

# Part 3. Bandit Feedback and Likelihood Models for Recommendation

David Rohde, Flavian Vasile  
Criteo Research

*June 26, 2019*

## Collaborators:

Special thanks to:

- Olivier Jeunen
- Dmytro Mykhaylov
- Martin Bompaire
- Stephen Bonner

# Notation

Symbol	Dimension	Description
$u$	Scalar	A given users id.
$t$	Scalar	sequential time.
$P$	Scalar	Total number of products.
$v_t$	Scalar	Product id viewed by user at time $t$ .
$r_t$	Scalar	Product id recommended at time $t$ .
$c_t$	Scalar	Product id recommended at time $t$ .

Table 1: Notations and Definitions

# Classic Reco

## Motivating Example



Product view



# Classic Reco

## Motivating Example



Product  
view



# Classic Reco

## Motivating Example

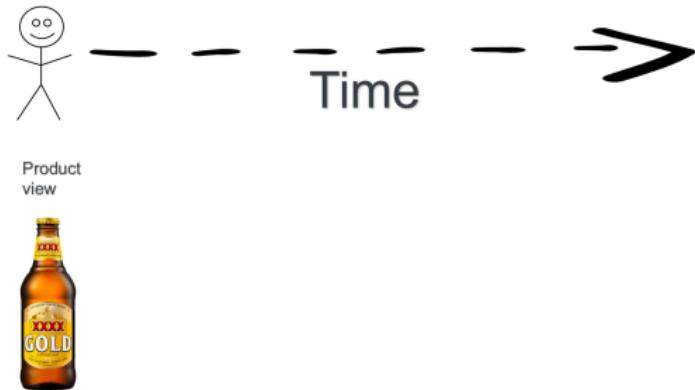


Product  
view



# Classic Reco

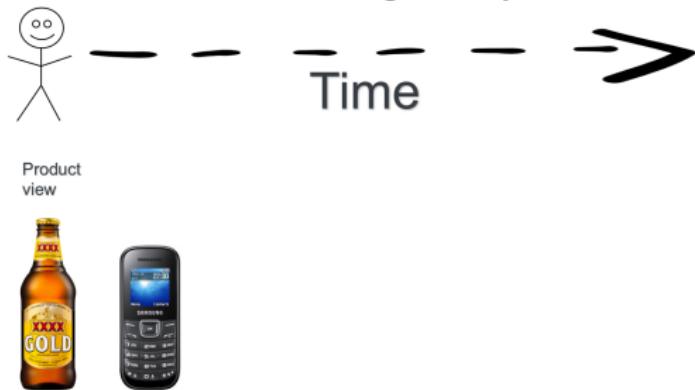
## Motivating Example



$$P(V_1 = \text{beer})$$

# Classic Reco

## Motivating Example



$$P(V_2 = \text{phone A} | V_1 = \text{beer})$$

# Classic Reco

## Motivating Example



$$P(V_3 = \text{phone B} | V_1 = \text{beer}, V_2 = \text{phone A})$$

## Next Item Prediction, hold out $V_3$

$$P(V_3 = \text{phone A} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.27$$

$$P(V_3 = \text{phone B} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.24$$

$$P(V_3 = \text{rice} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.10$$

$$P(V_3 = \text{couscous} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.09$$

$$P(V_3 = \text{beer} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.30$$

## Next Item Prediction, hold out $V_3$

$$P(V_3 = \text{phone A} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.27$$

$$P(V_3 = \text{phone B} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.24$$

$$P(V_3 = \text{rice} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.10$$

$$P(V_3 = \text{couscous} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.09$$

$$P(V_3 = \text{beer} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.30$$

What actually happened?

## Next Item Prediction, hold out $V_3$

$$P(V_3 = \text{phone A} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.27$$

$$P(V_3 = \text{phone B} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.24$$

$$P(V_3 = \text{rice} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.10$$

$$P(V_3 = \text{couscous} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.09$$

$$P(V_3 = \text{beer} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.30$$

What actually happened?  $V_3 = \text{phone B}$

## Next Item Prediction, hold out $V_3$

$$P(V_3 = \text{phone A} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.27$$

$$P(V_3 = \text{phone B} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.24$$

$$P(V_3 = \text{rice} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.10$$

$$P(V_3 = \text{couscous} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.09$$

$$P(V_3 = \text{beer} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.30$$

What actually happened?  $V_3 = \text{phone B}$

Our model assigned:  $P(V_3 = \text{phone B} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.24$

## Next Item Prediction, hold out $V_3$

$$P(V_3 = \text{phone A} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.27$$

$$P(V_3 = \text{phone B} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.24$$

$$P(V_3 = \text{rice} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.10$$

$$P(V_3 = \text{couscous} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.09$$

$$P(V_3 = \text{beer} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.30$$

What actually happened?  $V_3 = \text{phone B}$

Our model assigned:  $P(V_3 = \text{phone B} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.24$

Of course better would be higher, but how good is this?

## Next Item Prediction, hold out $V_3$ : Log likelihood

What actually happened?

## Next Item Prediction, hold out $V_3$ : Log likelihood

What actually happened?  $V_3$  = phone B

## Next Item Prediction, hold out $V_3$ : Log likelihood

What actually happened?  $V_3 = \text{phone B}$

Our model assigned:  $P(V_3 = \text{phone B} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.24$

## Next Item Prediction, hold out $V_3$ : Log likelihood

What actually happened?  $V_3 = \text{phone B}$

Our model assigned:  $P(V_3 = \text{phone B} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.24$

Log Likelihood:

$\log P(V_3 = \text{phone B} | V_1 = \text{beer}, V_2 = \text{phone A}) = -1.427116$

## Next Item Prediction, hold out $V_3$ : Log likelihood

What actually happened?  $V_3 = \text{phone B}$

Our model assigned:  $P(V_3 = \text{phone B} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.24$

Log Likelihood:

$$\log P(V_3 = \text{phone B} | V_1 = \text{beer}, V_2 = \text{phone A}) = -1.427116$$

Rationale: you want to assign high values to things that happen.

## Next Item Prediction, hold out $V_3$ : HR@K Recall@K Precision@K - K=2

$$P(V_3 = \text{phone A} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.27$$

$$P(V_3 = \text{phone B} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.24$$

$$P(V_3 = \text{rice} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.10$$

$$P(V_3 = \text{couscous} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.09$$

$$P(V_3 = \text{beer} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.30$$

What actually happened?

## Next Item Prediction, hold out $V_3$ : HR@K Recall@K Precision@K - K=2

$$P(V_3 = \text{phone A} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.27$$

$$P(V_3 = \text{phone B} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.24$$

$$P(V_3 = \text{rice} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.10$$

$$P(V_3 = \text{couscous} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.09$$

$$P(V_3 = \text{beer} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.30$$

What actually happened?  $V_3 = \text{phone B}$

## Next Item Prediction, hold out $V_3$ : HR@K Recall@K Precision@K - K=2

$$\begin{aligned} P(V_3 = \text{phone A} | V_1 = \text{beer}, V_2 = \text{phone A}) &= 0.27 \\ P(V_3 = \text{phone B} | V_1 = \text{beer}, V_2 = \text{phone A}) &= 0.24 \\ P(V_3 = \text{rice} | V_1 = \text{beer}, V_2 = \text{phone A}) &= 0.10 \\ P(V_3 = \text{couscous} | V_1 = \text{beer}, V_2 = \text{phone A}) &= 0.09 \\ P(V_3 = \text{beer} | V_1 = \text{beer}, V_2 = \text{phone A}) &= 0.30 \end{aligned}$$

What actually happened?  $V_3 = \text{phone B}$

Our model assigned:

$P(V_3 = \text{phone B} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.24$ , which is 3rd place

How good is 3rd place?

## Next Item Prediction, hold out $V_3$ : HR@K Recall@K Precision@K - K=2

$$\begin{aligned} P(V_3 = \text{phone A} | V_1 = \text{beer}, V_2 = \text{phone A}) &= 0.27 \\ P(V_3 = \text{phone B} | V_1 = \text{beer}, V_2 = \text{phone A}) &= 0.24 \\ P(V_3 = \text{rice} | V_1 = \text{beer}, V_2 = \text{phone A}) &= 0.10 \\ P(V_3 = \text{couscous} | V_1 = \text{beer}, V_2 = \text{phone A}) &= 0.09 \\ P(V_3 = \text{beer} | V_1 = \text{beer}, V_2 = \text{phone A}) &= 0.30 \end{aligned}$$

What actually happened?  $V_3 = \text{phone B}$

Our model assigned:

$P(V_3 = \text{phone B} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.24$ , which is 3rd place

How good is 3rd place? HR@K if K=2, our model assigned phone B the third highest spot so it just misses out.

## Next Item Prediction, hold out $V_3$ : HR@K Recall@K Precision@K - K=3

$$P(V_3 = \text{phone A} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.27$$

$$P(V_3 = \text{phone B} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.24$$

$$P(V_3 = \text{rice} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.10$$

$$P(V_3 = \text{couscous} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.09$$

$$P(V_3 = \text{beer} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.30$$

What actually happened?

## Next Item Prediction, hold out $V_3$ : HR@K Recall@K Precision@K - K=3

$$P(V_3 = \text{phone A} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.27$$

$$P(V_3 = \text{phone B} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.24$$

$$P(V_3 = \text{rice} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.10$$

$$P(V_3 = \text{couscous} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.09$$

$$P(V_3 = \text{beer} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.30$$

What actually happened?  $V_3 = \text{phone B}$

## Next Item Prediction, hold out $V_3$ : HR@K Recall@K Precision@K - K=3

$$P(V_3 = \text{phone A} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.27$$

$$P(V_3 = \text{phone B} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.24$$

$$P(V_3 = \text{rice} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.10$$

$$P(V_3 = \text{couscous} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.09$$

$$P(V_3 = \text{beer} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.30$$

What actually happened?  $V_3 = \text{phone B}$

Our model assigned:

$P(V_3 = \text{phone B} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.24$ , which is 3rd place

How good is 3rd place?

## Next Item Prediction, hold out $V_3$ : HR@K Recall@K Precision@K - K=3

$P(V_3 = \text{phone A}   V_1 = \text{beer}, V_2 = \text{phone A})$	= 0.27
$P(V_3 = \text{phone B}   V_1 = \text{beer}, V_2 = \text{phone A})$	= 0.24
$P(V_3 = \text{rice}   V_1 = \text{beer}, V_2 = \text{phone A})$	= 0.10
$P(V_3 = \text{couscous}   V_1 = \text{beer}, V_2 = \text{phone A})$	= 0.09
$P(V_3 = \text{beer}   V_1 = \text{beer}, V_2 = \text{phone A})$	= 0.30

What actually happened?  $V_3 = \text{phone B}$

Our model assigned:

$P(V_3 = \text{phone B} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.24$ , which is 3rd place

How good is 3rd place? HR@K if K=3, our model assigned phone B the third highest spot so it is in.

## Next Item Prediction, hold out $V_3$ : HR@K Recall@K Precision@K - K=3

$P(V_3 = \text{phone A}   V_1 = \text{beer}, V_2 = \text{phone A})$	= 0.27
$P(V_3 = \text{phone B}   V_1 = \text{beer}, V_2 = \text{phone A})$	= 0.24
$P(V_3 = \text{rice}   V_1 = \text{beer}, V_2 = \text{phone A})$	= 0.10
$P(V_3 = \text{couscous}   V_1 = \text{beer}, V_2 = \text{phone A})$	= 0.09
$P(V_3 = \text{beer}   V_1 = \text{beer}, V_2 = \text{phone A})$	= 0.30

What actually happened?  $V_3 = \text{phone B}$

Our model assigned:

$P(V_3 = \text{phone B} | V_1 = \text{beer}, V_2 = \text{phone A}) = 0.24$ , which is 3rd place

How good is 3rd place? HR@K if K=3, our model assigned phone B the third highest spot so it is in.

Rationale: you want a ranking that assigns high positions to events that occur. The position matters more than the probability.

# Real Reco

## Motivating Example



# Real Reco

## Motivating Example



Product view



Recommend



# Real Reco

## Motivating Example



Product view



Recommend



# The Real Reco Problem

## Motivating Example



$P(C_4 = 1 | A_4 = \text{couscous}, V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone B})$

# The Real Reco Problem

## Motivating Example



$$P(C_4 = 1 | A_4 = \text{phone B}, V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone B})$$

# The Real Reco Problem

## Motivating Example



$$P(C_4 = 1 | A_4 = \text{rice}, V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone B})$$

## What we have, and what we want:

Our model:

$$P(V_4 = \text{phone A} | V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone A}) = 0.27$$

$$P(V_4 = \text{phone B} | V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone A}) = 0.28$$

$$P(V_4 = \text{rice} | V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone A}) = .12$$

$$P(V_4 = \text{couscous} | V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone A}) = .14$$

$$P(V_4 = \text{beer} | V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone A}) = 0.19$$

## What we have, and what we want:

Our model:

$$P(V_4 = \text{phone A} | V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone A}) = 0.27$$

$$P(V_4 = \text{phone B} | V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone A}) = 0.28$$

$$P(V_4 = \text{rice} | V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone A}) = .12$$

$$P(V_4 = \text{couscous} | V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone A}) = .14$$

$$P(V_4 = \text{beer} | V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone A}) = 0.19$$

Which is a vector of size  $P$  that sums to 1.

We want:

## What we have, and what we want:

Our model:

$$P(V_4 = \text{phone A} | V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone A}) = 0.27$$

$$P(V_4 = \text{phone B} | V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone A}) = 0.28$$

$$P(V_4 = \text{rice} | V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone A}) = 0.12$$

$$P(V_4 = \text{couscous} | V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone A}) = 0.14$$

$$P(V_4 = \text{beer} | V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone A}) = 0.19$$

Which is a vector of size  $P$  that sums to 1.

We want:

$$P(C_4 = 1 | A_4 = \text{phone A}, V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone A}) = ?$$

$$P(C_4 = 1 | A_4 = \text{phone B}, V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone A}) = ?$$

$$P(C_4 = 1 | A_4 = \text{rice}, V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone A}) = ?$$

$$P(C_4 = 1 | A_4 = \text{couscous}, V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone A}) = ?$$

$$P(C_4 = 1 | A_4 = \text{beer}, V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone A}) = ?$$

## What we have, and what we want:

Our model:

$$P(V_4 = \text{phone A} | V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone A}) = 0.27$$

$$P(V_4 = \text{phone B} | V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone A}) = 0.28$$

$$P(V_4 = \text{rice} | V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone A}) = 0.12$$

$$P(V_4 = \text{couscous} | V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone A}) = 0.14$$

$$P(V_4 = \text{beer} | V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone A}) = 0.19$$

Which is a vector of size  $P$  that sums to 1.

We want:

$$P(C_4 = 1 | A_4 = \text{phone A}, V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone A}) = ?$$

$$P(C_4 = 1 | A_4 = \text{phone B}, V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone A}) = ?$$

$$P(C_4 = 1 | A_4 = \text{rice}, V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone A}) = ?$$

$$P(C_4 = 1 | A_4 = \text{couscous}, V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone A}) = ?$$

$$P(C_4 = 1 | A_4 = \text{beer}, V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone A}) = ?$$

Which is a vector of size  $P$  where each entry is between 0 and 1.

## An implicit assumption:

Our model predicts:

$$P(V_4 = \text{phone A} | V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone A}) = 0.27$$

$$P(V_4 = \text{phone B} | V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone A}) = 0.28$$

$$P(V_4 = \text{rice} | V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone A}) = .12$$

$$P(V_4 = \text{couscous} | V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone A}) = .14$$

$$P(V_4 = \text{beer} | V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone A}) = 0.19$$

So let's recommend phone B, it has the highest next item probability.

An implicit assumption:

If:

## An implicit assumption:

If:

$$P(V_n = a | V_1 = v_1 \dots V_{n-1} = v_{n-1}) > P(V_n = b | V_1 = v_1 \dots V_{n-1} = v_{n-1})$$

## An implicit assumption:

If:

$$P(V_n = a | V_1 = v_1..V_{n-1} = v_{n-1}) > P(V_n = b | V_1 = v_1..V_{n-1} = v_{n-1})$$

Assume:

$$\begin{aligned} P(C_n = 1 | A_n = a, V_1 = v_1..V_{n-1} = v_{n-1}) \\ > P(C_n = 1 | A_n = b, V_1 = v_1..V_{n-1} = v_{n-1}) \end{aligned}$$

## An implicit assumption:

If:

$$P(V_n = a | V_1 = v_1..V_{n-1} = v_{n-1}) > P(V_n = b | V_1 = v_1..V_{n-1} = v_{n-1})$$

Assume:

$$\begin{aligned} P(C_n = 1 | A_n = a, V_1 = v_1..V_{n-1} = v_{n-1}) \\ > P(C_n = 1 | A_n = b, V_1 = v_1..V_{n-1} = v_{n-1}) \end{aligned}$$

Vast amounts of academic Reco work assumes this implicitly!

## The assumption is testable

Our model predicts:

$$P(V_4 = \text{phone A} | V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone A}) = 0.27$$

$$P(V_4 = \text{phone B} | V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone A}) = 0.28$$

$$P(V_4 = \text{rice} | V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone A}) = .12$$

$$P(V_4 = \text{couscous} | V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone A}) = .14$$

$$P(V_4 = \text{beer} | V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone A}) = 0.19$$

So let's recommend phone B, it has the highest next item probability.

## The assumption is testable

Our model predicts:

$$P(V_4 = \text{phone A} | V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone A}) = 0.27$$

$$P(V_4 = \text{phone B} | V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone A}) = 0.28$$

$$P(V_4 = \text{rice} | V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone A}) = .12$$

$$P(V_4 = \text{couscous} | V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone A}) = .14$$

$$P(V_4 = \text{beer} | V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone A}) = 0.19$$

So let's recommend phone B, it has the highest next item probability. - but we will also sometimes choose other actions with a small probability, so we can check if these items have lower click through rates.

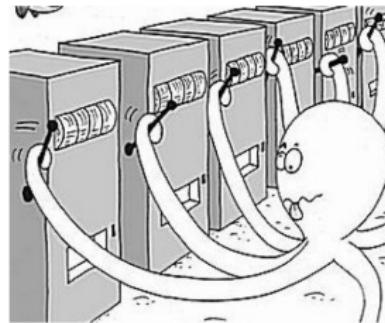
## History: Split History of Recommender Systems



Next Item or Missing Link Prediction  
e.g. Netflix Prize or Movie Lens

	movie 1	movie 2	movie 3	movie 4	movie 5	movie 6	movie 7	movie 8	movie 9	movie 10
user 1										
user 2	2	1	2							
user 3			3	3						
user 4	2			3		5	3		4	
user 5	4				5					
user 6		2								
user 7		2					4	2	3	
user 8	3	4			4					
user 9									3	
user 10		1		2						

Computational Advertising, e.g.  
Bandits, Counterfactual Risk  
Minimisation, Contextual Bandits,  
Reinforcement Learning.



# Classic RecSys Datasets: Are they logs of recommender systems?

- MovieLens:

# Classic RecSys Datasets: Are they logs of recommender systems?

- MovieLens: no, explicit feedback of movie ratings

# Classic RecSys Datasets: Are they logs of recommender systems?

- MovieLens: no, explicit feedback of movie ratings
- Netflix Prize:

# Classic RecSys Datasets: Are they logs of recommender systems?

- MovieLens: no, explicit feedback of movie ratings
- Netflix Prize: no, explicit feedback of movie ratings

# Classic RecSys Datasets: Are they logs of recommender systems?

- MovieLens: no, explicit feedback of movie ratings
- Netflix Prize: no, explicit feedback of movie ratings
- Yoochoose (RecSys 15):

# Classic RecSys Datasets: Are they logs of recommender systems?

- MovieLens: no, explicit feedback of movie ratings
- Netflix Prize: no, explicit feedback of movie ratings
- Yoochoose (RecSys 15): no, implicit session based behavior

# Classic RecSys Datasets: Are they logs of recommender systems?

- MovieLens: no, explicit feedback of movie ratings
- Netflix Prize: no, explicit feedback of movie ratings
- Yoochoose (RecSys 15): no, implicit session based behavior
- 30 Music:

# Classic RecSys Datasets: Are they logs of recommender systems?

- MovieLens: no, explicit feedback of movie ratings
- Netflix Prize: no, explicit feedback of movie ratings
- Yoochoose (RecSys 15): no, implicit session based behavior
- 30 Music: no, implicit session based behavior

# Classic RecSys Datasets: Are they logs of recommender systems?

- MovieLens: no, explicit feedback of movie ratings
- Netflix Prize: no, explicit feedback of movie ratings
- Yoochoose (RecSys 15): no, implicit session based behavior
- 30 Music: no, implicit session based behavior
- Yahoo News Feed Dataset

# Classic RecSys Datasets: Are they logs of recommender systems?

- MovieLens: no, explicit feedback of movie ratings
- Netflix Prize: no, explicit feedback of movie ratings
- Yoochoose (RecSys 15): no, implicit session based behavior
- 30 Music: no, implicit session based behavior
- Yahoo News Feed Dataset yes!
- Criteo Dataset for counterfactual evaluation of RecSys algorithms:

# Classic RecSys Datasets: Are they logs of recommender systems?

- MovieLens: no, explicit feedback of movie ratings
- Netflix Prize: no, explicit feedback of movie ratings
- Yoochoose (RecSys 15): no, implicit session based behavior
- 30 Music: no, implicit session based behavior
- Yahoo News Feed Dataset yes!
- Criteo Dataset for counterfactual evaluation of RecSys algorithms: yes!

# Classic RecSys Datasets: Are they logs of recommender systems?

- MovieLens: no, explicit feedback of movie ratings
- Netflix Prize: no, explicit feedback of movie ratings
- Yoochoose (RecSys 15): no, implicit session based behavior
- 30 Music: no, implicit session based behavior
- Yahoo News Feed Dataset yes!
- Criteo Dataset for counterfactual evaluation of RecSys algorithms: yes!

The Criteo Dataset shows a log of recommendations and if they were successful in getting users to click.

# Classic RecSys Evaluation Metrics: Do they evaluate the quality of recommendation?

- Recall@K, Precision@K, HR@K:

# Classic RecSys Evaluation Metrics: Do they evaluate the quality of recommendation?

- Recall@K, Precision@K, HR@K: How often is an item in the top k

# Classic RecSys Evaluation Metrics: Do they evaluate the quality of recommendation?

- Recall@K, Precision@K, HR@K: How often is an item in the top k no, evaluates next item prediction

# Classic RecSys Evaluation Metrics: Do they evaluate the quality of recommendation?

- Recall@K, Precision@K, HR@K: How often is an item in the top k no, evaluates next item prediction
- DCG:

# Classic RecSys Evaluation Metrics: Do they evaluate the quality of recommendation?

- Recall@K, Precision@K, HR@K: How often is an item in the top k no, evaluates next item prediction
- DCG: Are we assigning high score to an item

# Classic RecSys Evaluation Metrics: Do they evaluate the quality of recommendation?

- Recall@K, Precision@K, HR@K: How often is an item in the top k no, evaluates next item prediction
- DCG: Are we assigning high score to an item no, evaluates next item prediction

# Classic RecSys Evaluation Metrics: Do they evaluate the quality of recommendation?

- Recall@K, Precision@K, HR@K: How often is an item in the top k no, evaluates next item prediction
- DCG: Are we assigning high score to an item no, evaluates next item prediction
- Log likelihood:

# Classic RecSys Evaluation Metrics: Do they evaluate the quality of recommendation?

- Recall@K, Precision@K, HR@K: How often is an item in the top k no, evaluates next item prediction
- DCG: Are we assigning high score to an item no, evaluates next item prediction
- Log likelihood: unclear, but usually next item prediction

# Classic RecSys Evaluation Metrics: Do they evaluate the quality of recommendation?

- Recall@K, Precision@K, HR@K: How often is an item in the top k no, evaluates next item prediction
- DCG: Are we assigning high score to an item no, evaluates next item prediction
- Log likelihood: unclear, but usually next item prediction
- Inverse Propensity Score estimate of click through rate:

# Classic RecSys Evaluation Metrics: Do they evaluate the quality of recommendation?

- Recall@K, Precision@K, HR@K: How often is an item in the top k no, evaluates next item prediction
- DCG: Are we assigning high score to an item no, evaluates next item prediction
- Log likelihood: unclear, but usually next item prediction
- Inverse Propensity Score estimate of click through rate: to be explained later.

# Classic RecSys Evaluation Metrics: Do they evaluate the quality of recommendation?

- Recall@K, Precision@K, HR@K: How often is an item in the top k no, evaluates next item prediction
- DCG: Are we assigning high score to an item no, evaluates next item prediction
- Log likelihood: unclear, but usually next item prediction
- Inverse Propensity Score estimate of click through rate: to be explained later. yes! (although it is often noisy)

# Classic RecSys Evaluation Metrics: Do they evaluate the quality of recommendation?

- Recall@K, Precision@K, HR@K: How often is an item in the top k no, evaluates next item prediction
- DCG: Are we assigning high score to an item no, evaluates next item prediction
- Log likelihood: unclear, but usually next item prediction
- Inverse Propensity Score estimate of click through rate: to be explained later. yes! (although it is often noisy)
- AB Test:

# Classic RecSys Evaluation Metrics: Do they evaluate the quality of recommendation?

- Recall@K, Precision@K, HR@K: How often is an item in the top k no, evaluates next item prediction
- DCG: Are we assigning high score to an item no, evaluates next item prediction
- Log likelihood: unclear, but usually next item prediction
- Inverse Propensity Score estimate of click through rate: to be explained later. yes! (although it is often noisy)
- AB Test: i.e. run a randomized control trial live

# Classic RecSys Evaluation Metrics: Do they evaluate the quality of recommendation?

- Recall@K, Precision@K, HR@K: How often is an item in the top k no, evaluates next item prediction
- DCG: Are we assigning high score to an item no, evaluates next item prediction
- Log likelihood: unclear, but usually next item prediction
- Inverse Propensity Score estimate of click through rate: to be explained later. yes! (although it is often noisy)
- AB Test: i.e. run a randomized control trial liveyes, but the academic literature has no access to this

# Classic RecSys Evaluation Metrics: Do they evaluate the quality of recommendation?

- Recall@K, Precision@K, HR@K: How often is an item in the top k no, evaluates next item prediction
- DCG: Are we assigning high score to an item no, evaluates next item prediction
- Log likelihood: unclear, but usually next item prediction
- Inverse Propensity Score estimate of click through rate: to be explained later. yes! (although it is often noisy)
- AB Test: i.e. run a randomized control trial liveyes, but the academic literature has no access to this

If the dataset does not contain a log of recommendations and if they were successful, you cannot compute metrics of the recommendation quality.

## Offline evaluation that predicts an AB test result

Imagine we have a new recommendation strategy

$\pi_t(A_n = a_n | V_1 = v_1, \dots, V_{n-1} = v_{n-1})$ , can we predict how well it will perform if we deploy it? We collect logs from a different policy:

$\pi_0(A_n = a_n | V_1 = v_1, \dots, V_{n-1} = v_{n-1})$

## Offline evaluation that predicts an AB test result

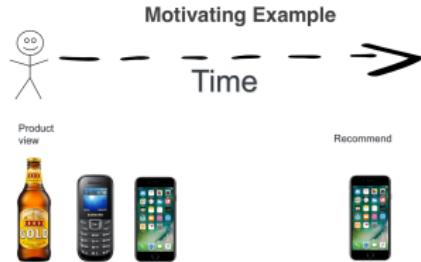
Imagine we have a new recommendation strategy

$\pi_t(A_n = a_n | V_1 = v_1, \dots, V_{n-1} = v_{n-1})$ , can we predict how well it will perform if we deploy it? We collect logs from a different policy:

$\pi_0(A_n = a_n | V_1 = v_1, \dots, V_{n-1} = v_{n-1})$

Let's examine some hypothetical logs...

# Offline evaluation IPS



Imagine the user clicks

$$\pi_t(A_4 = \text{phone B} | V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone B}) = 1$$

$$\pi_0(A_4 = \text{phone B} | V_1 = \text{beer}, V_2 = \text{phone A}, V_3 = \text{phone B}) = 0.8$$

The new policy will recommend “phone B” the  $\frac{1}{0.8} = 1.25 \times$  as often as the old policy.

## Offline evaluation IPS

Imagine the user clicks

$$\pi_t(A_2 = \text{rice} | V_1 = \text{couscous}) = 1$$

$$\pi_0(A_2 = \text{rice} | V_1 = \text{couscous}) = 0.01$$

The new policy will recommend “rice” the  $\frac{1}{0.01} = 100\times$  as often as the old policy.

## Offline evaluation IPS

$$\text{CTR estimate} = \frac{1}{N} \sum_n^N \frac{c_n \pi_t(A_n = a_n | V_1 = v_1..V_n = v_n)}{\pi_0(A_n = a_n | V_1 = v_1..V_n = v_n)}$$

## Offline evaluation IPS

$$\text{CTR estimate} = \frac{1}{N} \sum_n^N \frac{c_n \pi_t(A_n = a_n | V_1 = v_1..V_n = v_n)}{\pi_0(A_n = a_n | V_1 = v_1..V_n = v_n)}$$

We only look at the clicks i.e.  $c_n = 1$  (otherwise the contribution is 0).

## Offline evaluation IPS

$$\text{CTR estimate} = \frac{1}{N} \sum_n^N \frac{c_n \pi_t(A_n = a_n | V_1 = v_1..V_n = v_n)}{\pi_0(A_n = a_n | V_1 = v_1..V_n = v_n)}$$

We only look at the clicks i.e.  $c_n = 1$  (otherwise the contribution is 0).

When the new policy differs markedly from the old, the weights become very high (to compensate for the fact that these are rare examples in your sample). These rare high values contribute a lot to the variance of the estimator.

## Offline evaluation IPS

$$\text{CTR estimate} = \frac{1}{N} \sum_n^N \frac{c_n \pi_t(A_n = a_n | V_1 = v_1..V_n = v_n)}{\pi_0(A_n = a_n | V_1 = v_1..V_n = v_n)}$$

We only look at the clicks i.e.  $c_n = 1$  (otherwise the contribution is 0).

When the new policy differs markedly from the old, the weights become very high (to compensate for the fact that these are rare examples in your sample). These rare high values contribute a lot to the variance of the estimator.

IPS actually attempts to predict the AB test, but it often has less than spectacular results...

## Real world Reco

Real world reco is an anonymous hybrid of different methodologies.  
Often poor agreement between online and offline metrics are reported..

## Real world Reco

Real world reco is an anonymous hybrid of different methodologies.

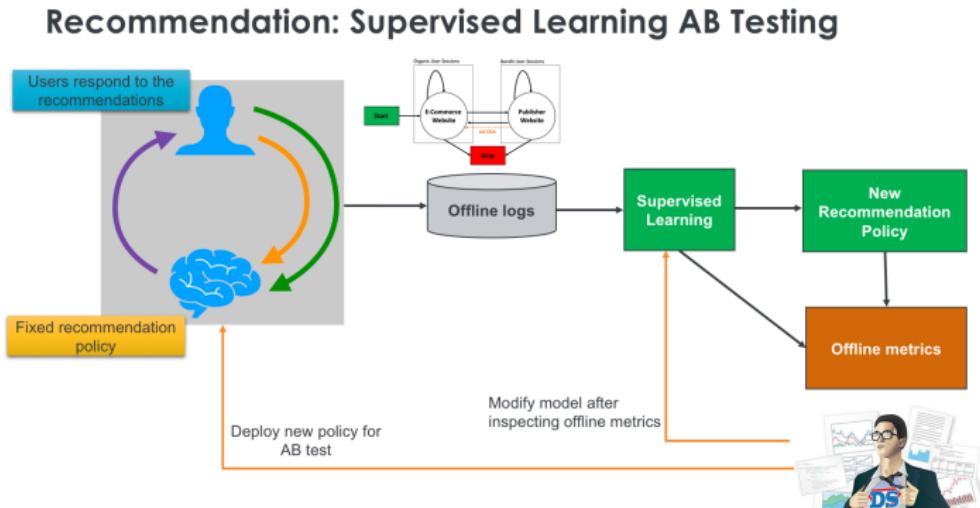
Often poor agreement between online and offline metrics are reported..  
why?

## Real world Reco

Real world reco is an anonymous hybrid of different methodologies.

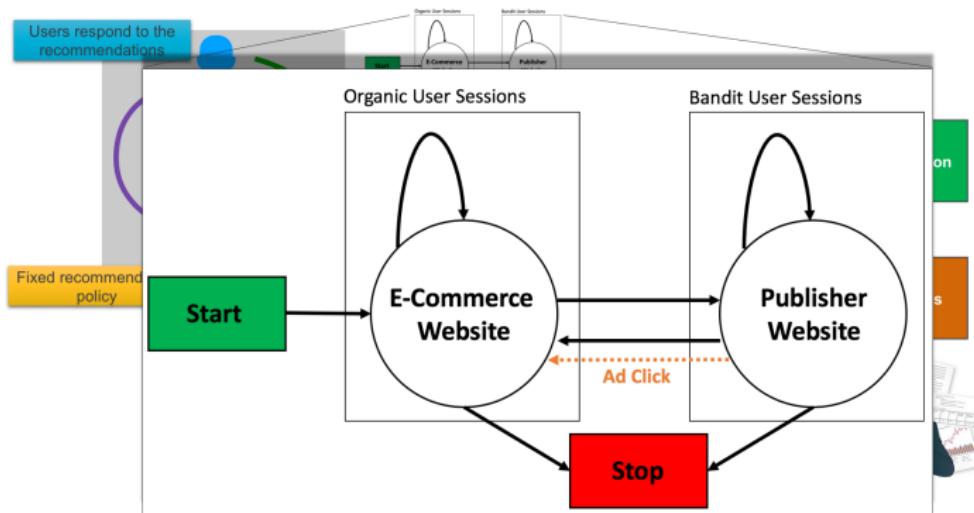
Often poor agreement between online and offline metrics are reported..  
why? What does real world reco look like?

# Recommendation as Supervised Learning and AB Testing



# Recommendation as Supervised Learning and AB Testing

## Recommendation: Supervised Learning AB Testing



Let's take a step back...

## How are we improving large-scale Recommender Systems in the Real World

- Learn a supervised model from past user activity
- Evaluate offline and decide whether to A/B test
- A/B test
- If positive and scalable, roll-out
- If not, try to understand what happened and try to create a better model of the world using the same past data
- Repeat

## Supervised learning with the wrong objective

We are operating under the assumption that the best recommendation policy is in some sense the **optimal auto-complete of natural user behavior**

## Supervised learning with the wrong objective

From the point of view of business metrics, **learning to autocomplete behavior is a great initial recommendation policy**, especially when no prior user feedback is available.

# Supervised learning with the wrong objective

From the point of view of business metrics, **learning to autocomplete behavior is a great initial recommendation policy**, especially when no prior user feedback is available.

**However**, after a first golden age, where all offline improvements turn into positive A/B tests, the *naive recommendation* optimization objective and the business objective will start to diverge.

## Aligning the Recommendation objective with the business

- Of course, we could start incorporating user feedback that we collected while running the initial recommendation policies
- We should be able to continue bringing improvements using feedback that is now aligned with our business metrics (ad CTR, post click sales, dwell time, number of videos watched, ...)

# Modern Recommendation Systems Research

**Q: How does the literature on Large Scale Recommendation Systems look right now?**

A: Most of the latest publications are talking about ways of using Deep Learning for Recommendation:

- **Matrix Factorization extensions:** Word2Vec, Deep and Wide, Neural MF, Node2Vec
- **Content-based recommendation:** CNNs for Image, Text, Sounds to compute item similarities
- **Next event prediction / user activity modeling:** RNNs, TCNs

# Modern Recommendation Systems Research

We see great improvements in offline metrics!

- **Regression/Classification metrics:** MSE, AUC, NLL
- **Ranking metrics:** MPR, Precision@k, NDCG

# What is wrong with this picture?

**In the same time, as practitioners, we see difficulties in improving the Real World Recommendation models!**

Offline - online metrics alignment for recommendation is a recognized problem:

- **Previous work:** Missing Not At Random (MNAR) framework for Matrix factorization, Bandits Literature
- **New avenues:** *REVEAL: Offline evaluation for recommender systems Workshop at RecSys 2019* (this September in Copenhagen)  
<https://recsys.acm.org/recsys19/reveal/>

# The relationship between Reinforcement Learning and Recommendation

- We are doing Reinforcement Learning by hand!

# The relationship between Reinforcement Learning and Recommendation

- We are doing Reinforcement Learning by hand!
- Furthermore, we are trying to do RL using the Supervised Learning Framework

# The relationship between Reinforcement Learning and Recommendation

- We are doing Reinforcement Learning by hand!
- Furthermore, we are trying to do RL using the Supervised Learning Framework
- Standard test data sets do not let us explore this aspect of recommender systems

# The relationship between Reinforcement Learning and Recommendation

- We are doing Reinforcement Learning by hand!
  - Furthermore, we are trying to do RL using the Supervised Learning Framework
  - Standard test data sets do not let us explore this aspect of recommender systems
- .. but how do you evaluate a reinforcement learning algorithm?

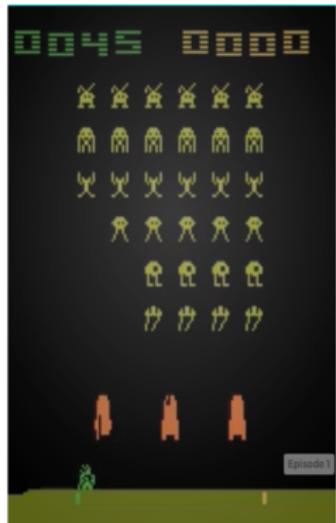
# Open AI Gym



**Problem:** The reinforcement learning community lacked a common set of tools to allow comparison of reinforcement learning algorithms.

**Open AI Gym: 2016 (Brockman et al.):** Software standard (Python api) that allows comparison of the performance of reinforcement learning algorithms.

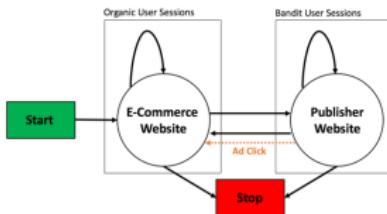
**Core idea:** environments (the problem) and agents (the RL algorithm) interact.



# Introducing RecoGym



**RecoGym:** Simulates User Behavior (both Organic and Bandit)



- Allows online evaluation of new recommendation policies in a simulated environment
- Holistic view of user (organic and bandit) provides a framework for categorizing recommendation algorithms
- Key feature: adjustable parameter that changes the effectiveness of "Pure Organic" or next item prediction algorithms.

11 •

**Algorithm 1:** A simple Simulator

---

**Input :**  $S \in \mathcal{R}^{3 \times 3}$  transition matrix between organic and bandit,  $\Gamma \in \mathcal{R}^{P \times K}$  organic embeddings,  $\mu_\Gamma$  organic popularity  $\beta \in \mathcal{R}^{P \times K}$  bandit embeddings,  $\mu_\beta$  non-personalised ctr contribution  $f(\cdot)$  monotonic increasing function accounting for ad fatigue  $m$ ,  $U$  number of users,  $P$  number of products.

**Output:** Sequence of organic and bandit events

```

1 for  $u \in 1..U$  do
2    $t \leftarrow 0$ 
3    $z_{u,0} \leftarrow$  organic
4    $r_{u,0} \leftarrow$  undef
5    $c_{u,0} \leftarrow$  undef
6    $\omega_{u,0} \sim N(0_{K \times 1}, I_K)$ 
7    $v_{u,0} \sim \text{Categorical}(\text{softmax}(\Gamma \omega_{u,0}))$ 
8   while  $z_{u,t} \neq \text{stop}$  do
9      $t \leftarrow t + 1$ 
10     $\omega_{u,t} \sim N(\omega_{u,t-1}, \sigma_w^2 I_K)$ 
11    if  $c_{u,t-1} = 1$  then
12       $z_{u,t} \sim \text{Categorical}(S_{z_{u,t-1}, \text{organic}}, S_{z_{u,t-1}, \text{bandit}}, S_{z_{u,t-1}, \text{stop}})$ 
13    else
14       $z_{u,t} = \text{organic}$ 
15    end
16    if  $z_{u,t} = \text{organic}$  then
17       $v_{u,t} \sim \text{Categorical}(\text{softmax}(\Gamma \omega_{u,t} + \mu_\Gamma))$ 
18       $r_{u,t} \leftarrow$  undef
19       $c_{u,t} \leftarrow$  undef
20    end
21    if  $z_{u,t} = \text{bandit}$  then
22       $r_{u,t}$  is generated from the policy
23       $c_{u,t} \sim \text{Bernoulli}([f(\beta \omega_{u,t} + \mu_\beta)] r_{u,t})$ 
24    end
25  end
26 end

```

---



$$v_0 \sim \text{categorical} \left( \text{softmax} \left( \begin{pmatrix} \Gamma_{1,1} & \dots & \Gamma_{1,K} \\ \vdots & \ddots & \vdots \\ \Gamma_{P,1} & \dots & \Gamma_{P,K} \end{pmatrix} \begin{pmatrix} \omega_{0,1} \\ \vdots \\ \omega_{0,K} \end{pmatrix} + \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_P \end{pmatrix} \right) \right)$$

$$v_1 \sim \text{categorical} \left( \text{softmax} \left( \begin{pmatrix} \Gamma_{1,1} & \dots & \Gamma_{1,K} \\ \vdots & \ddots & \vdots \\ \Gamma_{P,1} & \dots & \Gamma_{P,K} \end{pmatrix} \begin{pmatrix} \omega_{1,1} \\ \vdots \\ \omega_{1,K} \end{pmatrix} + \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_P \end{pmatrix} \right) \right)$$

$$v_2 \sim \text{categorical} \left( \text{softmax} \left( \begin{pmatrix} \Gamma_{1,1} & \dots & \Gamma_{1,K} \\ \vdots & \ddots & \vdots \\ \Gamma_{P,1} & \dots & \Gamma_{P,K} \end{pmatrix} \begin{pmatrix} \omega_{2,1} \\ \vdots \\ \omega_{2,K} \end{pmatrix} + \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_P \end{pmatrix} \right) \right)$$

$f$  is an increasing function

$$f : \mathbb{R} \rightarrow (0, 1)$$

$$c_3|r_3 \sim \text{Bernoulli} \left( f \left( \begin{pmatrix} \beta_{1,1} & \dots & \beta_{1,K} \\ \vdots & \ddots & \vdots \\ \beta_{P,1} & \dots & \beta_{P,K} \end{pmatrix} \begin{pmatrix} \omega_{3,1} \\ \vdots \\ \omega_{3,K} \end{pmatrix} + \begin{pmatrix} \mu_1^* \\ \vdots \\ \mu_P^* \end{pmatrix} \right) \Big| r_3 \right)$$

$$c_4|r_4 \sim \text{Bernoulli} \left( f \left( \begin{pmatrix} \beta_{1,1} & \dots & \beta_{1,K} \\ \vdots & \ddots & \vdots \\ \beta_{P,1} & \dots & \beta_{P,K} \end{pmatrix} \begin{pmatrix} \omega_{4,1} \\ \vdots \\ \omega_{4,K} \end{pmatrix} + \begin{pmatrix} \mu_1^* \\ \vdots \\ \mu_P^* \end{pmatrix} \right) \Big| r_4 \right)$$

$$v_5 \sim \text{categorical} \left( \text{softmax} \left( \begin{pmatrix} \Gamma_{1,1} & \dots & \Gamma_{1,K} \\ \vdots & \ddots & \vdots \\ \Gamma_{P,1} & \dots & \Gamma_{P,K} \end{pmatrix} \begin{pmatrix} \omega_{5,1} \\ \vdots \\ \omega_{5,K} \end{pmatrix} + \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_P \end{pmatrix} \right) \right)$$

time



criteo

# The RecoGym Session

```
In [1]: import gym, reco_gym

# env_0_args is a dictionary of default parameters (i.e. number of products)
from reco_gym import env_1_args, Configuration

# You can overwrite environment arguments here:
env_1_args['random_seed'] = 41

# Initialize the gym for the first time by calling .make() and .init_gym()
env = gym.make('reco-gym-v1')
env.init_gym(env_1_args)

env.reset() # we call request to move to the first user
```

```
In [2]: observation, reward, done, info = env.step(None)
# We specify None because we have need to learn about the user before we can act
```

```
In [3]: observation.current_sessions
```

```
Out[3]: [{t': 0, 'u': 0, 'z': 'pageview', 'v': 4},
          {'t': 1, 'u': 0, 'z': 'pageview', 'v': 4},
          {'t': 2, 'u': 0, 'z': 'pageview', 'v': 4},
          {'t': 3, 'u': 0, 'z': 'pageview', 'v': 0}]
```

```
In [4]: # We see the user is interested in product id 4 (they viewed it 3 times) and product id 0 (they viewed it once)
```

# The RecoGym Session

```
In [5]: observation, reward, done, info = env.step(0)

In [6]: reward # the user does not click on our recommendation of product 0
Out[6]: 0

In [7]: observation.current_sessions # the user does not view any more products, so we learn nothing more about them
Out[7]: []
```

# The RecoGym Session

```
In [8]: observation, reward, done, info = env.step(0) # we recommend product 0 again
In [9]: reward # they do not click again, ctr are usually low - this is not surprising
Out[9]: 0
In [10]: observation.current_sessions # again no additional product views the bandit session continues
Out[10]: []
```

# The RecoGym Session

```
In [11]: observation, reward, done, info = env.step(4) # we now recommend product 4
In [12]: reward # they clicked! We must have done something right on that last recommendation
Out[12]: 1
In [13]: observation.current_sessions # the user moved to the retailer website and viewed the product
Out[13]: [{t: 7, u: 0, z: 'pageview', v: 4}]
```

# Unlike RL we continue to use logs

Let's start with a random logging policy (a crazy thing to do, but a useful theoretical concept).

```
In [17]: env.generate_logs(100)
```

```
Out[17]:
```

	a	c	ps	ps-a	t	u	v	z
0	NaN	NaN	NaN	None	0	0	2.0	organic
1	NaN	NaN	NaN	None	1	0	4.0	organic
2	NaN	NaN	NaN	None	2	0	4.0	organic
3	NaN	NaN	NaN	None	3	0	4.0	organic
4	NaN	NaN	NaN	None	4	0	2.0	organic
5	NaN	NaN	NaN	None	5	0	4.0	organic
6	NaN	NaN	NaN	None	6	0	5.0	organic
7	NaN	NaN	NaN	None	7	0	4.0	organic
8	NaN	NaN	NaN	None	8	0	4.0	organic
9	NaN	NaN	NaN	None	9	0	4.0	organic
10	3.0	0.0	0.1	[0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, ...	10	0	NaN	bandit
11	1.0	0.0	0.1	[0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, ...	11	0	NaN	bandit
12	2.0	0.0	0.1	[0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, ...	12	0	NaN	bandit
13	1.0	0.0	0.1	[0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, ...	13	0	NaN	bandit
14	8.0	0.0	0.1	[0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, ...	14	0	NaN	bandit
15	2.0	0.0	0.1	[0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, ...	15	0	NaN	bandit
16	5.0	0.0	0.1	[0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, ...	16	0	NaN	bandit
17	1.0	0.0	0.1	[0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, ...	17	0	NaN	bandit
18	7.0	0.0	0.1	[0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, ...	18	0	NaN	bandit
19	7.0	0.0	0.1	[0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, ...	19	0	NaN	bandit
20	5.0	0.0	0.1	[0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, ...	20	0	NaN	bandit

# Organic Best Of vs Bandit Best Of

## Exercise - Getting Started with RecoGym

Go to the notebook “Module III A.ipynb”

- Examine the logs, to start with we will look at non-personalized behavior only.
- Plot a histogram of organic product popularity. What is the most popular product?
- Plot the (non-personalized) click through rate of each product (with an error analysis). What is the most clicked-on product?
- Plot popularity against click through rate. How good a proxy is popularity for click through rate?
- Simulate an AB test comparing a simple recommender system (agent) that always recommends the highest ctr product with a recommender system that always recommends the organically most popular product.

## Harder questions

Reflect on how the logging policy would affect these results.

## Harder questions

Reflect on how the logging policy would affect these results.

What is personalization? What impact does personalization bring?

## Harder questions

Reflect on how the logging policy would affect these results.

What is personalization? What impact does personalization bring?

Where does traditional academic recommender systems research fit into this picture?

Evaluate an organic agent using the  
Bandit Signal

## The cold start scenario

- You need to build a recommender system for a client, they have an existing system with existing organic traffic.

## The cold start scenario

- You need to build a recommender system for a client, they have an existing system with existing organic traffic. This means you can build standard next item predictions to build models of user behavior and understand how user's preferences cluster together.

## The cold start scenario

- You need to build a recommender system for a client, they have an existing system with existing organic traffic. This means you can build standard next item predictions to build models of user behavior and understand how user's preferences cluster together.
- They are currently running a recommender system with a random policy.

## The cold start scenario

- You need to build a recommender system for a client, they have an existing system with existing organic traffic. This means you can build standard next item predictions to build models of user behavior and understand how user's preferences cluster together.
- They are currently running a recommender system with a random policy.
- Your task is to build a next item prediction model as a proxy for a recommendation algorithm. Then produce offline metrics to convince the engineering team to run an AB test.

## The cold start scenario

- You need to build a recommender system for a client, they have an existing system with existing organic traffic. This means you can build standard next item predictions to build models of user behavior and understand how user's preferences cluster together.
- They are currently running a recommender system with a random policy.
- Your task is to build a next item prediction model as a proxy for a recommendation algorithm. Then produce offline metrics to convince the engineering team to run an AB test.
- Finally you evaluate your performance against production.

## The cold start scenario

- You need to build a recommender system for a client, they have an existing system with existing organic traffic. This means you can build standard next item predictions to build models of user behavior and understand how user's preferences cluster together.
- They are currently running a recommender system with a random policy.
- Your task is to build a next item prediction model as a proxy for a recommendation algorithm. Then produce offline metrics to convince the engineering team to run an AB test.
- Finally you evaluate your performance against production.

Please look at notebook “Module III B.ipynb”

# Likelihood Based Agents

## Likelihood Based Agent

$$c_n \sim \text{Bernoulli} \left( \sigma \left( \Phi([\mathbf{X}_n \ \mathbf{a}_n])^T \boldsymbol{\beta} \right) \right)$$

where

- $\boldsymbol{\beta}$  are the parameters;
- $\sigma(\cdot)$  is the logistic sigmoid;
- $\Phi(\cdot)$  is a function that maps  $\mathbf{X}, \mathbf{a}$  to a higher dimensional space and includes some interaction terms between  $\mathbf{X}_n$  and  $\mathbf{a}_n$ .

## Likelihood Based Agent

$$c_n \sim \text{Bernoulli} \left( \sigma \left( \Phi([\mathbf{X}_n \ \mathbf{a}_n])^T \boldsymbol{\beta} \right) \right)$$

where

- $\boldsymbol{\beta}$  are the parameters;
- $\sigma(\cdot)$  is the logistic sigmoid;
- $\Phi(\cdot)$  is a function that maps  $\mathbf{X}, \mathbf{a}$  to a higher dimensional space and includes some interaction terms between  $\mathbf{X}_n$  and  $\mathbf{a}_n$ . *Why?*

## Likelihood Based Agent

Or using  $\Phi([\mathbf{X}_n \ \mathbf{a}_n]) = \mathbf{X}_n \otimes \mathbf{a}_n$ :

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{\text{lh}} = & \operatorname{argmax}_{\boldsymbol{\beta}} \sum_n c_n \log \sigma \left( (\mathbf{X}_n \otimes \mathbf{a}_n)^T \boldsymbol{\beta} \right) \\ & + (1 - c_n) \log \left( 1 - \sigma \left( (\mathbf{X}_n \otimes \mathbf{a}_n)^T \boldsymbol{\beta} \right) \right)\end{aligned}$$

## Feature Engineering to get the context vector $X$

The above model requires us to specify the context  $X$ , how can we do this?

# Feature Engineering

a	c	ps	ps-a	t	u	z		
0	NaN	NaN	NaN	None	0	0.0	organic	
1	3.0	0.0	0.1	[0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, ...	1	NaN	band1	
2	4.0	0.0	0.1	[0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, ...	2	NaN	band1	
3	5.0	0.0	0.1	[0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, ...	3	NaN	band1	
4	NaN	NaN	NaN	None	0	1	1.0	organic
5	2.0	0.0	0.1	[0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, ...	1	NaN	band1	
6	8.0	0.0	0.1	[0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, ...	2	NaN	band1	
7	4.0	0.0	0.1	[0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, ...	3	NaN	band1	
8	NaN	NaN	NaN	None	4	1	6.0	organic
9	NaN	NaN	NaN	None	5	1	6.0	organic
10	NaN	NaN	NaN	None	6	1	4.0	organic
11	NaN	NaN	NaN	None	7	1	6.0	organic
12	NaN	NaN	NaN	None	8	1	2.0	organic
13	NaN	NaN	NaN	None	9	1	6.0	organic
14	NaN	NaN	NaN	None	10	1	2.0	organic
15	NaN	NaN	NaN	None	11	1	1.0	organic
16	NaN	NaN	NaN	None	12	1	6.0	organic
17	NaN	NaN	NaN	None	13	1	6.0	organic
18	NaN	NaN	NaN	None	14	1	4.0	organic
19	NaN	NaN	NaN	None	15	1	1.0	organic
20	NaN	NaN	NaN	None	16	1	1.0	organic
21	NaN	NaN	NaN	None	17	1	1.0	organic
22	NaN	NaN	NaN	None	18	1	6.0	organic
23	NaN	NaN	NaN	None	19	1	6.0	organic
24	NaN	NaN	NaN	None	20	1	1.0	organic
25	NaN	NaN	NaN	None	21	1	6.0	organic
26	3.0	0.0	0.1	[0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, ...	22	NaN	band1	

```
history[0:8]
```

```
[array([1., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.]),  
 array([1., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.]),  
 array([1., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.]),  
 array([0., 1., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.]),  
 array([0., 1., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.]),  
 array([0., 1., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.]),  
 array([0., 1., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.]),  
 array([0., 6., 2., 0., 2., 0., 9., 0., 0., 0., 0., 0.]),  
 array([0., 6., 2., 0., 2., 0., 9., 0., 0., 0., 0., 0.])]
```

```
actions[0:8]
```

```
[array([0, 0, 0, 1, 0, 0, 0, 0, 0, 0], dtype=int8),  
 array([0, 0, 0, 1, 0, 0, 0, 0, 0, 0], dtype=int8),  
 array([0, 0, 0, 0, 1, 0, 0, 0, 0, 0], dtype=int8),  
 array([0, 1, 0, 0, 0, 0, 0, 0, 0, 0], dtype=int8),  
 array([0, 0, 0, 0, 0, 0, 0, 0, 1, 0], dtype=int8),  
 array([0, 0, 0, 1, 0, 0, 0, 0, 0, 0], dtype=int8),  
 array([0, 0, 1, 0, 0, 0, 0, 0, 0, 0], dtype=int8),  
 array([1, 0, 0, 0, 0, 0, 0, 0, 0, 0], dtype=int8)]
```

# Feature Engineering

a	c	ps	ps-a	t	u	z		
0	NaN	NaN	NaN	None	0	0.0	organic	
1	3.0	0.0	0.1	[0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, ...	1	NaN	band1	
2	4.0	0.0	0.1	[0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, ...	2	NaN	band1	
3	5.0	0.0	0.1	[0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, ...	3	NaN	band1	
4	NaN	NaN	NaN	None	0	1	1.0	organic
5	2.0	0.0	0.1	[0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, ...	1	NaN	band1	
6	8.0	0.0	0.1	[0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, ...	2	NaN	band1	
7	4.0	0.0	0.1	[0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, ...	3	NaN	band1	
8	NaN	NaN	NaN	None	4	1	6.0	organic
9	NaN	NaN	NaN	None	5	1	6.0	organic
10	NaN	NaN	NaN	None	6	1	4.0	organic
11	NaN	NaN	NaN	None	7	1	6.0	organic
12	NaN	NaN	NaN	None	8	1	2.0	organic
13	NaN	NaN	NaN	None	9	1	6.0	organic
14	NaN	NaN	NaN	None	10	1	2.0	organic
15	NaN	NaN	NaN	None	11	1	1.0	organic
16	NaN	NaN	NaN	None	12	1	6.0	organic
17	NaN	NaN	NaN	None	13	1	6.0	organic
18	NaN	NaN	NaN	None	14	1	4.0	organic
19	NaN	NaN	NaN	None	15	1	1.0	organic
20	NaN	NaN	NaN	None	16	1	1.0	organic
21	NaN	NaN	NaN	None	17	1	1.0	organic
22	NaN	NaN	NaN	None	18	1	6.0	organic
23	NaN	NaN	NaN	None	19	1	6.0	organic
24	NaN	NaN	NaN	None	20	1	1.0	organic
25	NaN	NaN	NaN	None	21	1	6.0	organic
26	3.0	0.0	0.1	[0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, ...	22	1	NaN	band1

```
history[0:8]
```

```
actions[0:8]
```

```
[array([0, 0, 0, 1, 0, 0, 0, 0, 0, 0], dtype=int8),  
 array([0, 0, 0, 1, 0, 0, 0, 0, 0, 0], dtype=int8),  
 array([0, 0, 0, 0, 1, 0, 0, 0, 0, 0], dtype=int8),  
 array([0, 1, 0, 0, 0, 0, 0, 0, 0, 0], dtype=int8),  
 array([0, 0, 0, 0, 0, 0, 0, 1, 0, 0], dtype=int8),  
 array([0, 0, 0, 1, 0, 0, 0, 0, 0, 0], dtype=int8),  
 array([0, 0, 1, 0, 0, 0, 0, 0, 0, 0], dtype=int8),  
 array([1, 0, 0, 0, 0, 0, 0, 0, 0, 0], dtype=int8)]
```

How would you interact the features?

# Feature Engineering

a	c	ps	ps-a	t	u	v	z	
0	NaN	NaN	NaN	None	0	0.0	organic	
1	3.0	0.0	0.1	[0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, ...	1	0	NaN	bandit
2	4.0	0.0	0.1	[0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, ...	2	0	NaN	bandit
3	5.0	0.0	0.1	[0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, ...	3	0	NaN	bandit
4	NaN	NaN	NaN	None	0	1.0	organic	
5	2.0	0.0	0.1	[0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, ...	1	1	NaN	bandit
6	8.0	0.0	0.1	[0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, ...	2	1	NaN	bandit
7	4.0	0.0	0.1	[0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, ...	3	1	NaN	bandit
8	NaN	NaN	NaN	None	4	1.6	organic	
9	NaN	NaN	NaN	None	5	1.6	organic	
10	NaN	NaN	NaN	None	6	1.4	organic	
11	NaN	NaN	NaN	None	7	1.6	organic	
12	NaN	NaN	NaN	None	8	1.2	organic	
13	NaN	NaN	NaN	None	9	1.6	organic	
14	NaN	NaN	NaN	None	10	1.2	organic	
15	NaN	NaN	NaN	None	11	1.0	organic	
16	NaN	NaN	NaN	None	12	1.6	organic	
17	NaN	NaN	NaN	None	13	1.6	organic	
18	NaN	NaN	NaN	None	14	1.4	organic	
19	NaN	NaN	NaN	None	15	1.1	organic	
20	NaN	NaN	NaN	None	16	1.0	organic	
21	NaN	NaN	NaN	None	17	1.0	organic	
22	NaN	NaN	NaN	None	18	1.6	organic	
23	NaN	NaN	NaN	None	19	1.6	organic	
24	NaN	NaN	NaN	None	20	1.1	organic	
25	NaN	NaN	NaN	None	21	1.6	organic	
26	3.0	0.0	0.1	[0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, ...	22	1	NaN	bandit

```
history[0:8]
```

```
[array([1., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.]),  
 array([1., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.]),  
 array([1., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.]),  
 array([0., 1., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.]),  
 array([0., 1., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.]),  
 array([0., 1., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.]),  
 array([0., 6., 2., 0., 2., 0., 9., 0., 0., 0., 0., 0., 0., 0.]),  
 array([0., 6., 2., 0., 2., 0., 9., 0., 0., 0., 0., 0., 0., 0.])]
```

```
actions[0:8]
```

```
[array([0, 0, 0, 1, 0, 0, 0, 0, 0, 0], dtype=int8),  
 array([0, 0, 0, 1, 0, 0, 0, 0, 0, 0], dtype=int8),  
 array([0, 0, 0, 0, 1, 0, 0, 0, 0, 0], dtype=int8),  
 array([0, 1, 0, 0, 0, 0, 0, 0, 0, 0], dtype=int8),  
 array([0, 0, 0, 0, 0, 0, 0, 0, 1, 0], dtype=int8),  
 array([0, 0, 0, 1, 0, 0, 0, 0, 0, 0], dtype=int8),  
 array([0, 0, 1, 0, 0, 0, 0, 0, 0, 0], dtype=int8),  
 array([1, 0, 0, 0, 0, 0, 0, 0, 0, 0], dtype=int8)]
```

How would you interact the features?

Is this a good feature engineering scheme? Can you improve it?

## Pure Bandit Feedback

- You need to build a recommender system for a client, they have an existing system with existing organic traffic.

## Pure Bandit Feedback

- You need to build a recommender system for a client, they have an existing system with existing organic traffic. This means you can build standard next item predictions to build models of user behavior and understand how user's preferences cluster together.

## Pure Bandit Feedback

- You need to build a recommender system for a client, they have an existing system with existing organic traffic. This means you can build standard next item predictions to build models of user behavior and understand how user's preferences cluster together.
- They are currently running a recommender system with a random policy.

## Pure Bandit Feedback

- You need to build a recommender system for a client, they have an existing system with existing organic traffic. This means you can build standard next item predictions to build models of user behavior and understand how user's preferences cluster together.
- They are currently running a recommender system with a random policy.
- Your task is to build a next item prediction model as a proxy for a recommendation algorithm. Then produce offline metrics to convince the engineering team to run an AB test.

## Pure Bandit Feedback

- You need to build a recommender system for a client, they have an existing system with existing organic traffic. This means you can build standard next item predictions to build models of user behavior and understand how user's preferences cluster together.
- They are currently running a recommender system with a random policy.
- Your task is to build a next item prediction model as a proxy for a recommendation algorithm. Then produce offline metrics to convince the engineering team to run an AB test.
- Finally you evaluate your performance against production.

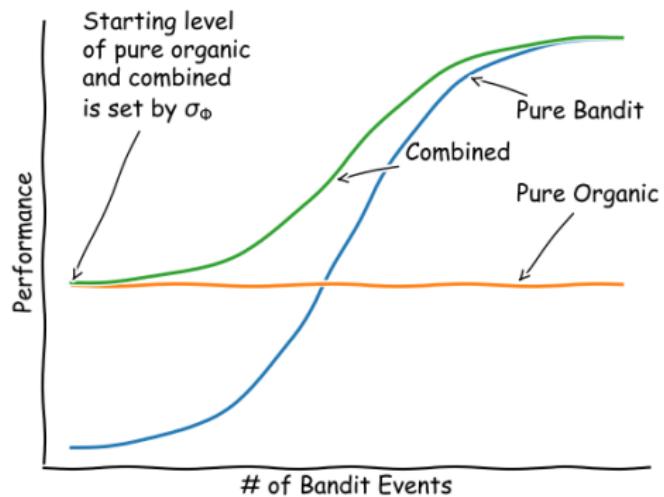
## Pure Bandit Feedback

- You need to build a recommender system for a client, they have an existing system with existing organic traffic. This means you can build standard next item predictions to build models of user behavior and understand how user's preferences cluster together.
- They are currently running a recommender system with a random policy.
- Your task is to build a next item prediction model as a proxy for a recommendation algorithm. Then produce offline metrics to convince the engineering team to run an AB test.
- Finally you evaluate your performance against production.

Please look at notebook “Module III C.ipynb”

# Combining Organic Signal with Bandit Signal

# Pure Organic vs Pure Bandit



# Can we get the best of both worlds?

A simple algorithm:

# Can we get the best of both worlds?

A simple algorithm:

Use one of the methods from the first part of the course on the organic data to produce product embeddings. Classic methods are SVD or word2vec.

## Can we get the best of both worlds?

A simple algorithm:

Use one of the methods from the first part of the course on the organic data to produce product embeddings. Classic methods are SVD or word2vec.

A user's history is then a list of embeddings rather than non-descript identifiers.

## Can we get the best of both worlds?

A simple algorithm:

Use one of the methods from the first part of the course on the organic data to produce product embeddings. Classic methods are SVD or word2vec.

A user's history is then a list of embeddings rather than non-descript identifiers.

The action space is also in the same embedding space.

## Can we get the best of both worlds?

A simple algorithm:

Use one of the methods from the first part of the course on the organic data to produce product embeddings. Classic methods are SVD or word2vec.

A user's history is then a list of embeddings rather than non-descript identifiers.

The action space is also in the same embedding space.

Exercise: implement this approach, use module III B and C as a reference.

## Can we get the best of both worlds?

A simple algorithm:

Use one of the methods from the first part of the course on the organic data to produce product embeddings. Classic methods are SVD or word2vec.

A user's history is then a list of embeddings rather than non-descript identifiers.

The action space is also in the same embedding space.

Exercise: implement this approach, use module III B and C as a reference.

*What advantages or challenges do you have in this approach?*

Thank You!