A plausible model of recognition and postdiction in dynamic environment

Kevin Li, Maneesh Sahani

Gatsby Computational Neuroscience Unit, University College London

January 6, 2020

Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

January 6, 2020 1/20

1. Introduction

Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

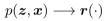
January 6, 2020 2/20

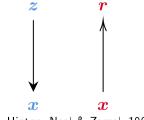
- 32

イロト イヨト イヨト イヨ

Inference using an internal model (Helmholtz machine)

static world





Dayan, Hinton, Neal & Zemel, 1995

Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

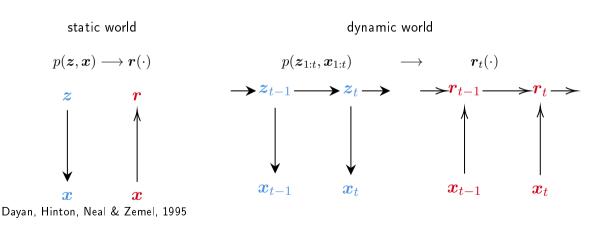
Model of recognition and postdiction

January 6, 2020 3 / 20

- 32

(日)

Inference using an internal model (Helmholtz machine)



- 34

イロト イポト イヨト イヨト

Illusion 1

Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

January 6, 2020 4 / 20

◆□ → ◆檀 → ◆臣 → ◆臣 → □臣

Illusion 1

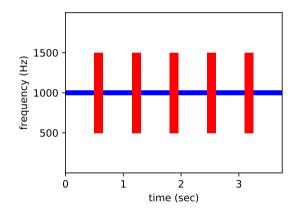
Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

January 6, 2020 4 / 20

◆□ → ◆檀 → ◆臣 → ◆臣 → □臣

Illusion 1



Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

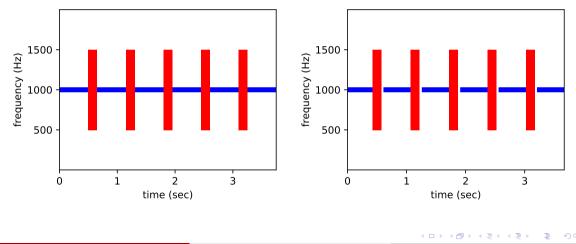
January 6, 2020 4 / 20

3

イロト イヨト イヨト イヨ

Introduction

Illusion 1



Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

January 6, 2020 4 / 20

Illusion 2: cutaneous rabbit

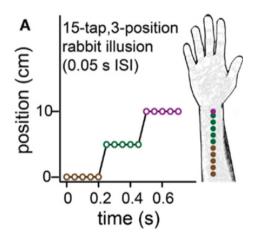
In the course of designing some experiments on the cutaneous perception of mechanical pulses delivered to the back of the forearm, it was discovered that, under some conditions of timing, the taps produced seemed not to be properly localized under the contactors. [...] They will seem to be distributed, with more or less uniform spacing, from the region of the first contactor to that of the third. There is a smooth progression of jumps up the arm, as if a tiny rabbit were hopping from elbow to wrist.

Geldard & Sherrick, 1972, Science

Illusion 2: cutaneous rabbit

In the course of designing some experiments on the cutaneous perception of mechanical pulses delivered to the back of the forearm, it was discovered that, under some conditions of timing, the taps produced seemed not to be properly localized under the contactors. [...] They will seem to be distributed, with more or less uniform spacing, from the region of the first contactor to that of the third. There is a smooth progression of jumps up the arm, as if a tiny rabbit were hopping from elbow to wrist.

Geldard & Sherrick, 1972, Science



Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

January 6, 2020 6/20

(日)、(同)、(日)、(日)、(日)、

• Percept of the past is changed by new, future observation

Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

January 6, 2020 6/20

- 32

イロト イヨト イヨト イヨ

- Percept of the past is changed by new, future observation
- "Law of Continuity"

Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

3

イロト イヨト イヨト イヨ

- Percept of the past is changed by new, future observation
- "Law of Continuity"

What could be the neural basis for these statistical computation?

Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

3

A B A B A B
 A
 B
 A
 A
 B
 A
 A
 B
 A
 A
 B
 A
 A
 B
 A
 A
 B
 A
 A
 B
 A
 A
 B
 A
 A
 B
 A
 A
 B
 A
 A
 B
 A
 A
 B
 A
 A
 B
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A

- Percept of the past is changed by new, future observation
- "Law of Continuity"

What could be the neural basis for these statistical computation?

• representing beliefs as distributional objects

3

- Percept of the past is changed by new, future observation
- "Law of Continuity"
- What could be the neural basis for these statistical computation?
 - representing beliefs as distributional objects
 - updating beliefs of the past based on new evidence in real time?

- Percept of the past is changed by new, future observation
- "Law of Continuity"

What could be the neural basis for these statistical computation?

- representing beliefs as distributional objects
- updating beliefs of the past based on new evidence in real time?
- learning to do all the above

2. Distributed distributional code

Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

January 6, 2020 7 / 20

イロト イヨト イヨト イヨ

Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

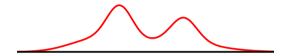
January 6, 2020 8/20

3

メロト メポト メヨト メヨ

A DDC encodes a probability distribution:

 $m{q}(m{z})$

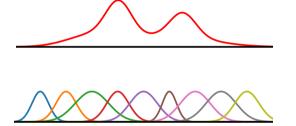


A DDC encodes a probability distribution:

 $\boldsymbol{q}(\boldsymbol{z})$

by a set of tuning functions

 $oldsymbol{\gamma}(oldsymbol{z}) := \left[\gamma_1\left(oldsymbol{z}
ight), oldsymbol{\gamma_2}\left(oldsymbol{z}
ight), \gamma_3\left(oldsymbol{z}
ight), ..., oldsymbol{\gamma_K}\left(oldsymbol{z}
ight)
ight]$



Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

January 6, 2020 8/20

A DDC encodes a probability distribution:

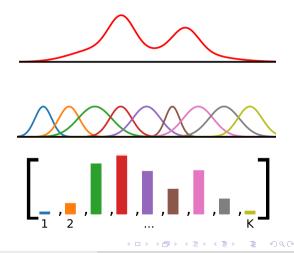
 $\boldsymbol{q}(\boldsymbol{z})$

by a set of tuning functions

 $oldsymbol{\gamma}(oldsymbol{z}) := \left[\gamma_{1}\left(oldsymbol{z}
ight), rac{\gamma_{2}\left(oldsymbol{z}
ight), \gamma_{3}\left(oldsymbol{z}
ight), ..., rac{\gamma_{K}\left(oldsymbol{z}
ight)}{\gamma_{2}\left(oldsymbol{z}
ight), \gamma_{3}\left(oldsymbol{z}
ight), ..., rac{\gamma_{K}\left(oldsymbol{z}
ight)}{\gamma_{K}\left(oldsymbol{z}
ight)}
ight]$

into a set of expectations

$$m{r}:=\mathbb{E}_{m{q}(m{z})}\left[\gamma_{1}\left(m{z}
ight),rac{\gamma_{2}\left(m{z}
ight),...,\gamma_{K}\left(m{z}
ight)
ight]$$



Model of recognition and postdiction

A DDC encodes a probability distribution:

q(z)

by a set of tuning functions

$$oldsymbol{\gamma}(oldsymbol{z}) := \left[\gamma_1\left(oldsymbol{z}
ight), oldsymbol{\gamma}_2\left(oldsymbol{z}
ight), \gamma_3\left(oldsymbol{z}
ight), ..., oldsymbol{\gamma}_K\left(oldsymbol{z}
ight)
ight]$$

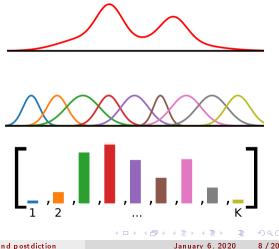
into a set of expectations

$$m{r}:=\mathbb{E}_{m{q}(m{z})}\left[\gamma_{1}\left(m{z}
ight),\gamma_{2}\left(m{z}
ight),...,\gamma_{K}\left(m{z}
ight)
ight]$$

Zemel, Dayan & Pouget (1998); Sahani & Dayan (2003), Vértes & Sahani (2018)

Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction



given $\mathbb{E}_{q}\left[\gamma_{k}\left(\boldsymbol{z}\right)\right] = r_{k}, \quad \forall k \in \{1, 2, \dots K\}$

Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

January 6, 2020 9/20

- 34

メロト メポト メヨト メヨト

given $\mathbb{E}_{q}\left[\gamma_{k}\left(\boldsymbol{z}\right)\right] = r_{k}, \quad \forall k \in \{1, 2, \dots K\}$ $q(\boldsymbol{z}) = \arg \max H[q]$

- 34

given
$$\mathbb{E}_{q}\left[\gamma_{k}\left(\boldsymbol{z}\right)\right] = r_{k}, \quad \forall k \in \{1, 2, \dots K\}$$

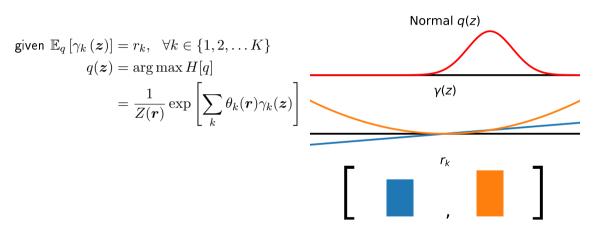
 $q(\boldsymbol{z}) = rg \max H[q]$
 $= \frac{1}{Z(\boldsymbol{r})} \exp \left[\sum_{k} \theta_{k}(\boldsymbol{r})\gamma_{k}(\boldsymbol{z})\right]$

Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

January 6, 2020 9/20

(日)



Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

January 6, 2020 9/20

3

Image: A matrix

given
$$\mathbb{E}_{q}\left[\gamma_{k}\left(\boldsymbol{z}\right)\right] = r_{k}, \quad \forall k \in \{1, 2, \dots K\}$$

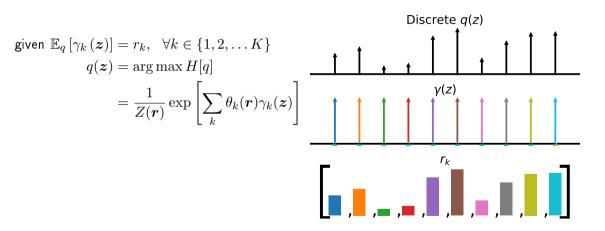
 $q(\boldsymbol{z}) = rg \max H[q]$
 $= \frac{1}{Z(\boldsymbol{r})} \exp \left[\sum_{k} \theta_{k}(\boldsymbol{r})\gamma_{k}(\boldsymbol{z})\right]$

Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

January 6, 2020 9/20

(日)



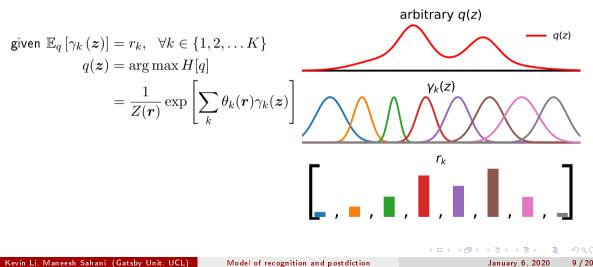
Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

January 6, 2020 9/20

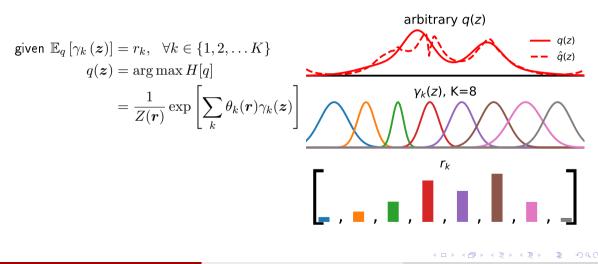
- 31

A D N A B N A B N A B



Kevin Li. Maneesh Sahani (Gatsby Unit. UCL)

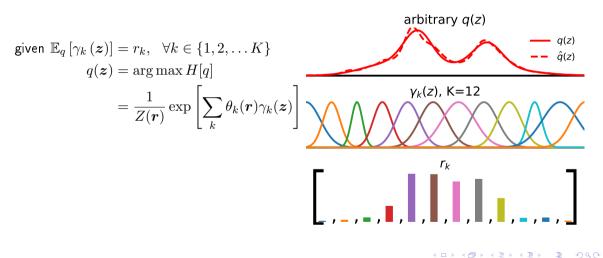
Model of recognition and postdiction



Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

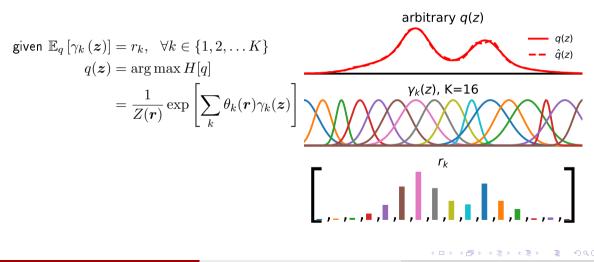
January 6, 2020 9/20



Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

January 6, 2020 9/20



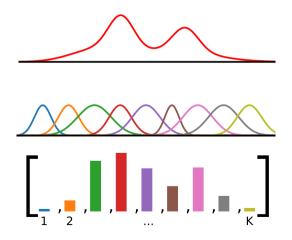
Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

January 6, 2020 9/20

Why DDC? It makes computation simple for neurons

$$oldsymbol{r}:=\mathbb{E}_{oldsymbol{q}\left(oldsymbol{z}
ight)}\left[\gamma_{1}\left(oldsymbol{z}
ight),\gamma_{2}\left(oldsymbol{z}
ight),...,\gamma_{K}\left(oldsymbol{z}
ight)
ight]$$



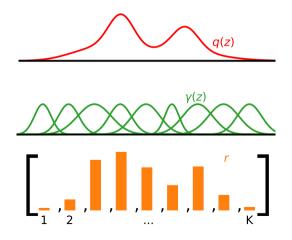
Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

January 6, 2020 10 / 20

Why DDC? It makes computation simple for neurons

$$oldsymbol{r}:=\mathbb{E}_{oldsymbol{q}\left(oldsymbol{z}
ight)}\left[\gamma_{1}\left(oldsymbol{z}
ight),\gamma_{2}\left(oldsymbol{z}
ight),...,\gamma_{K}\left(oldsymbol{z}
ight)
ight]$$



Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

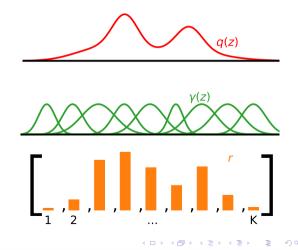
Model of recognition and postdiction

January 6, 2020 10 / 20

Why DDC? It makes computation simple for neurons

$$oldsymbol{r}:=\mathbb{E}_{oldsymbol{q}(oldsymbol{z})}\left[\gamma_{1}\left(oldsymbol{z}
ight),\gamma_{2}\left(oldsymbol{z}
ight),...,\gamma_{K}\left(oldsymbol{z}
ight)
ight]$$

Key computations involve expected values:



Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

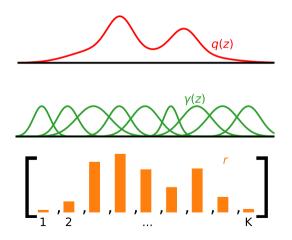
January 6, 2020 10 / 20

$$oldsymbol{r}:=\mathbb{E}_{oldsymbol{q}\left(oldsymbol{z}
ight)}\left[\gamma_{1}\left(oldsymbol{z}
ight),\gamma_{2}\left(oldsymbol{z}
ight),...,\gamma_{K}\left(oldsymbol{z}
ight)
ight]$$

Key computations involve expected values:

• Message passing:

 $q(z_2) = \mathbb{E}_{q(z_1)} \left[p(z_2|z_1) \right]$

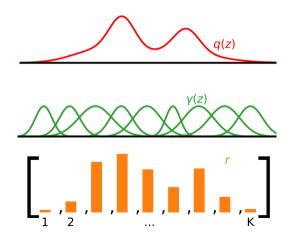


$$oldsymbol{r}:=\mathbb{E}_{oldsymbol{q}\left(oldsymbol{z}
ight)}\left[\gamma_{1}\left(oldsymbol{z}
ight),\gamma_{2}\left(oldsymbol{z}
ight),...,\gamma_{K}\left(oldsymbol{z}
ight)
ight]$$

Key computations involve expected values:

- Message passing: $q(z_2) = \mathbb{E}_{q(z_1)} \left[p(z_2|z_1) \right]$
- Marginalization:

$$q(z_2 \in (a, b)) = \mathbb{E}_{q(z_1, z_2)} [1(a < z_2 < b)]$$



$$oldsymbol{r}:=\mathbb{E}_{oldsymbol{q}\left(oldsymbol{z}
ight)}\left[\gamma_{1}\left(oldsymbol{z}
ight),\gamma_{2}\left(oldsymbol{z}
ight),...,\gamma_{K}\left(oldsymbol{z}
ight)
ight]$$

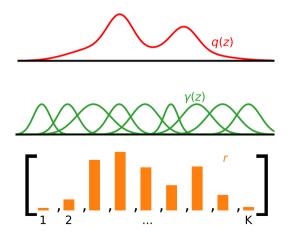
Key computations involve expected values:

- Message passing: $q(z_2) = \mathbb{E}_{q(z_1)} \left[p(z_2 | z_1) \right]$
- Marginalization:

$$q(z_2 \in (a, b)) = \mathbb{E}_{q(z_1, z_2)} \left[1(a < z_2 < b) \right]$$

Action evaluation:

 $Q(a) = \mathbb{E}_{q(s)} \left[R(s, a) \right]$



January 6, 2020

10 / 20

$$oldsymbol{r}:=\mathbb{E}_{oldsymbol{q}\left(oldsymbol{z}
ight)}\left[\gamma_{1}\left(oldsymbol{z}
ight),\gamma_{2}\left(oldsymbol{z}
ight),...,\gamma_{K}\left(oldsymbol{z}
ight)
ight]$$

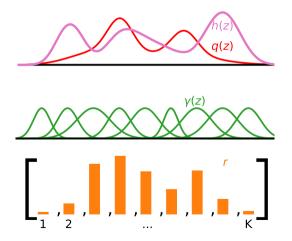
Key computations involve expected values:

- Message passing: $q(z_2) = \mathbb{E}_{q(z_1)} \left[p(z_2 | z_1) \right]$
- Marginalization:

$$q(z_2 \in (a, b)) = \mathbb{E}_{q(z_1, z_2)} \left[1(a < z_2 < b) \right]$$

- Action evaluation:
 - $Q(a) = \mathbb{E}_{q(s)} \left[R(s, a) \right]$

How to compute $\mathbb{E}\left[h(\boldsymbol{z})\right]$?



$$oldsymbol{r}:=\mathbb{E}_{oldsymbol{q}\left(oldsymbol{z}
ight)}\left[\gamma_{1}\left(oldsymbol{z}
ight),\gamma_{2}\left(oldsymbol{z}
ight),...,\gamma_{K}\left(oldsymbol{z}
ight)
ight]$$

Key computations involve expected values:

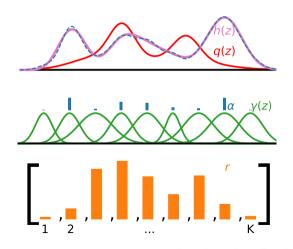
- Message passing: $q(z_2) = \mathbb{E}_{q(z_1)} \left[p(z_2 | z_1) \right]$
- Marginalization:

$$q(z_2 \in (a, b)) = \mathbb{E}_{q(z_1, z_2)} [1(a < z_2 < b)]$$

• Action evaluation: $Q(z) = \mathbb{E} \left[\frac{B(z)}{2} \right]$

 $Q(a) = \mathbb{E}_{q(s)} \left[R(s, a) \right]$ How to compute $\mathbb{E} \left[h(\boldsymbol{z}) \right]$?

if
$$h(\boldsymbol{z}) \approx \sum_i \alpha_i \gamma_i(\boldsymbol{z})$$



Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

January 6, 2020 10 / 20

$$oldsymbol{r}:=\mathbb{E}_{oldsymbol{q}\left(oldsymbol{z}
ight)}\left[\gamma_{1}\left(oldsymbol{z}
ight),\gamma_{2}\left(oldsymbol{z}
ight),...,\gamma_{K}\left(oldsymbol{z}
ight)
ight]$$

Key computations involve expected values:

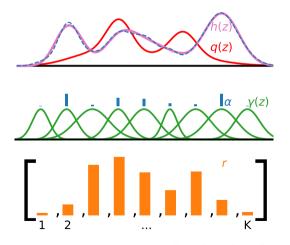
- Message passing: $q(z_2) = \mathbb{E}_{q(z_1)} \left[p(z_2 | z_1) \right]$
- Marginalization:

$$q(z_2 \in (a, b)) = \mathbb{E}_{q(z_1, z_2)} \left[1(a < z_2 < b) \right]$$

- Action evaluation:
 - $Q(a) = \mathbb{E}_{q(s)} \left[R(s, a) \right]$

How to compute $\mathbb{E}[h(z)]$?

$$\text{if } h(\boldsymbol{z}) \approx \sum_i \alpha_i \gamma_i(\boldsymbol{z}) \Rightarrow \mathbb{E}\left[h(\boldsymbol{z})\right] \approx \sum_i \alpha_i r_i$$



• Previously, we saw marginal $q(m{z}) \leftrightarrow \mathbb{E}_{q(m{z})}[m{\gamma}(m{z})]$

Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

3

A B A B A B A B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A

- ullet Previously, we saw marginal $q(m{z}) \leftrightarrow \mathbb{E}_{q(m{z})}\left[m{\gamma}(m{z})
 ight]$
- For $p(\boldsymbol{z}, \boldsymbol{x}) = p(\boldsymbol{z})p(\boldsymbol{x}|\boldsymbol{z}), \; p(\boldsymbol{z}|\boldsymbol{x})$ is intractable

simulation



Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

- Previously, we saw marginal $q(m{z}) \leftrightarrow \mathbb{E}_{q(m{z})}\left[m{\gamma}(m{z})
 ight]$
- For $p(\boldsymbol{z}, \boldsymbol{x}) = p(\boldsymbol{z})p(\boldsymbol{x}|\boldsymbol{z}), \; p(\boldsymbol{z}|\boldsymbol{x})$ is intractable
- How to obtain approx. $q(m{z}|m{x}) \leftrightarrow \mathbb{E}_{q(m{z}|m{x})}\left[m{\gamma}(m{z})
 ight]?$



Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

- Previously, we saw marginal $q(m{z}) \leftrightarrow \mathbb{E}_{q(m{z})}\left[m{\gamma}(m{z})
 ight]$
- For $p(\boldsymbol{z}, \boldsymbol{x}) = p(\boldsymbol{z})p(\boldsymbol{x}|\boldsymbol{z}), \ p(\boldsymbol{z}|\boldsymbol{x})$ is intractable
- How to obtain approx. $q(m{z}|m{x}) \leftrightarrow \mathbb{E}_{q(m{z}|m{x})}\left[m{\gamma}(m{z})
 ight]?$
- Clue: posterior mean...

$$\mathbb{E}_{p(oldsymbol{z}|oldsymbol{x})}\left[oldsymbol{\gamma}(oldsymbol{z})
ight] = rgmin_{oldsymbol{\phi}}\mathbb{E}_{p(oldsymbol{z}|oldsymbol{x})}\left[\|oldsymbol{\gamma}(oldsymbol{z}) - oldsymbol{\phi}\|_2^2
ight]$$

simulation



- Previously, we saw marginal $q(m{z}) \leftrightarrow \mathbb{E}_{q(m{z})}\left[m{\gamma}(m{z})
 ight]$
- For $p(\boldsymbol{z}, \boldsymbol{x}) = p(\boldsymbol{z})p(\boldsymbol{x}|\boldsymbol{z}), \; p(\boldsymbol{z}|\boldsymbol{x})$ is intractable
- How to obtain approx. $q(m{z}|m{x}) \leftrightarrow \mathbb{E}_{q(m{z}|m{x})}\left[m{\gamma}(m{z})
 ight]?$
- Clue: posterior mean...

$$\mathbb{E}_{p(oldsymbol{z}|oldsymbol{x})}\left[oldsymbol{\gamma}(oldsymbol{z})
ight] = rgmin_{oldsymbol{\phi}}\mathbb{E}_{p(oldsymbol{z}|oldsymbol{x})}\left[\|oldsymbol{\gamma}(oldsymbol{z})-oldsymbol{\phi}\|_2^2
ight]$$

• "Amortize" using
$$oldsymbol{\phi}_{oldsymbol{W}}\left(oldsymbol{x}
ight):=oldsymbol{W}oldsymbol{\sigma}(oldsymbol{x})$$

simulation



Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

- Previously, we saw marginal $q(m{z}) \leftrightarrow \mathbb{E}_{q(m{z})}\left[m{\gamma}(m{z})
 ight]$
- For $p(\boldsymbol{z}, \boldsymbol{x}) = p(\boldsymbol{z})p(\boldsymbol{x}|\boldsymbol{z}), \; p(\boldsymbol{z}|\boldsymbol{x})$ is intractable
- How to obtain approx. $q(m{z}|m{x}) \leftrightarrow \mathbb{E}_{q(m{z}|m{x})}\left[m{\gamma}(m{z})
 ight]?$
- Clue: posterior mean... $\mathbb{E}_{p(\boldsymbol{z}|\boldsymbol{x})} \left[\boldsymbol{\gamma}(\boldsymbol{z}) \right] = \operatorname*{arg\,min}_{\boldsymbol{\phi}} \mathbb{E}_{p(\boldsymbol{z}|\boldsymbol{x})} \left[\| \boldsymbol{\gamma}(\boldsymbol{z}) - \boldsymbol{\phi} \|_{2}^{2} \right]$

• "Amortize" using $oldsymbol{\phi}_{oldsymbol{W}}\left(oldsymbol{x}
ight):=oldsymbol{W}oldsymbol{\sigma}(oldsymbol{x})$

$$oldsymbol{W}^* = rgmin_{oldsymbol{W}} \mathbb{E}_{p(oldsymbol{z},oldsymbol{x})} \left[\|oldsymbol{\gamma}(oldsymbol{z}) - oldsymbol{W}oldsymbol{\sigma}(oldsymbol{x})\|_2^2
ight]$$

simulation



Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

January 6, 2020 11/20

- Previously, we saw marginal $q(m{z}) \leftrightarrow \mathbb{E}_{q(m{z})}\left[m{\gamma}(m{z})
 ight]$
- For $p(\boldsymbol{z}, \boldsymbol{x}) = p(\boldsymbol{z})p(\boldsymbol{x}|\boldsymbol{z}), \; p(\boldsymbol{z}|\boldsymbol{x})$ is intractable
- How to obtain approx. $q(m{z}|m{x}) \leftrightarrow \mathbb{E}_{q(m{z}|m{x})}\left[m{\gamma}(m{z})
 ight]?$
- Clue: posterior mean... $\mathbb{E}_{p(\boldsymbol{z}|\boldsymbol{x})} \left[\boldsymbol{\gamma}(\boldsymbol{z}) \right] = \operatorname*{arg\,min}_{\boldsymbol{\phi}} \mathbb{E}_{p(\boldsymbol{z}|\boldsymbol{x})} \left[\| \boldsymbol{\gamma}(\boldsymbol{z}) - \boldsymbol{\phi} \|_2^2 \right]$

• "Amortize" using $oldsymbol{\phi}_{oldsymbol{W}}\left(oldsymbol{x}
ight):=oldsymbol{W}oldsymbol{\sigma}(oldsymbol{x})$

$$egin{aligned} m{W}^* &= rg\min_{m{W}} \mathbb{E}_{p(m{z},m{x})} \left[\|m{\gamma}(m{z}) - m{W}m{\sigma}(m{x})\|_2^2
ight] \ m{r}(m{x}) &:= m{W}^*m{\sigma}(m{x}) = \mathbb{E}_{m{q}(m{z}|m{x})} \left[m{\gamma}(m{z})
ight] \end{aligned}$$

simulation



Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

January 6, 2020 11/20

- Previously, we saw marginal $q(m{z}) \leftrightarrow \mathbb{E}_{q(m{z})}\left[m{\gamma}(m{z})
 ight]$
- For $p(\boldsymbol{z}, \boldsymbol{x}) = p(\boldsymbol{z})p(\boldsymbol{x}|\boldsymbol{z}), \; p(\boldsymbol{z}|\boldsymbol{x})$ is intractable
- How to obtain approx. $q(m{z}|m{x}) \leftrightarrow \mathbb{E}_{q(m{z}|m{x})}\left[m{\gamma}(m{z})
 ight]?$
- Clue: posterior mean... $\mathbb{E}_{p(\boldsymbol{z}|\boldsymbol{x})} \left[\boldsymbol{\gamma}(\boldsymbol{z}) \right] = \operatorname*{arg\,min}_{\boldsymbol{\phi}} \mathbb{E}_{p(\boldsymbol{z}|\boldsymbol{x})} \left[\| \boldsymbol{\gamma}(\boldsymbol{z}) - \boldsymbol{\phi} \|_{2}^{2} \right]$

• "Amortize" using $oldsymbol{\phi}_{oldsymbol{W}}\left(oldsymbol{x}
ight):=oldsymbol{W}oldsymbol{\sigma}(oldsymbol{x})$

$$egin{aligned} m{W}^* &= rgmin_{m{W}} \mathbb{E}_{p(m{z},m{x})} \left[\|m{\gamma}(m{z}) - m{W}m{\sigma}(m{x})\|_2^2
ight] \ m{r}(m{x}) &:= m{W}^*m{\sigma}(m{x}) = \mathbb{E}_{m{q}(m{z}|m{x})} \left[m{\gamma}(m{z})
ight] \end{aligned}$$

• Find W* by the delta rule:

$$\Delta \boldsymbol{W} \propto (\boldsymbol{\gamma}(\boldsymbol{z}) - \boldsymbol{\phi}_{\boldsymbol{W}}(\boldsymbol{x}))\boldsymbol{\sigma}(\boldsymbol{x})^{\mathsf{T}}, \quad \{\boldsymbol{z}, \boldsymbol{x}\} \sim p$$

Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

simulation $\gamma(z)$ 2 \boldsymbol{x}

January 6, 2020 11/20

- Previously, we saw marginal $q(m{z}) \leftrightarrow \mathbb{E}_{q(m{z})}\left[m{\gamma}(m{z})
 ight]$
- For $p(\boldsymbol{z}, \boldsymbol{x}) = p(\boldsymbol{z})p(\boldsymbol{x}|\boldsymbol{z}), \ p(\boldsymbol{z}|\boldsymbol{x})$ is intractable
- How to obtain approx. $q(m{z}|m{x}) \leftrightarrow \mathbb{E}_{q(m{z}|m{x})}\left[m{\gamma}(m{z})
 ight]?$
- Clue: posterior mean... $\mathbb{E}_{p(\boldsymbol{z}|\boldsymbol{x})} \left[\boldsymbol{\gamma}(\boldsymbol{z}) \right] = \operatorname*{arg\,min}_{\boldsymbol{\phi}} \mathbb{E}_{p(\boldsymbol{z}|\boldsymbol{x})} \left[\| \boldsymbol{\gamma}(\boldsymbol{z}) - \boldsymbol{\phi} \|_{2}^{2} \right]$

• "Amortize" using $oldsymbol{\phi}_{oldsymbol{W}}\left(oldsymbol{x}
ight):=oldsymbol{W}oldsymbol{\sigma}(oldsymbol{x})$

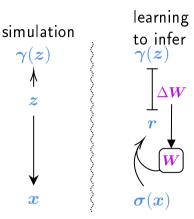
$$egin{aligned} m{W}^* &= rg\min_{m{W}} \mathbb{E}_{p(m{z},m{x})} \left[\|m{\gamma}(m{z}) - m{W}m{\sigma}(m{x})\|_2^2
ight] \ m{r}(m{x}) &:= m{W}^*m{\sigma}(m{x}) = \mathbb{E}_{m{q}(m{z}|m{x})} \left[m{\gamma}(m{z})
ight] \end{aligned}$$

• Find W^* by the **delta rule**:

 $\Delta W \propto (\gamma(z) - \phi_W(x))\sigma(x)^{\mathsf{T}}, \quad \{z, x\} \sim p$

Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction



January 6, 2020 11 / 20

Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

January 6, 2020 12 / 20

- 32

イロト イポト イヨト イヨト

Definition

DDC of $q(m{z})$ associated tuning functions $m{\gamma}(m{z})$ is $m{r}:=\mathbb{E}_{q(m{z})}\left[m{\gamma}(m{z})
ight]$

Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

January 6, 2020 12 / 20

3

A = A = A = A = A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A

Definition

DDC of $q(m{z})$ associated tuning functions $m{\gamma}(m{z})$ is $m{r}:=\mathbb{E}_{q(m{z})}\left[m{\gamma}(m{z})
ight]$

MaxEnt interpretation

$$r \stackrel{\mathbb{E}[m{\gamma}(m{z})]}{\longleftarrow} q(m{z}) \propto \exp\left[m{ heta} \cdot m{\gamma}(m{z})
ight]$$

Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

January 6, 2020 12 / 20

Definition

DDC of $q(m{z})$ associated tuning functions $m{\gamma}(m{z})$ is $m{r}:=\mathbb{E}_{q(m{z})}\left[m{\gamma}(m{z})
ight]$

MaxEnt interpretation

$$r \stackrel{\mathbb{E}[\gamma(oldsymbol{z})]}{\underbrace{}{\gamma(oldsymbol{z}),\mathsf{MaxEnt}}} q(oldsymbol{z}) \propto \exp\left[oldsymbol{ heta} \cdot oldsymbol{\gamma}(oldsymbol{z})
ight]$$

Expectation approximation

$$h(\boldsymbol{z}) pprox \boldsymbol{lpha} \cdot \boldsymbol{\gamma}(\boldsymbol{z}) \implies \mathbb{E}\left[h(\boldsymbol{z})\right] pprox \boldsymbol{lpha} \cdot \boldsymbol{r}$$

Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

January 6, 2020 12 / 20

Definition

DDC of $q(m{z})$ associated tuning functions $m{\gamma}(m{z})$ is $m{r}:=\mathbb{E}_{q(m{z})}\left[m{\gamma}(m{z})
ight]$

MaxEnt interpretation

$$r \stackrel{\mathbb{E}[m{\gamma}(m{z})]}{\longleftarrow} q(m{z}) \propto \exp\left[m{ heta} \cdot m{\gamma}(m{z})
ight]$$

Expectation approximation

$$h(\boldsymbol{z}) \approx \boldsymbol{\alpha} \cdot \boldsymbol{\gamma}(\boldsymbol{z}) \implies \mathbb{E}\left[h(\boldsymbol{z})\right] pprox \boldsymbol{\alpha} \cdot \boldsymbol{r}$$

Learning to infer given p(z, x)

$$oldsymbol{r}(oldsymbol{x}) = \mathbb{E}_{q(oldsymbol{z}|oldsymbol{x})} [oldsymbol{\gamma}(oldsymbol{z})] = oldsymbol{W}^* oldsymbol{\sigma}(oldsymbol{x}), \qquad \Delta oldsymbol{W} \propto (oldsymbol{\gamma} - oldsymbol{\phi}_{oldsymbol{W}}) oldsymbol{\sigma}^\intercal, \quad \{oldsymbol{z},oldsymbol{x}\} \sim p(oldsymbol{z},oldsymbol{x})$$

Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

January 6, 2020 12 / 20

э

・ロット 全部 マイロット 中国

3. Online recognition and postdiction

Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

January 6, 2020 13 / 20

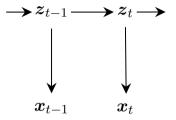
э

Image: A math a math

A generic dynamic internal model

We assume a generic internal model

$$egin{aligned} oldsymbol{z}_t &= oldsymbol{f}(oldsymbol{z}_{t-1}, \xi^{(z)}) \ oldsymbol{x}_t &= oldsymbol{g}(oldsymbol{z}_t, \xi^{(x)}) \end{aligned}$$



Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

January 6, 2020 14 / 20

- 32

メロト メポト メヨト メヨト

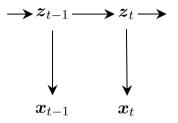
A generic dynamic internal model

We assume a generic internal model

$$egin{aligned} oldsymbol{z}_t &= oldsymbol{f}(oldsymbol{z}_{t-1}, \xi^{(z)}) \ oldsymbol{x}_t &= oldsymbol{g}(oldsymbol{z}_t, \xi^{(x)}) \end{aligned}$$

Assumptions

- Discrete-time
- Markov property
- Stationarity



Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

ъ January 6, 2020 14 / 20

• Online recognition (filtering): maintain $q(\boldsymbol{z}_t|\boldsymbol{x}_{1:t})$

Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

- 31

• Online recognition (filtering): maintain $q(\boldsymbol{z}_{1:t}|\boldsymbol{x}_{1:t})$ to allow postdiction

Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

January 6, 2020 15 / 20

- 31

- Online recognition (filtering): maintain $q(\boldsymbol{z}_{1:t}|\boldsymbol{x}_{1:t})$ to allow postdiction
- Define temporally extended encoding function $\psi_t := \psi(z_{1:t})$ for DDC

- Online recognition (filtering): maintain $q(m{z}_{1:t}|m{x}_{1:t})$ to allow postdiction
- Define temporally extended encoding function $oldsymbol{\psi}_t := oldsymbol{\psi}(oldsymbol{z}_{1:t})$ for DDC
- ullet A plausible $oldsymbol{\psi}_t$

$$egin{aligned} oldsymbol{\psi}_1 &:= oldsymbol{\gamma}\left(oldsymbol{z}_1
ight) \ oldsymbol{\psi}_t &:= oldsymbol{U}oldsymbol{\psi}_{t-1} + oldsymbol{\gamma}\left(oldsymbol{z}_t
ight) \end{aligned}$$

- 34

(日) (四) (日) (日) (日)

- Online recognition (filtering): maintain $q(\boldsymbol{z}_{1:t}|\boldsymbol{x}_{1:t})$ to allow postdiction
- Define temporally extended encoding function $\psi_t := \psi\left(oldsymbol{z}_{1:t}
 ight)$ for DDC
- ullet A plausible $oldsymbol{\psi}_t$

$$egin{aligned} oldsymbol{\psi}_{1} &:= oldsymbol{\gamma}\left(oldsymbol{z}_{1}
ight) \ oldsymbol{\psi}_{t} &:= oldsymbol{U}oldsymbol{\psi}_{t-1} + oldsymbol{\gamma}\left(oldsymbol{z}_{t}
ight) \end{aligned}$$

• Update beliefs about the past: compute

$$oldsymbol{r}_t = \mathbb{E}_{q(oldsymbol{z}_{1:t} | oldsymbol{x}_{1:t})} \left[oldsymbol{\psi}_t
ight]$$

from $oldsymbol{r}_{t-1}$ and $oldsymbol{x}_t$

イロン 不得 とうせい かほう 一日

- Online recognition (filtering): maintain $q(m{z}_{1:t}|m{x}_{1:t})$ to allow postdiction
- Define temporally extended encoding function $oldsymbol{\psi}_t := oldsymbol{\psi}(oldsymbol{z}_{1:t})$ for DDC
- ullet A plausible $oldsymbol{\psi}_t$

$$egin{aligned} oldsymbol{\psi}_{1} &:= oldsymbol{\gamma}\left(oldsymbol{z}_{1}
ight) \ oldsymbol{\psi}_{t} &:= oldsymbol{U}oldsymbol{\psi}_{t-1} + oldsymbol{\gamma}\left(oldsymbol{z}_{t}
ight) \end{aligned}$$

• Update beliefs about the past: compute

$$oldsymbol{r}_t = \mathbb{E}_{q(oldsymbol{z}_{1:t} | oldsymbol{x}_{1:t})} \left[oldsymbol{\psi}_t
ight]$$

from $oldsymbol{r}_{t-1}$ and $oldsymbol{x}_t$

• **Postiction:** readout statistics

$$h(\boldsymbol{z}_{t-\tau}) pprox \boldsymbol{\alpha} \cdot \boldsymbol{\psi}(\boldsymbol{z}_{1:t}) \implies \mathbb{E}_{q(\boldsymbol{z}_{t-\tau})}[h(\boldsymbol{z}_{t-\tau})] pprox \boldsymbol{\alpha} \cdot \boldsymbol{r}_{t}$$

Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

K A T K A T K

• Want $oldsymbol{r}_t(oldsymbol{x}_{1:t}) = \mathbb{E}_{q_t}\left[oldsymbol{\psi}_t\left(oldsymbol{z}_{1:t}
ight)
ight]$

Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

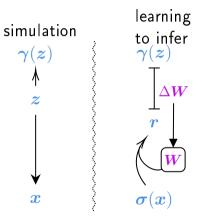
Model of recognition and postdiction

January 6, 2020 16 / 20

- 20

イロト 不得下 イヨト イヨト

- Want $oldsymbol{r}_t(oldsymbol{x}_{1:t}) = \mathbb{E}_{q_t}\left[oldsymbol{\psi}_t\left(oldsymbol{z}_{1:t}
 ight)
 ight]$
- Recall for $p(oldsymbol{z},oldsymbol{x})$ $oldsymbol{r}:=oldsymbol{\phi}_{oldsymbol{W}^*}(oldsymbol{x})$



Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

- Want $m{r}_t(m{x}_{1:t}) = \mathbb{E}_{q_t}\left[m{\psi}_t\left(m{z}_{1:t}
 ight)
 ight]$
- Recall for $p(oldsymbol{z},oldsymbol{x})$ $oldsymbol{r}:=oldsymbol{\phi}_{oldsymbol{W}^*}(oldsymbol{x})$
- Likewise, for SSM $p(oldsymbol{z}_{1:t},oldsymbol{x}_{1:t})$

$$\boldsymbol{r}_t := \boldsymbol{\phi}_{\boldsymbol{W}_t}(\boldsymbol{x}_{1:t})$$

- 34

(日) (四) (日) (日) (日)

- Want $oldsymbol{r}_t(oldsymbol{x}_{1:t}) = \mathbb{E}_{q_t}\left[oldsymbol{\psi}_t\left(oldsymbol{z}_{1:t}
 ight)
 ight]$
- Recall for $p(oldsymbol{z},oldsymbol{x})$ $oldsymbol{r}:=oldsymbol{\phi}_{oldsymbol{W}^*}(oldsymbol{x})$
- Likewise, for SSM $p(oldsymbol{z}_{1:t},oldsymbol{x}_{1:t})$

$$oldsymbol{r}_t := oldsymbol{\phi}_{oldsymbol{W}_t}(oldsymbol{r}_{t-1},oldsymbol{x}_t)$$

- 34

(日) (四) (日) (日) (日)

- Want $m{r}_t(m{x}_{1:t}) = \mathbb{E}_{q_t}\left[m{\psi}_t\left(m{z}_{1:t}
 ight)
 ight]$
- Recall for $p(oldsymbol{z},oldsymbol{x})$ $oldsymbol{r}:=oldsymbol{\phi}_{oldsymbol{W}^*}(oldsymbol{x})$
- Likewise, for SSM $p(oldsymbol{z}_{1:t},oldsymbol{x}_{1:t})$

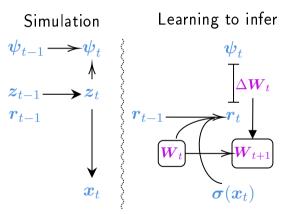
$$egin{aligned} m{r}_t &:= m{\phi}_{m{W}_t}(m{r}_{t-1},m{x}_t) \ &= m{W}_t \cdot (m{r}_{t-1} \otimes m{\sigma}(m{x}_t)) \end{aligned}$$

- 32

A D N A B N A B N A B

- Want $oldsymbol{r}_t(oldsymbol{x}_{1:t}) = \mathbb{E}_{q_t}\left[oldsymbol{\psi}_t\left(oldsymbol{z}_{1:t}
 ight)
 ight]$
- Recall for $p(oldsymbol{z},oldsymbol{x})$ $oldsymbol{r}:=oldsymbol{\phi}_{oldsymbol{W}^*}(oldsymbol{x})$
- Likewise, for SSM $p(oldsymbol{z}_{1:t},oldsymbol{x}_{1:t})$

 $egin{aligned} m{r}_t &:= m{\phi}_{m{W}_t}(m{r}_{t-1},m{x}_t) \ &= m{W}_t \cdot (m{r}_{t-1} \otimes m{\sigma}(m{x}_t)) \end{aligned}$

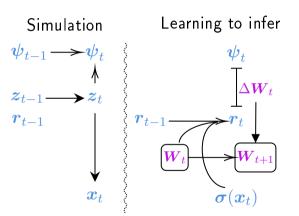


э

- Want $\boldsymbol{r}_t(\boldsymbol{x}_{1:t}) = \mathbb{E}_{a_t} \left[\boldsymbol{\psi}_t \left(\boldsymbol{z}_{1:t} \right) \right]$
- Recall for $p(\boldsymbol{z}, \boldsymbol{x})$ $oldsymbol{r}:=oldsymbol{\phi}_{oldsymbol{W}^*}(oldsymbol{x})$
- Likewise, for SSM $p(z_{1:t}, x_{1:t})$

 $\boldsymbol{r}_t := \boldsymbol{\phi}_{\boldsymbol{W}_t}(\boldsymbol{r}_{t-1}, \boldsymbol{x}_t)$ $= \mathbf{W}_t \cdot (\mathbf{r}_{t-1} \otimes \boldsymbol{\sigma}(\mathbf{x}_t))$

• Learning by the delta rule $\Delta W_t \leftarrow (\psi_t - \phi_t)(r_{t-1} \otimes \sigma_t)^{\mathsf{T}}$



< 口 > < 同

Kevin Li. Maneesh Sahani (Gatsby Unit. UCL)

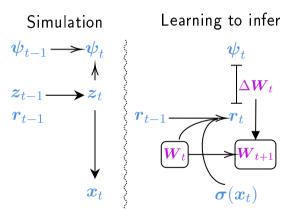
э

Learning to postdict online

- Want $oldsymbol{r}_t(oldsymbol{x}_{1:t}) = \mathbb{E}_{q_t}\left[oldsymbol{\psi}_t\left(oldsymbol{z}_{1:t}
 ight)
 ight]$
- Recall for $p(oldsymbol{z},oldsymbol{x})$ $oldsymbol{r}:=oldsymbol{\phi}_{oldsymbol{W}^*}(oldsymbol{x})$
- Likewise, for SSM $p(oldsymbol{z}_{1:t},oldsymbol{x}_{1:t})$

 $egin{aligned} m{r}_t &:= m{\phi}_{m{W}_t}(m{r}_{t-1},m{x}_t) \ &= m{W}_t \cdot (m{r}_{t-1} \otimes m{\sigma}(m{x}_t)) \end{aligned}$

• Learning by the delta rule $\Delta \boldsymbol{W}_t \leftarrow (\boldsymbol{\psi}_t - \boldsymbol{\phi}_t)(\boldsymbol{r}_{t-1} \otimes \boldsymbol{\sigma}_t)^{\mathsf{T}}$ $\{\boldsymbol{\psi}_t, \boldsymbol{x}_t, \boldsymbol{r}_{t-1}\} \sim p(\boldsymbol{z}_{1:t}, \boldsymbol{x}_{1:t}), \{\boldsymbol{h}_{\boldsymbol{W}_i}\}_{i=1}^{t-1}$



< <p>Image: A <

э

▶ ∢ ⊒

< E.

4. Testing DDC filtering on simulated experiments

Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

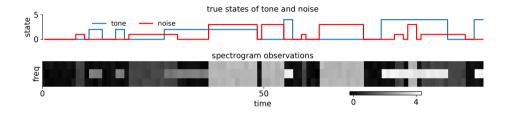
January 6, 2020 17 / 20

Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

January 6, 2020 18 / 20

- 32



Model of recognition and postdiction

January 6, 2020 18 / 20

э

< □ > < @ >



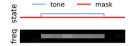
Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

January 6, 2020 18 / 20

3

・ロト ・ 理 ト ・ ヨ ト ・



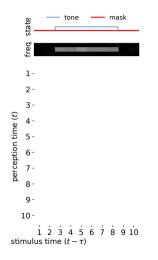
Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

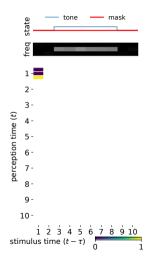
January 6, 2020 18 / 20

3

・ロト ・ 理 ト ・ ヨ ト ・

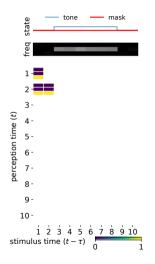


Э



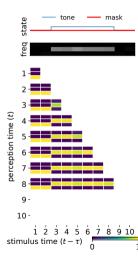
Э

Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)



January 6, 2020 18 / 20

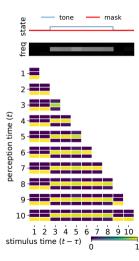
Э



Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

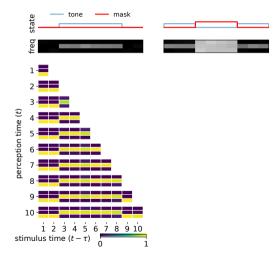
Э January 6, 2020 18 / 20



Model of recognition and postdiction

January 6, 2020 18 / 20

Э

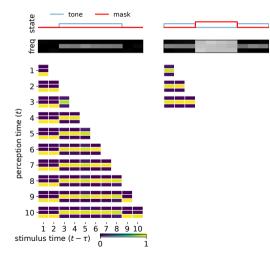


Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

January 6, 2020 18 / 20

э

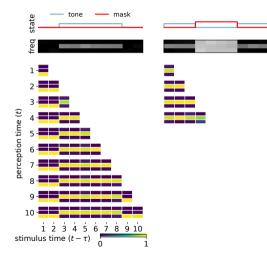


Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

January 6, 2020 18 / 20

э

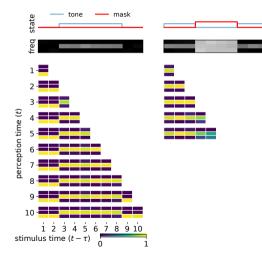


Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

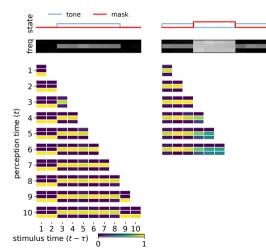
January 6, 2020 18 / 20

э



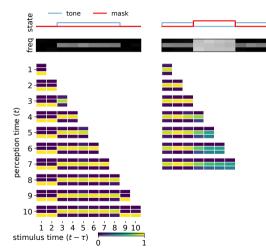
Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction



Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

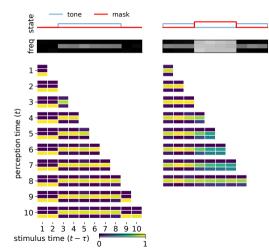


Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

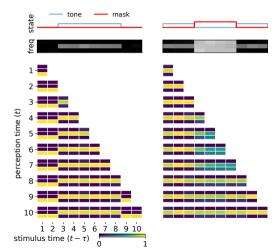
January 6, 2020 18 / 20

э



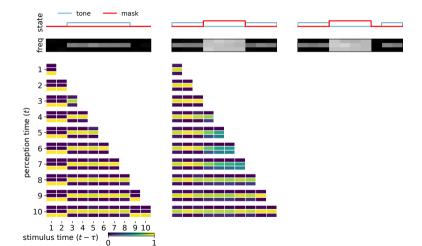
Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction



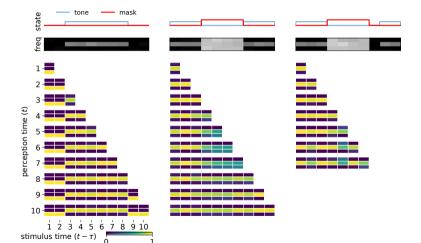
Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction



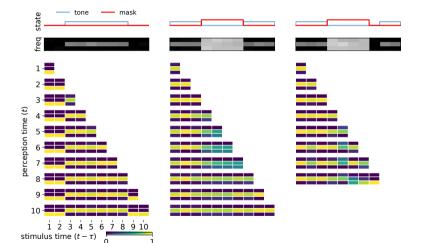
Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction



Model of recognition and postdiction

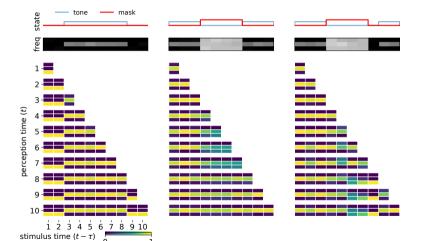
<ロト < 部ト < 言ト < 言ト 言 の < C January 6, 2020 18 / 20



Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

<ロト < 部ト < 言ト < 言ト 言 の < C January 6, 2020 18 / 20



'n

Model of recognition and postdiction

January 6, 2020 18 / 20

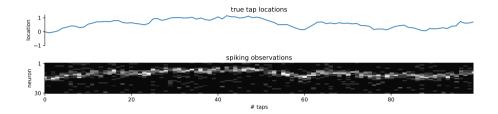
3

Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

January 6, 2020 19 / 20

12



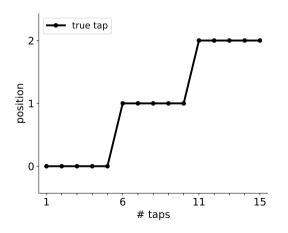
Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

January 6, 2020 19 / 20

3

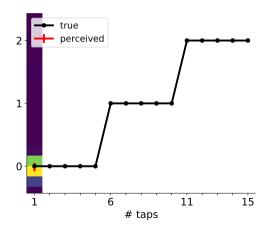
イロト 不得下 イヨト イヨ



Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

January 6, 2020 19 / 20

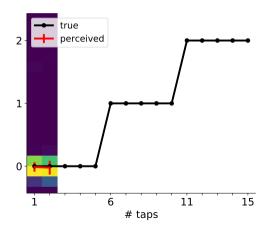
э



Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

January 6, 2020 19 / 20

3

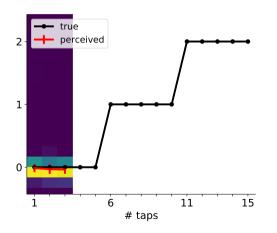


Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

January 6, 2020 19 / 20

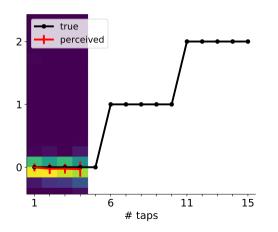
3



Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

January 6, 2020 19 / 20

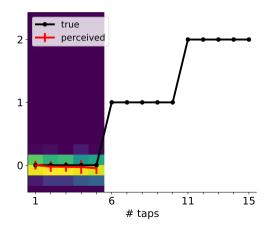
3



Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

January 6, 2020 19 / 20

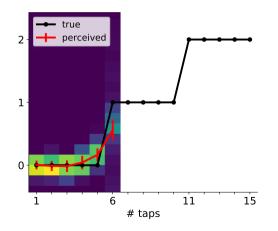
3



Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

January 6, 2020 19 / 20

3

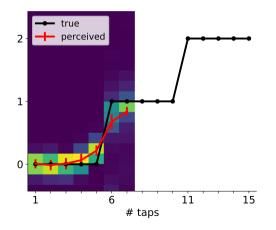


Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

January 6, 2020 19 / 20

3

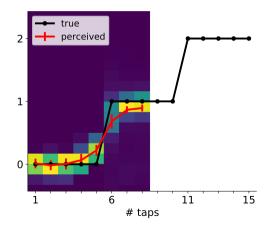


Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

January 6, 2020 19 / 20

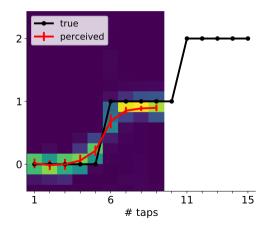
3



Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

January 6, 2020 19 / 20

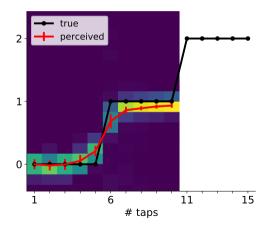
3



Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

January 6, 2020 19 / 20

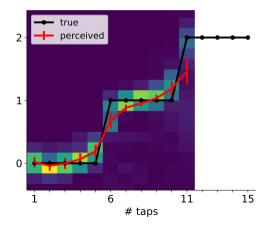
3



Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

January 6, 2020 19 / 20

3

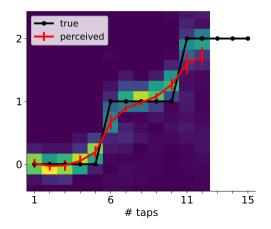


Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

January 6, 2020 19 / 20

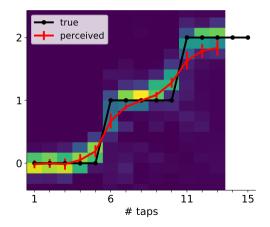
3



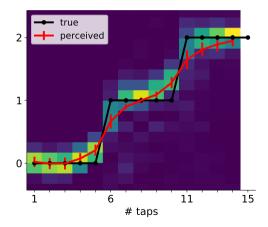
Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

January 6, 2020 19 / 20

3

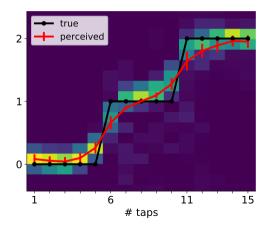


3



January 6, 2020 19 / 20

3

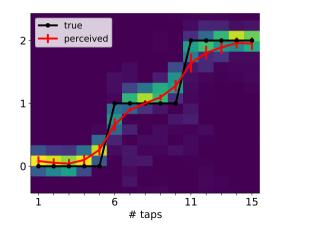


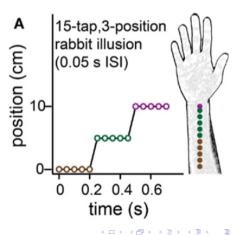
Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

January 6, 2020 19 / 20

3





Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Representation: DDC $\boldsymbol{r}_t := \mathbb{E}_{q(\boldsymbol{z}_{1:t} | \boldsymbol{x}_{1:t})} \left[\boldsymbol{\psi}_t(\boldsymbol{z}_{1:t})
ight]$

, or linear $\mathbb{E}_q\left[h(oldsymbol{z}_{t- au})
ight] pprox oldsymbol{lpha}^{\intercal} oldsymbol{r}_t$

Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

January 6, 2020 20 / 20

Representation: DDC $\boldsymbol{r}_t := \mathbb{E}_{q(\boldsymbol{z}_{1:t}|\boldsymbol{x}_{1:t})} [\psi_t(\boldsymbol{z}_{1:t})]$ Computation: bilinear $\boldsymbol{r}_t := \boldsymbol{W}^* \cdot (\boldsymbol{r}_{t-1} \otimes \sigma(\boldsymbol{x}_t))$, or linear $\mathbb{E}_q [h(\boldsymbol{z}_{t-\tau})] \approx \boldsymbol{\alpha}^{\mathsf{T}} \boldsymbol{r}_t$

Representation: DDC $\boldsymbol{r}_t := \mathbb{E}_{q(\boldsymbol{z}_{1:t}|\boldsymbol{x}_{1:t})} [\boldsymbol{\psi}_t(\boldsymbol{z}_{1:t})]$ Computation: bilinear $\boldsymbol{r}_t := \boldsymbol{W}^* \cdot (\boldsymbol{r}_{t-1} \otimes \boldsymbol{\sigma}(\boldsymbol{x}_t))$, or linear $\mathbb{E}_q [h(\boldsymbol{z}_{t-\tau})] \approx \boldsymbol{\alpha}^{\mathsf{T}} \boldsymbol{r}_t$ Learning to infer: delta rule $\Delta \boldsymbol{W} \propto (\boldsymbol{\psi}_t - \boldsymbol{\phi}_{\boldsymbol{W}})(\boldsymbol{r}_{t-1} \otimes \boldsymbol{\sigma}_t)$, similar for readout $\boldsymbol{\alpha}$

Kevin Li, Maneesh Sahani (Gatsby Unit, UCL)

Model of recognition and postdiction

January 6, 2020 20 / 20

Representation: DDC $r_t := \mathbb{E}_{q(\boldsymbol{z}_{1:t}|\boldsymbol{x}_{1:t})} [\psi_t(\boldsymbol{z}_{1:t})]$ Computation: bilinear $r_t := W^* \cdot (\boldsymbol{r}_{t-1} \otimes \sigma(\boldsymbol{x}_t))$, or linear $\mathbb{E}_q [h(\boldsymbol{z}_{t-\tau})] \approx \boldsymbol{\alpha}^{\mathsf{T}} \boldsymbol{r}_t$ Learning to infer: delta rule $\Delta W \propto (\psi_t - \phi_W)(\boldsymbol{r}_{t-1} \otimes \sigma_t)$, similar for readout $\boldsymbol{\alpha}$ Questions:

Representation: DDC $r_t := \mathbb{E}_{q(\boldsymbol{z}_{1:t}|\boldsymbol{x}_{1:t})} [\psi_t(\boldsymbol{z}_{1:t})]$ Computation: bilinear $r_t := \boldsymbol{W}^* \cdot (\boldsymbol{r}_{t-1} \otimes \sigma(\boldsymbol{x}_t))$, or linear $\mathbb{E}_q [h(\boldsymbol{z}_{t-\tau})] \approx \boldsymbol{\alpha}^{\mathsf{T}} \boldsymbol{r}_t$ Learning to infer: delta rule $\Delta \boldsymbol{W} \propto (\psi_t - \phi_{\boldsymbol{W}})(\boldsymbol{r}_{t-1} \otimes \sigma_t)$, similar for readout $\boldsymbol{\alpha}$ Questions:

• Is the encoding function $oldsymbol{\psi}_t = oldsymbol{U} oldsymbol{\psi}_{t-1} + oldsymbol{\gamma}(oldsymbol{z}_t)$ the best?

Representation: DDC $r_t := \mathbb{E}_{q(\boldsymbol{z}_{1:t}|\boldsymbol{x}_{1:t})} [\psi_t(\boldsymbol{z}_{1:t})]$ Computation: bilinear $r_t := \boldsymbol{W}^* \cdot (\boldsymbol{r}_{t-1} \otimes \sigma(\boldsymbol{x}_t))$, or linear $\mathbb{E}_q [h(\boldsymbol{z}_{t-\tau})] \approx \boldsymbol{\alpha}^{\mathsf{T}} \boldsymbol{r}_t$ Learning to infer: delta rule $\Delta \boldsymbol{W} \propto (\psi_t - \phi_{\boldsymbol{W}})(\boldsymbol{r}_{t-1} \otimes \sigma_t)$, similar for readout $\boldsymbol{\alpha}$ Questions:

- Is the encoding function $oldsymbol{\psi}_t = oldsymbol{U} oldsymbol{\psi}_{t-1} + oldsymbol{\gamma}(oldsymbol{z}_t)$ the best?
- Does the brain encode the joint $q(\boldsymbol{z}_{1:t}|\boldsymbol{x}_{1:t})$?

Representation: DDC $r_t := \mathbb{E}_{q(\boldsymbol{z}_{1:t}|\boldsymbol{x}_{1:t})} [\psi_t(\boldsymbol{z}_{1:t})]$ Computation: bilinear $r_t := \boldsymbol{W}^* \cdot (\boldsymbol{r}_{t-1} \otimes \sigma(\boldsymbol{x}_t))$, or linear $\mathbb{E}_q [h(\boldsymbol{z}_{t-\tau})] \approx \boldsymbol{\alpha}^{\mathsf{T}} \boldsymbol{r}_t$ Learning to infer: delta rule $\Delta \boldsymbol{W} \propto (\psi_t - \phi_{\boldsymbol{W}})(\boldsymbol{r}_{t-1} \otimes \sigma_t)$, similar for readout $\boldsymbol{\alpha}$ Questions:

- Is the encoding function $oldsymbol{\psi}_t = oldsymbol{U} oldsymbol{\psi}_{t-1} + oldsymbol{\gamma}(oldsymbol{z}_t)$ the best?
- Does the brain encode the joint $q(\boldsymbol{z}_{1:t}|\boldsymbol{x}_{1:t})$?
- Any theoretical argument for using the bilinear rule?

Representation: DDC $r_t := \mathbb{E}_{q(\boldsymbol{z}_{1:t}|\boldsymbol{x}_{1:t})} [\psi_t(\boldsymbol{z}_{1:t})]$ Computation: bilinear $r_t := \boldsymbol{W}^* \cdot (\boldsymbol{r}_{t-1} \otimes \sigma(\boldsymbol{x}_t))$, or linear $\mathbb{E}_q [h(\boldsymbol{z}_{t-\tau})] \approx \boldsymbol{\alpha}^{\mathsf{T}} \boldsymbol{r}_t$ Learning to infer: delta rule $\Delta \boldsymbol{W} \propto (\psi_t - \phi_{\boldsymbol{W}})(\boldsymbol{r}_{t-1} \otimes \sigma_t)$, similar for readout $\boldsymbol{\alpha}$ Questions:

- Is the encoding function $oldsymbol{\psi}_t = oldsymbol{U} oldsymbol{\psi}_{t-1} + oldsymbol{\gamma}(oldsymbol{z}_t)$ the best?
- Does the brain encode the joint $q(\boldsymbol{z}_{1:t}|\boldsymbol{x}_{1:t})$?
- Any theoretical argument for using the bilinear rule?
- ullet Is there a way to also adapt $oldsymbol{U}$ in a plausible way?

Representation: DDC $r_t := \mathbb{E}_{q(\boldsymbol{z}_{1:t}|\boldsymbol{x}_{1:t})} [\psi_t(\boldsymbol{z}_{1:t})]$ Computation: bilinear $r_t := \boldsymbol{W}^* \cdot (\boldsymbol{r}_{t-1} \otimes \sigma(\boldsymbol{x}_t))$, or linear $\mathbb{E}_q [h(\boldsymbol{z}_{t-\tau})] \approx \boldsymbol{\alpha}^{\mathsf{T}} \boldsymbol{r}_t$ Learning to infer: delta rule $\Delta \boldsymbol{W} \propto (\psi_t - \phi_{\boldsymbol{W}})(\boldsymbol{r}_{t-1} \otimes \sigma_t)$, similar for readout $\boldsymbol{\alpha}$ Questions:

- Is the encoding function $oldsymbol{\psi}_t = oldsymbol{U} oldsymbol{\psi}_{t-1} + oldsymbol{\gamma}(oldsymbol{z}_t)$ the best?
- Does the brain encode the joint $q(\boldsymbol{z}_{1:t}|\boldsymbol{x}_{1:t})$?
- Any theoretical argument for using the bilinear rule?
- Is there a way to also adapt U in a plausible way?
- Can we encode the internal model by DDC? (talk to Eszter Vértes)