CSCI567 Machine Learning (Fall 2025)
Viterbi Algorithm Exercises

Haipeng Luo

University of Southern California

Oct 31, 2025

Exercise 1

What is the most likely sequence $z^*_{1:T_0}$ given $x_{1:T_0}$ for some $T_0 < T$?

Exercise 1

What is the most likely sequence $z^*_{1:T_0}$ given $x_{1:T_0}$ for some $T_0 < T$?
(imagine you only hear the first few words of a sentence)

Exercise 1

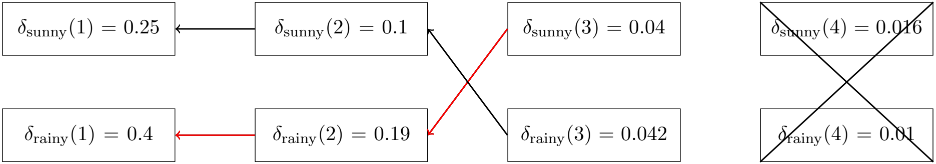
What is the most likely sequence $z^*_{1:T_0}$ given $x_{1:T_0}$ for some $T_0 < T$? (imagine you only hear the first few words of a sentence)

- Is it the first T_0 outputs of the Viterbi algorithm (with all data)?

Exercise 1

What is the most likely sequence $z^*_{1:T_0}$ given $x_{1:T_0}$ for some $T_0 < T$?
(imagine you only hear the first few words of a sentence)

- Is it the first T_0 outputs of the Viterbi algorithm (with all data)?



The answer for $T_0 = 3$ is: rainy, rainy, sunny?

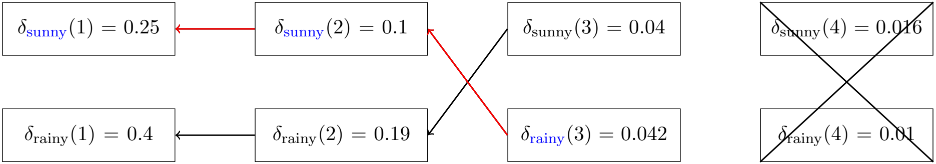
Exercise 1

What is the most likely sequence $z^*_{1:T_0}$ given $x_{1:T_0}$ for some $T_0 < T$? (imagine you only hear the first few words of a sentence)

- Is it the first T_0 outputs of the Viterbi algorithm (with all data)?

No. It should be

- $z^*_{T_0} = \operatorname{argmax}_s \delta_s(T_0)$
- for each $t = T_0, \ldots, 2$: $z^*_{t-1} = \Delta_{z^*_t}(t)$



The answer for $T_0 = 3$ is: **"sunny, sunny, rainy"**

Exercise 2

What is the most likely sequence $z^*_{1:T_0}$ given $x_{1:T}$ for some $T_0 < T$?

Exercise 2

What is the most likely sequence $z^*_{1:T_0}$ given $x_{1:T}$ for some $T_0 < T$?
(imagine you hear a sentence but only care about the first few words)

Exercise 2

What is the most likely sequence $z^*_{1:T_0}$ given $x_{1:T}$ for some $T_0 < T$?
(imagine you hear a sentence but only care about the first few words)

- Is it the same as Exercise 1?

Exercise 2

What is the most likely sequence $z^*_{1:T_0}$ given $x_{1:T}$ for some $T_0 < T$?
(imagine you hear a sentence but only care about the first few words)

- Is it the same as Exercise 1?
- Is it the first T_0 outputs of the Viterbi algorithm (with all data)?

Exercise 2

What is the most likely sequence $z^*_{1:T_0}$ given $x_{1:T}$ for some $T_0 < T$?
(imagine you hear a sentence but only care about the first few words)

- Is it the same as Exercise 1?
- Is it the first T_0 outputs of the Viterbi algorithm (with all data)?

Neither. It should be

- $z^*_{T_0} = \mathrm{argmax}_s \, \delta_s(T_0)\beta_s(T_0)$
- for each $t = T_0, \ldots, 2$: $z^*_{t-1} = \Delta_{z^*_t}(t)$

Exercise 2 (cont.)

Reasoning:

$$z^*_{T_0} = \operatorname*{argmax}_{s} \max_{z_{1:T_0-1}} P(Z_{T_0} = s, Z_{1:T_0-1} = z_{1:T_0-1}, X_{1:T} = x_{1:T})$$

Exercise 2 (cont.)

Reasoning:

$$
\begin{aligned}
z^*_{T_0} &= \operatorname*{argmax}_s \max_{z_{1:T_0-1}} P(Z_{T_0} = s, Z_{1:T_0-1} = z_{1:T_0-1}, X_{1:T} = x_{1:T}) \\
&= \operatorname*{argmax}_s \max_{z_{1:T_0-1}} P(Z_{T_0} = s, Z_{1:T_0-1} = z_{1:T_0-1}, X_{1:T_0} = x_{1:T_0}) \times \\
&\qquad P(X_{T_0+1,T} = x_{T_0+1:T} \mid Z_{T_0} = s, Z_{1:T_0-1} = z_{1:T_0-1}, X_{1:T_0} = x_{1:T_0})
\end{aligned}
$$

Exercise 2 (cont.)

Reasoning:

$$
\begin{aligned}
z^*_{T_0} &= \operatorname*{argmax}_s \max_{z_{1:T_0-1}} P(Z_{T_0} = s, Z_{1:T_0-1} = z_{1:T_0-1}, X_{1:T} = x_{1:T}) \\
&= \operatorname*{argmax}_s \max_{z_{1:T_0-1}} P(Z_{T_0} = s, Z_{1:T_0-1} = z_{1:T_0-1}, X_{1:T_0} = x_{1:T_0}) \times \\
&\qquad P(X_{T_0+1,T} = x_{T_0+1:T} \mid Z_{T_0} = s, Z_{1:T_0-1} = z_{1:T_0-1}, X_{1:T_0} = x_{1:T_0}) \\
&= \operatorname*{argmax}_s \left(\max_{z_{1:T_0-1}} P(Z_{T_0} = s, Z_{1:T_0-1} = z_{1:T_0-1}, X_{1:T_0} = x_{1:T_0}) \right) \times \\
&\qquad P(X_{T_0+1,T} = x_{T_0+1:T} \mid Z_{T_0} = s)
\end{aligned}
$$

Exercise 2 (cont.)

Reasoning:

$$
\begin{aligned}
z^*_{T_0} &= \operatorname*{argmax}_s \max_{z_{1:T_0-1}} P(Z_{T_0} = s, Z_{1:T_0-1} = z_{1:T_0-1}, X_{1:T} = x_{1:T}) \\
&= \operatorname*{argmax}_s \max_{z_{1:T_0-1}} P(Z_{T_0} = s, Z_{1:T_0-1} = z_{1:T_0-1}, X_{1:T_0} = x_{1:T_0}) \times \\
&\qquad P(X_{T_0+1,T} = x_{T_0+1:T} \mid Z_{T_0} = s, Z_{1:T_0-1} = z_{1:T_0-1}, X_{1:T_0} = x_{1:T_0}) \\
&= \operatorname*{argmax}_s \left(\max_{z_{1:T_0-1}} P(Z_{T_0} = s, Z_{1:T_0-1} = z_{1:T_0-1}, X_{1:T_0} = x_{1:T_0}) \right) \times \\
&\qquad P(X_{T_0+1,T} = x_{T_0+1:T} \mid Z_{T_0} = s) \\
&= \operatorname*{argmax}_s \ \delta_s(T_0)\beta_s(T_0)
\end{aligned}
$$

Exercise 3

What is the most likely sequence $z^*_{1:T}$ given $x_{1:T_0}$ for some $T_0 < T$?

Exercise 3

What is the most likely sequence $z^*_{1:T}$ given $x_{1:T_0}$ for some $T_0 < T$? (only hear the first few words but want to recover the full sentence)

Exercise 3

What is the most likely sequence $z^*_{1:T}$ given $x_{1:T_0}$ for some $T_0 < T$? (only hear the first few words but want to recover the full sentence)

- Is it the same as the Viterbi algorithm (with all data)?

Exercise 3

What is the most likely sequence $z^*_{1:T}$ given $x_{1:T_0}$ for some $T_0 < T$? (only hear the first few words but want to recover the full sentence)

- Is it the same as the Viterbi algorithm (with all data)?
- Are the first T_0 states the same as Exercise 1?

Exercise 3

What is the most likely sequence $z^*_{1:T}$ given $x_{1:T_0}$ for some $T_0 < T$? (only hear the first few words but want to recover the full sentence)

- Is it the same as the Viterbi algorithm (with all data)?
- Are the first T_0 states the same as Exercise 1?

Again, neither is true.

Exercise 3 (cont.)

Viterbi Algorithm with partial data $x_{1:T_0}$

For each $s \in [S]$, compute $\delta_s(1) = \pi_s b_{s,x_1}$.

Exercise 3 (cont.)

Viterbi Algorithm with partial data $x_{1:T_0}$

For each $s \in [S]$, compute $\delta_s(1) = \pi_s b_{s,x_1}$.

For each $t = 2, \ldots, T$,

- for each $s \in [S]$, compute

$$\delta_s(t) = \begin{cases} b_{s,x_t} \max_{s'} a_{s',s}\delta_{s'}(t-1) & \text{if } t \leq T_0 \\ & \end{cases}$$

Exercise 3 (cont.)

Viterbi Algorithm with partial data $x_{1:T_0}$

For each $s \in [S]$, compute $\delta_s(1) = \pi_s b_{s,x_1}$.

For each $t = 2, \ldots, T$,

- for each $s \in [S]$, compute

$$\delta_s(t) = \begin{cases} b_{s,x_t} \max_{s'} a_{s',s}\delta_{s'}(t-1) & \text{if } t \leq T_0 \\ \max_{s'} a_{s',s}\delta_{s'}(t-1) & \text{else} \end{cases}$$

Exercise 3 (cont.)

Viterbi Algorithm with partial data $x_{1:T_0}$

For each $s \in [S]$, compute $\delta_s(1) = \pi_s b_{s,x_1}$.

For each $t = 2, \ldots, T$,

- for each $s \in [S]$, compute

$$\delta_s(t) = \begin{cases} b_{s,x_t} \max_{s'} a_{s',s}\delta_{s'}(t-1) & \text{if } t \leq T_0 \\ \max_{s'} a_{s',s}\delta_{s'}(t-1) & \text{else} \end{cases}$$

$$\Delta_s(t) = \operatorname*{argmax}_{s'} a_{s',s}\delta_{s'}(t-1).$$

Exercise 3 (cont.)

Viterbi Algorithm with partial data $x_{1:T_0}$

For each $s \in [S]$, compute $\delta_s(1) = \pi_s b_{s,x_1}$.

For each $t = 2, \ldots, T$,

- for each $s \in [S]$, compute

$$\delta_s(t) = \begin{cases} b_{s,x_t} \max_{s'} a_{s',s} \delta_{s'}(t-1) & \text{if } t \leq T_0 \\ \max_{s'} a_{s',s} \delta_{s'}(t-1) & \text{else} \end{cases}$$

$$\Delta_s(t) = \operatorname*{argmax}_{s'} a_{s',s} \delta_{s'}(t-1).$$

Backtracking: let $z_T^* = \operatorname{argmax}_s \delta_s(T)$.

Exercise 3 (cont.)

Viterbi Algorithm with partial data $x_{1:T_0}$

For each $s \in [S]$, compute $\delta_s(1) = \pi_s b_{s,x_1}$.

For each $t = 2, \ldots, T$,

- for each $s \in [S]$, compute

$$\delta_s(t) = \begin{cases} b_{s,x_t} \max_{s'} a_{s',s} \delta_{s'}(t-1) & \text{if } t \leq T_0 \\ \max_{s'} a_{s',s} \delta_{s'}(t-1) & \text{else} \end{cases}$$

$$\Delta_s(t) = \underset{s'}{\operatorname{argmax}}\, a_{s',s} \delta_{s'}(t-1).$$

Backtracking: let $z_T^* = \operatorname{argmax}_s \delta_s(T)$.
For each $t = T, \ldots, 2$: set $z_{t-1}^* = \Delta_{z_t^*}(t)$.

Output the most likely path $z_1^*, \ldots, z_T^*$.

Recap

Difference compared to "most likely path $z^*_{1:T}$ given $x_{1:T}$":

Recap

Difference compared to "most likely path $z^*_{1:T}$ given $x_{1:T}$":

- most likely path $z^*_{1:T_0}$ given $x_{1:T_0}$

Recap

Difference compared to "most likely path $z^*_{1:T}$ given $x_{1:T}$":

- most likely path $z^*_{1:T_0}$ given $x_{1:T_0}$
  - exact same δ, Δ
  - find the last state using $\delta_s(T_0)$

Recap

Difference compared to "most likely path $z^*_{1:T}$ given $x_{1:T}$":

- most likely path $z^*_{1:T_0}$ given $x_{1:T_0}$
  - exact same δ, Δ
  - find the last state using $\delta_s(T_0)$
- most likely path $z^*_{1:T_0}$ given $x_{1:T}$

Recap

Difference compared to "most likely path $z^*_{1:T}$ given $x_{1:T}$":

- most likely path $z^*_{1:T_0}$ given $x_{1:T_0}$
  - exact same δ, Δ
  - find the last state using $\delta_s(T_0)$
- most likely path $z^*_{1:T_0}$ given $x_{1:T}$
  - exact same δ, Δ
  - find the last state using $\delta_s(T_0)\beta_s(T_0)$

Recap

Difference compared to "most likely path $z^*_{1:T}$ given $x_{1:T}$":

- most likely path $z^*_{1:T_0}$ given $x_{1:T_0}$
  - exact same δ, Δ
  - find the last state using $\delta_s(T_0)$
- most likely path $z^*_{1:T_0}$ given $x_{1:T}$
  - exact same δ, Δ
  - find the last state using $\delta_s(T_0)\beta_s(T_0)$
- most likely path $z^*_{1:T}$ given $x_{1:T_0}$

Recap

Difference compared to "most likely path $z^*_{1:T}$ given $x_{1:T}$":

- most likely path $z^*_{1:T_0}$ given $x_{1:T_0}$
  - exact same δ, Δ
  - find the last state using $\delta_s(T_0)$
- most likely path $z^*_{1:T_0}$ given $x_{1:T}$
  - exact same δ, Δ
  - find the last state using $\delta_s(T_0)\beta_s(T_0)$
- most likely path $z^*_{1:T}$ given $x_{1:T_0}$
  - different δ, Δ
  - find the last state using $\delta_s(T)$

Recap

Difference compared to "most likely path $z^*_{1:T}$ given $x_{1:T}$":

- most likely path $z^*_{1:T_0}$ given $x_{1:T_0}$
  - exact same δ, Δ
  - find the last state using $\delta_s(T_0)$
- most likely path $z^*_{1:T_0}$ given $x_{1:T}$
  - exact same δ, Δ
  - find the last state using $\delta_s(T_0)\beta_s(T_0)$
- most likely path $z^*_{1:T}$ given $x_{1:T_0}$
  - different δ, Δ
  - find the last state using $\delta_s(T)$

The rest of the backtracking is always the same: **follow the arrows!**