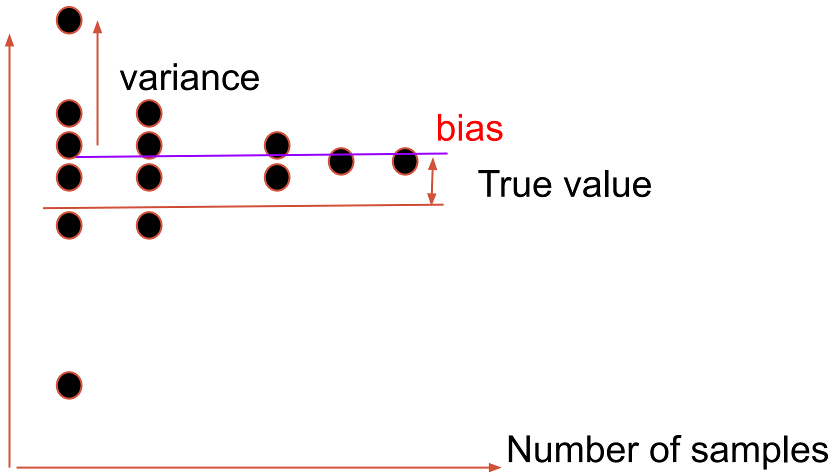
True value



Estimator variance

$$E[(\hat{X} - E[\hat{X}])^2] = Var[\hat{X}]$$

Expected estimator



Summary

What is RL

Probability review

Homework

Read chapter 1 of Sutton

Answer questions 1.1-1.5