

Data Mining and Machine Learning

Assignment Project Exam Help

Speech Recognition HMMs –

Viterbi D

<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro

Peter Jančovič

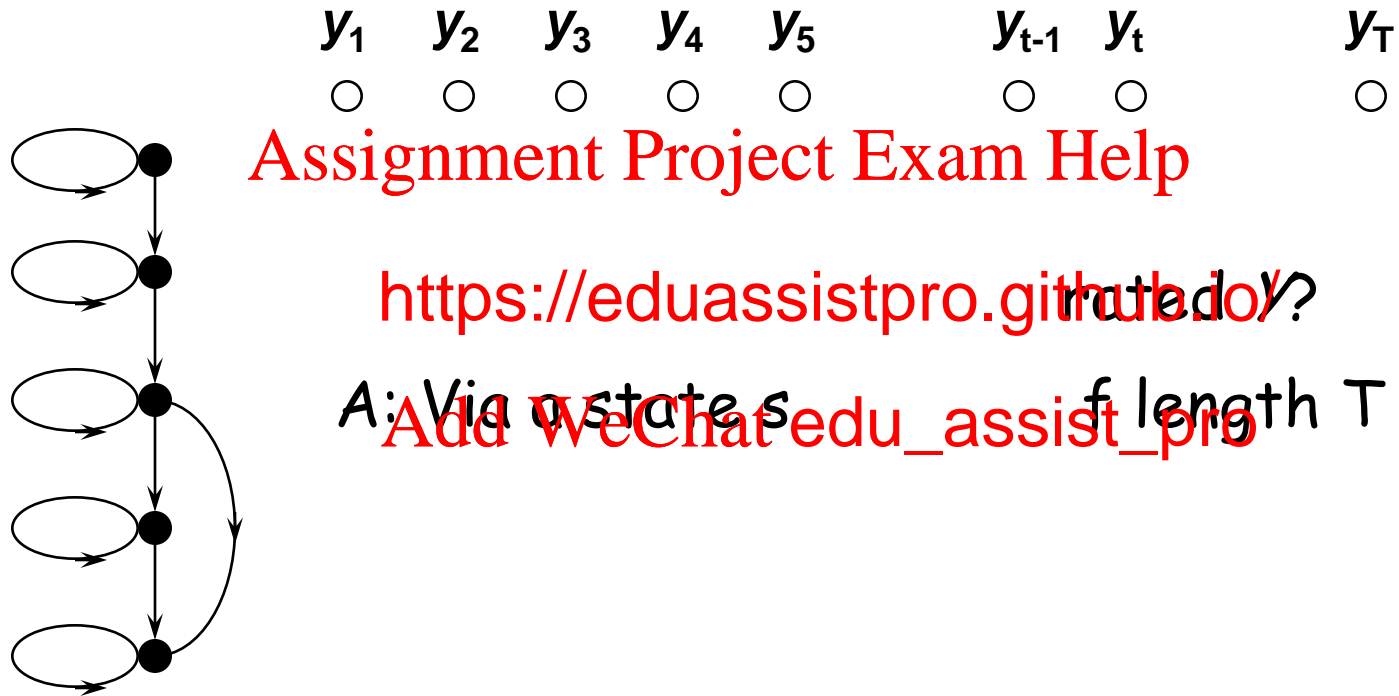


Viterbi Decoding

- Viterbi Decoding is the algorithm which is used to find the sequence of HMM states (or HMMs) which have generated the sequence
- Similar to the Forward probability calculation



Viterbi Decoding (1)



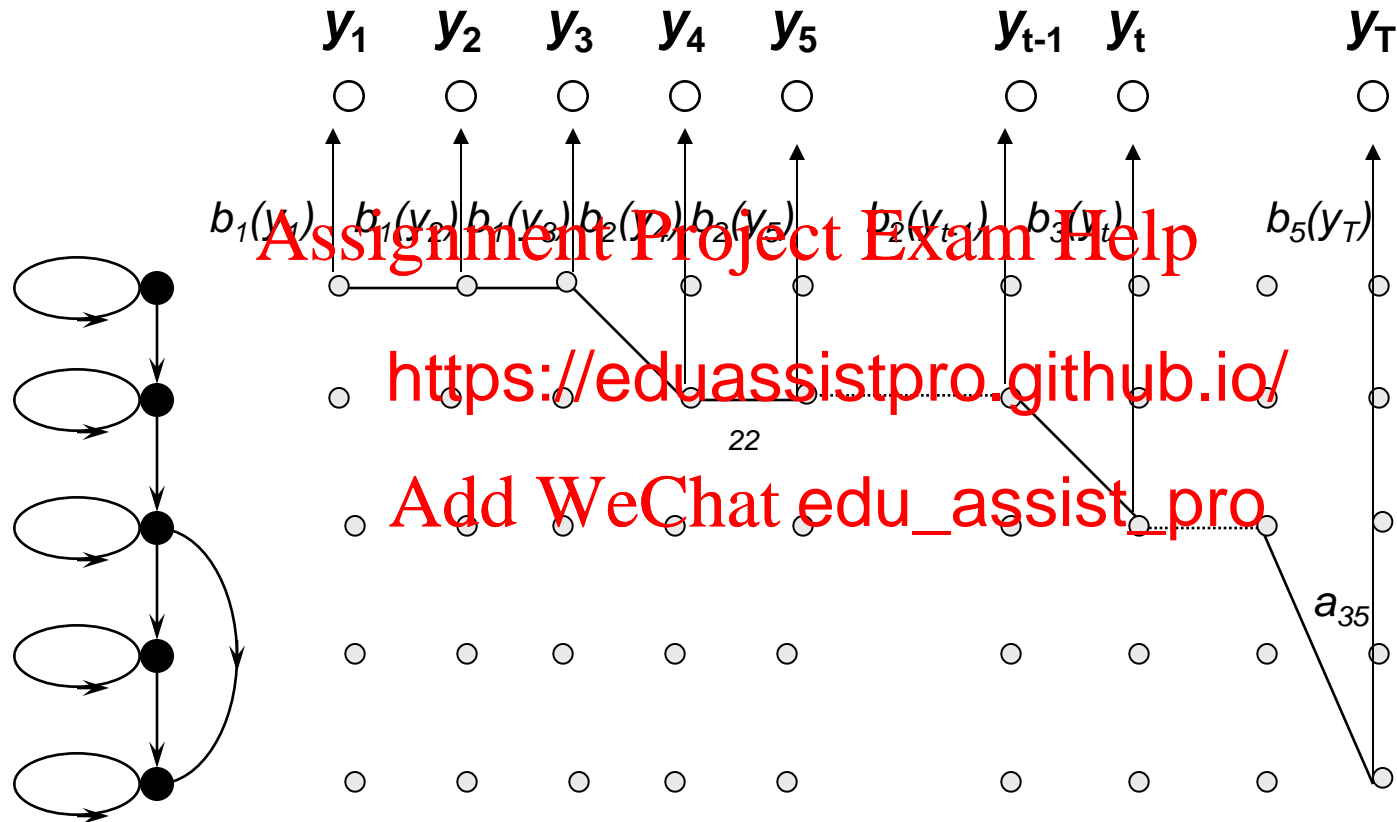
Assignment Project Exam Help

<https://eduassistpro.github.io/>

A: Via a state s of length T

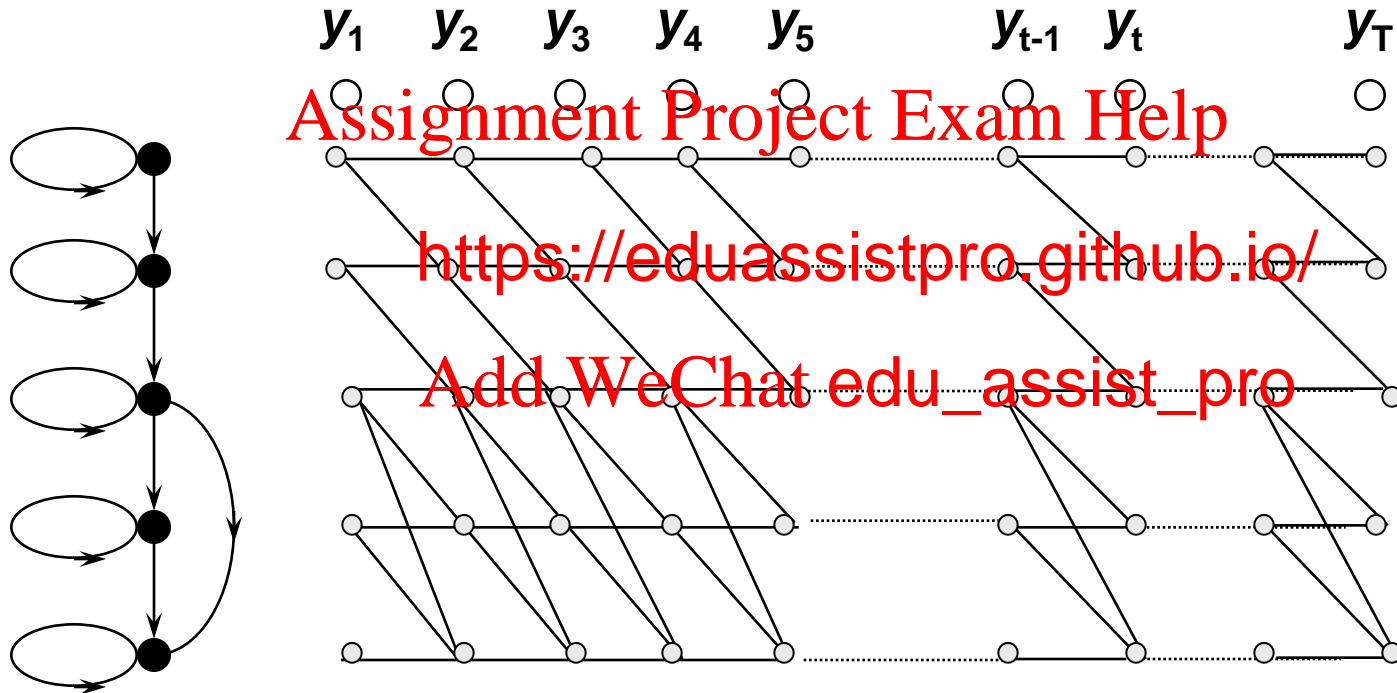


Basic Probability Calculation



Viterbi Decoding (2)

- Construction of 'state-time trellis'



Viterbi Decoding (3)

- Let $X = \{x_1, \dots, x_T\}$ be a state sequence of length T
- The joint probability of Y and X is given by:

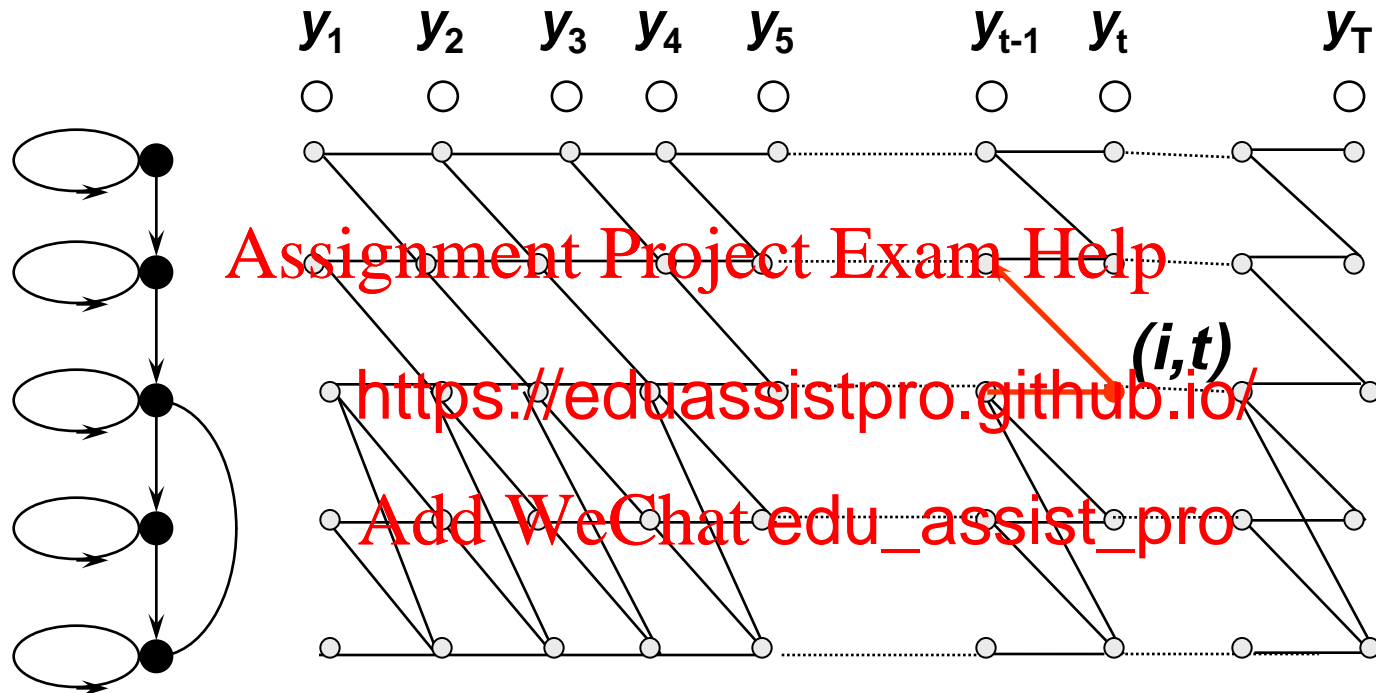
$$p(Y, X) = b_{x_1}(y_1) \prod_{t=2}^T a_{x_{t-1}x_t} b_{x_t}(y_t)$$

<https://eduassistpro.github.io/>

- i.e. the product of the state-out transition probabilities along the state sequence
- The optimal state sequence is the sequence X such that $p(Y, X)$ is maximized
- $p(Y)$ is the sum of $p(Y, X)$ over all sequences X



Viterbi Decoding (4)

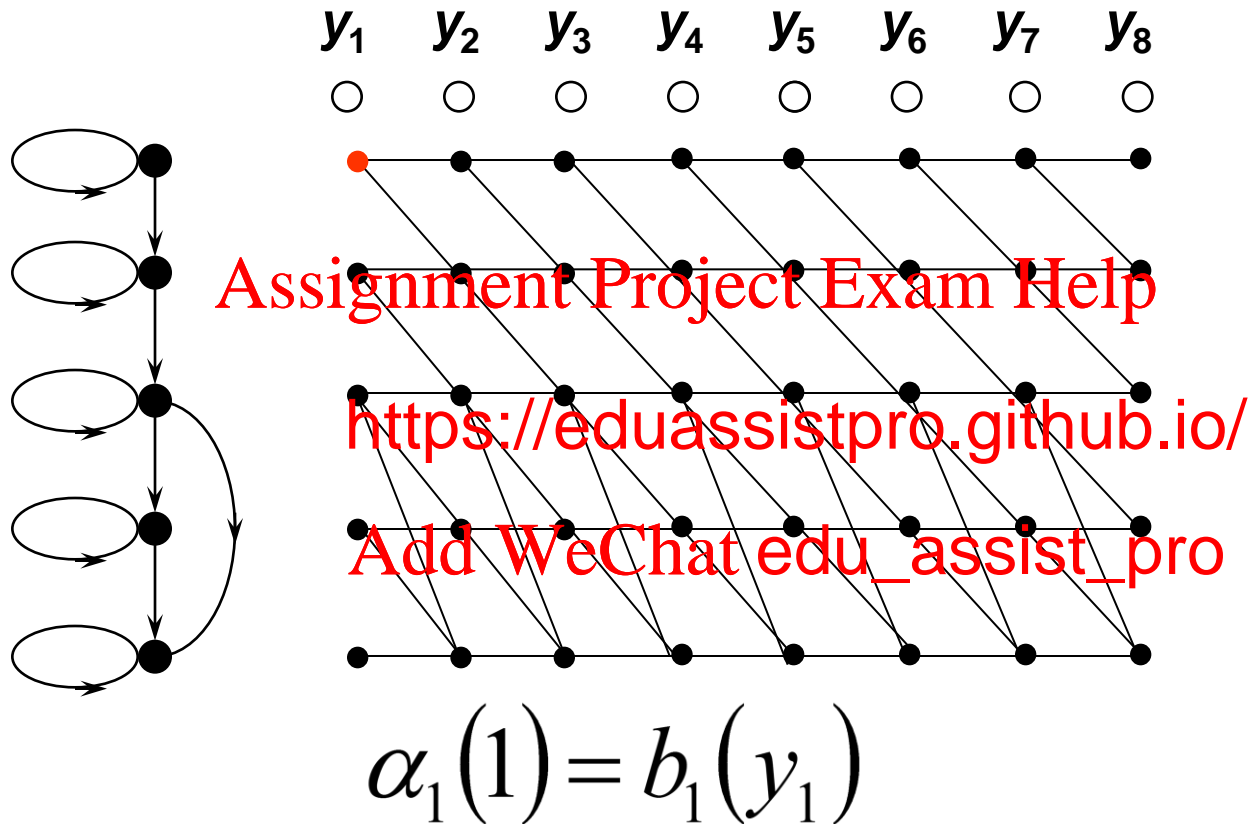


$$\alpha_t(i) = \text{Prob}(y_1, \dots, y_t, \text{opt sequence to } (i, t))$$

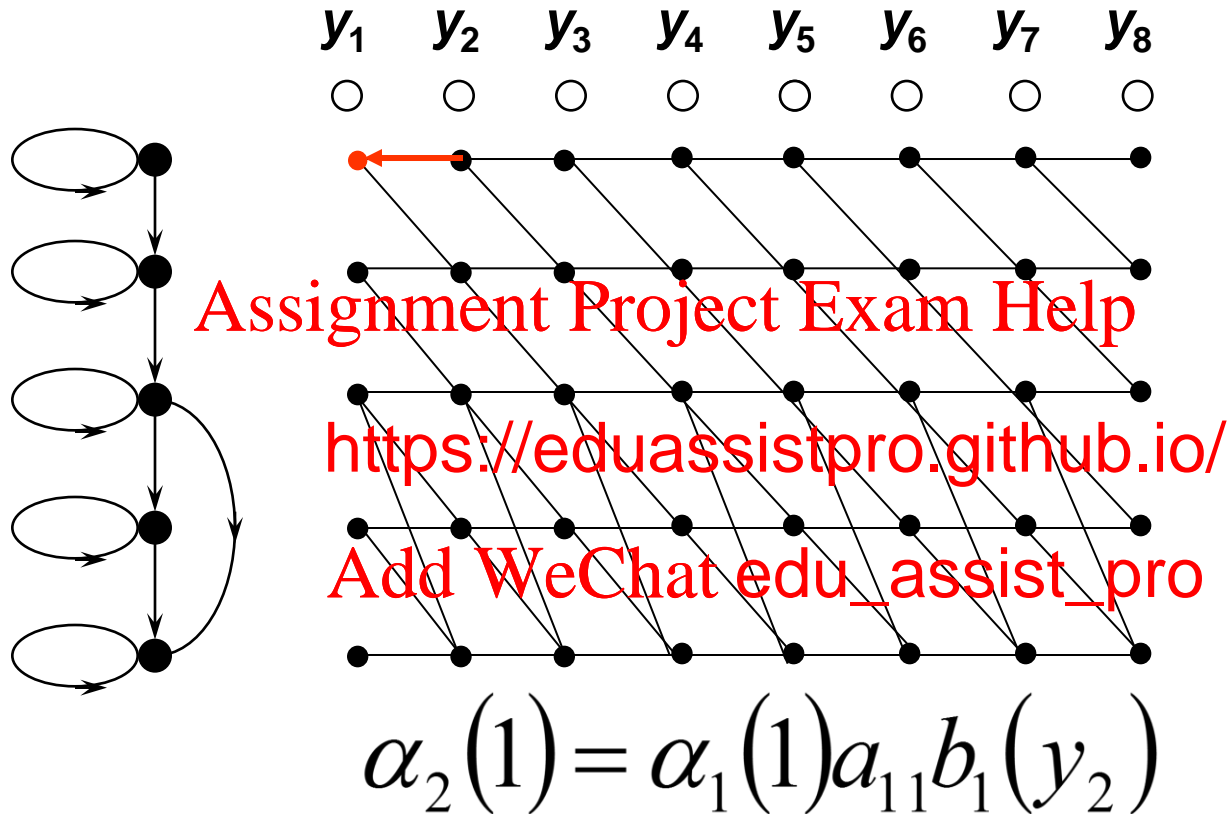
$$\alpha_t(i) = \max \{ \alpha_{t-1}(i-1) a_{i-1,i}, \alpha_{t-1}(i) a_{i,i} \} b_i(y_t)$$



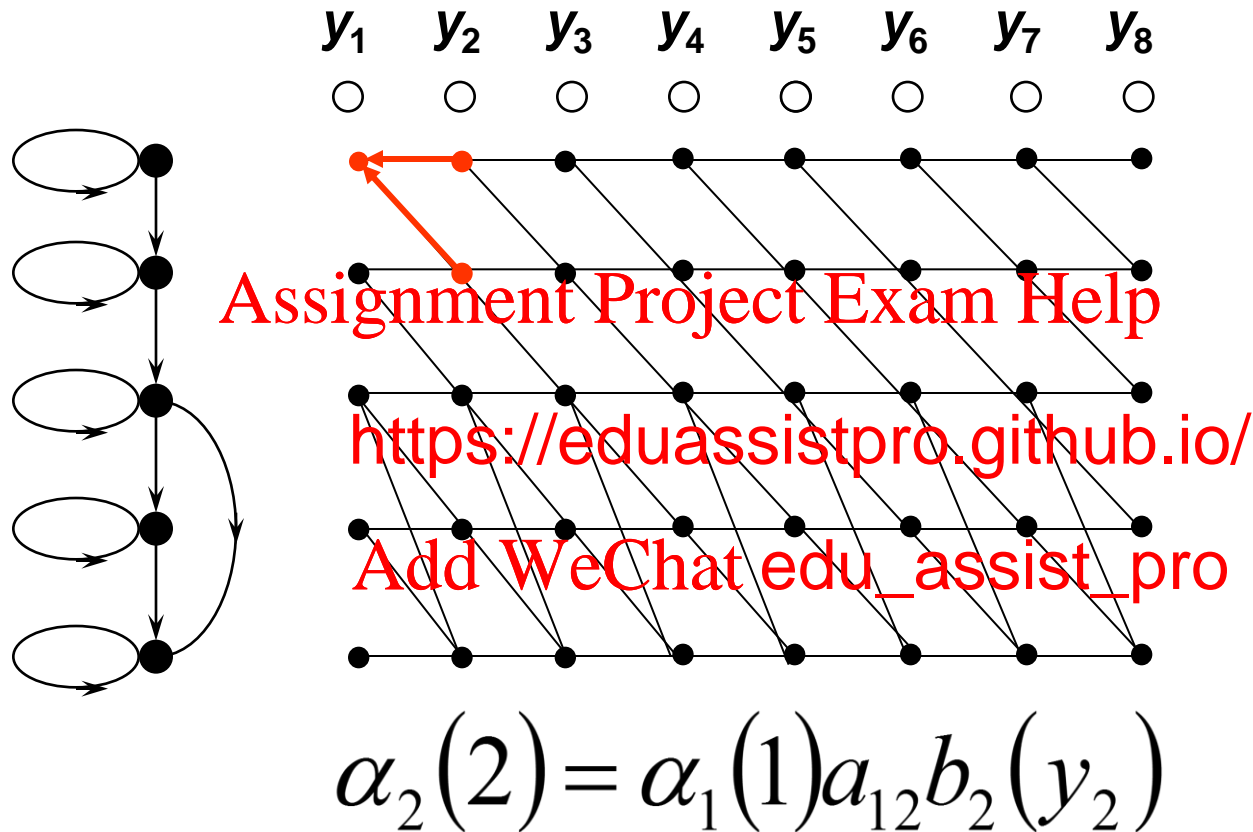
Viterbi Decoding – example



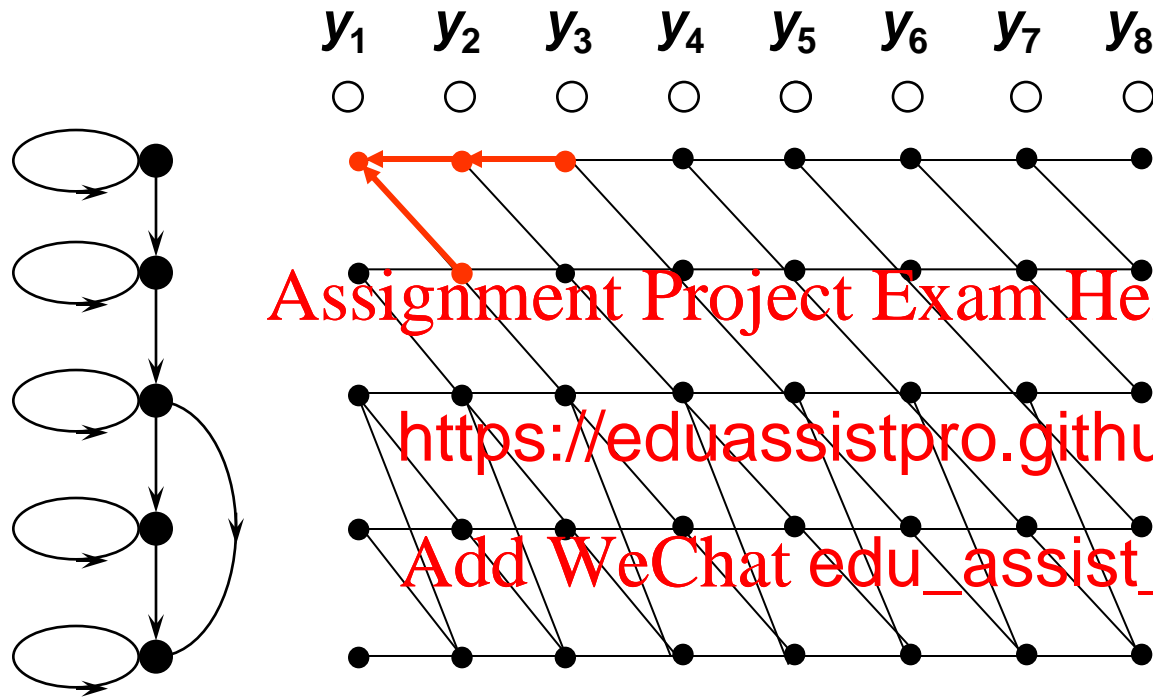
Viterbi Decoding – example



Viterbi Decoding – example



Viterbi Decoding – example



Assignment Project Exam Help

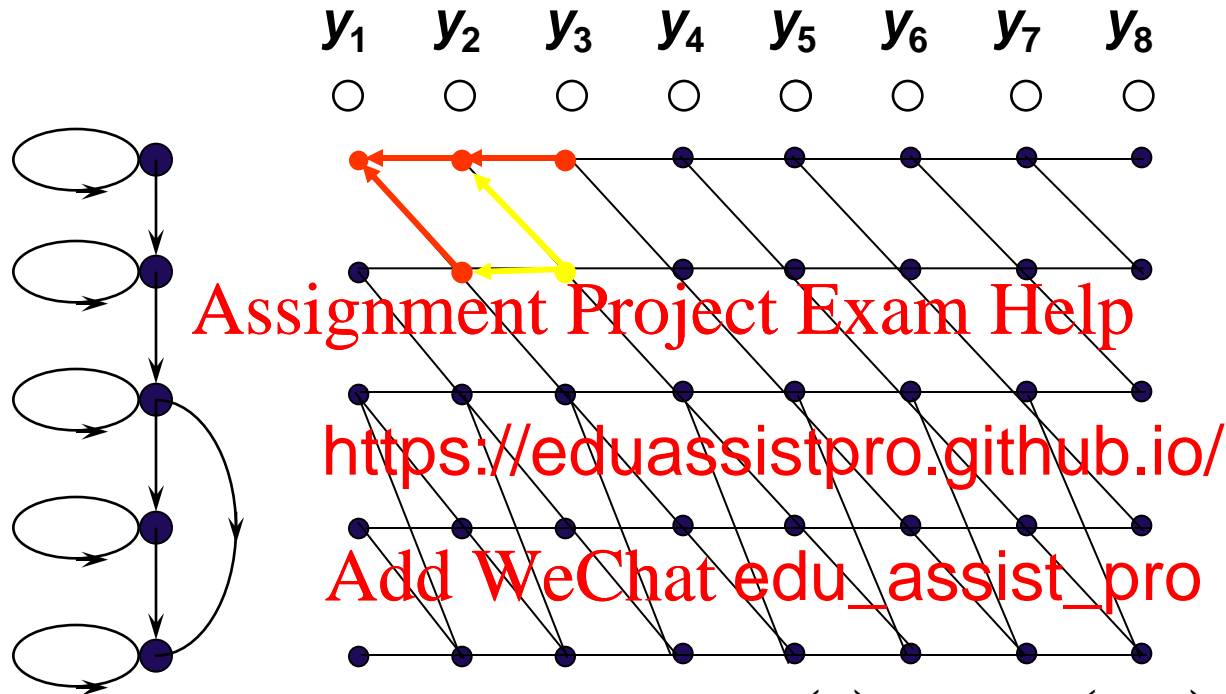
<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro

$$\alpha_3(1) = \alpha_2(1) a_{11} b_1(y_3)$$



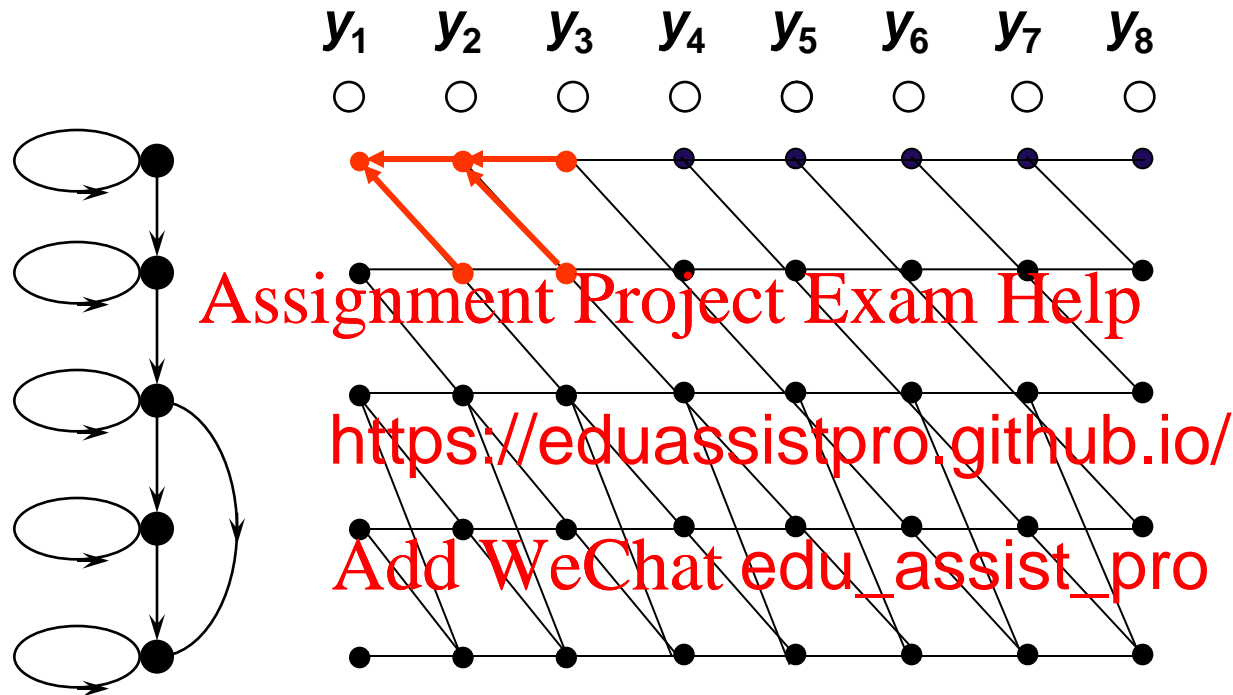
Viterbi Decoding – example



$$\alpha_3(2) = \max \begin{cases} \alpha_2(1) a_{12} b_2(y_3) \\ \alpha_2(2) a_{22} b_2(y_3) \end{cases}$$



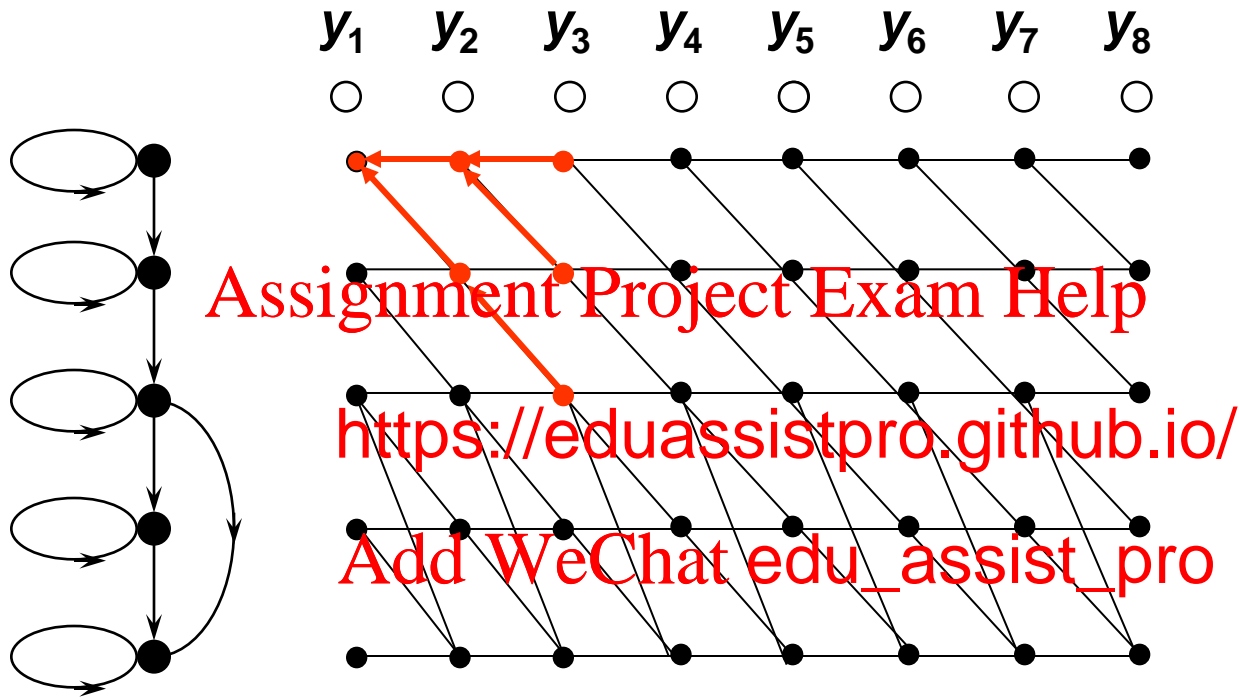
Viterbi Decoding – example



$$\alpha_3(2) = \max \begin{cases} \alpha_2(1) a_{12} b_2(y_3) \\ \alpha_2(2) a_{22} b_2(y_3) \end{cases}$$



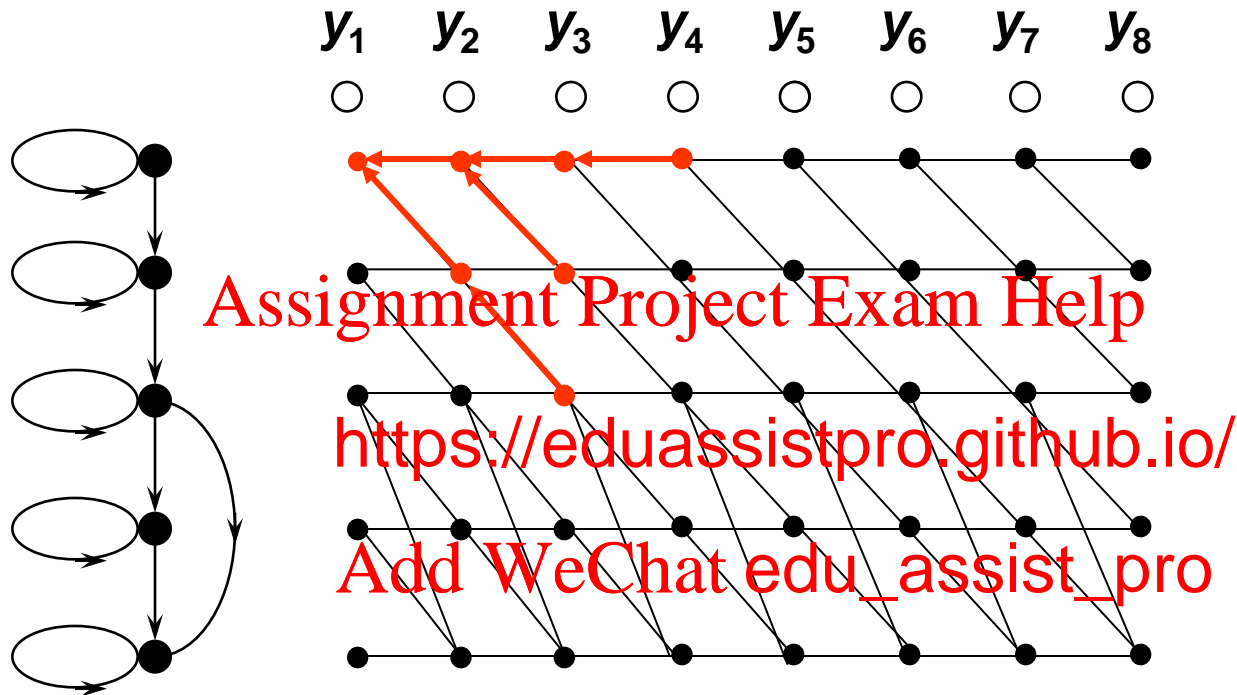
Viterbi Decoding – example



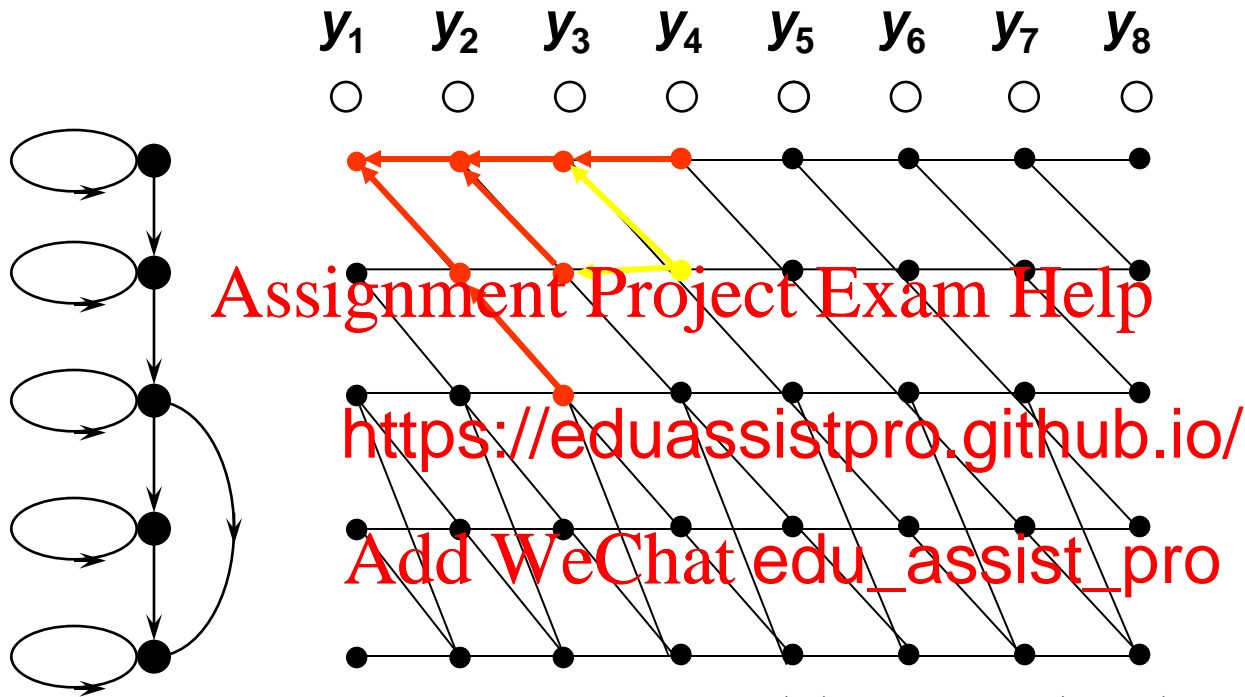
$$\alpha_3(3) = \alpha_2(2) a_{23} b_3(y_3)$$



Viterbi Decoding – example



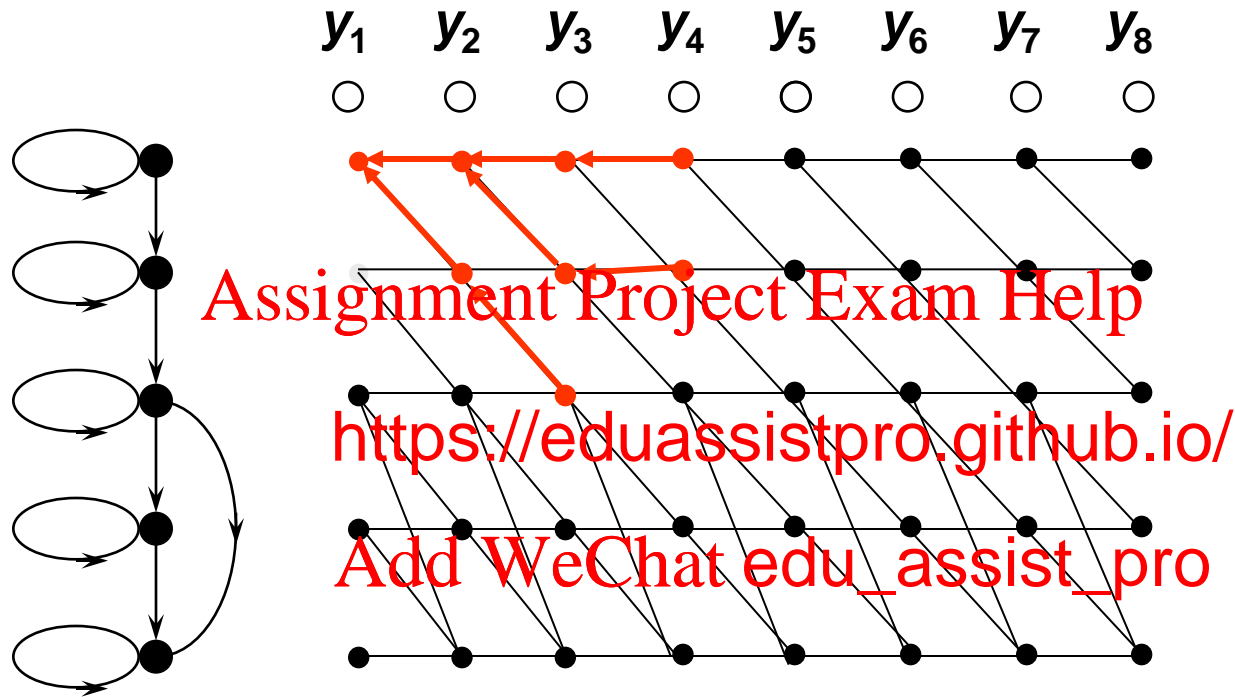
Viterbi Decoding – example



$$\alpha_4(2) = \max \begin{cases} \alpha_3(1) a_{12} b_2(y_4) \\ \alpha_3(2) a_{22} b_2(y_4) \end{cases}$$



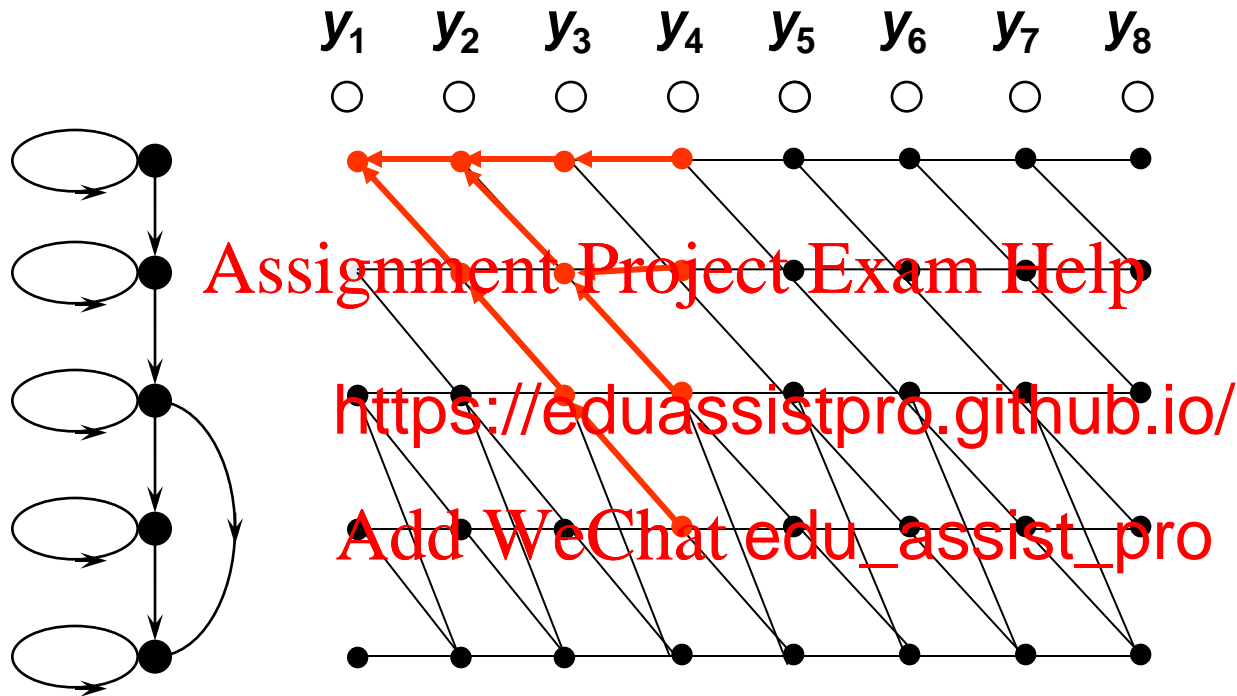
Viterbi Decoding – example



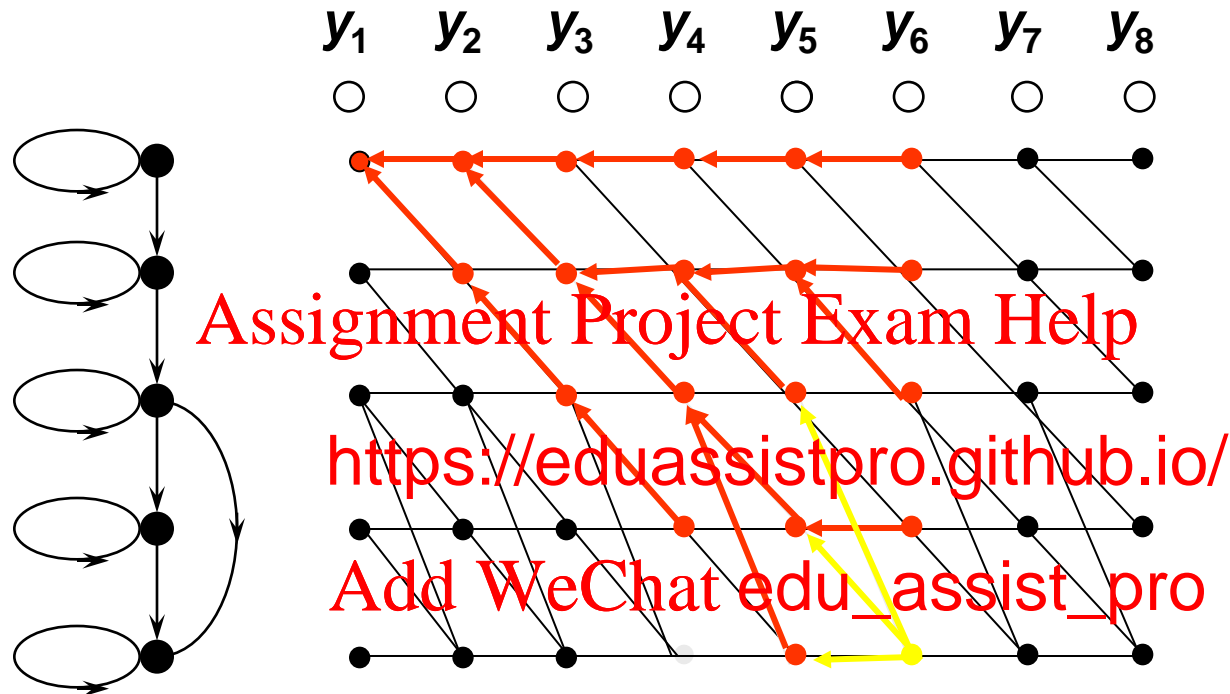
$$\alpha_4(2) = \max \begin{cases} \alpha_3(1) a_{12} b_2(y_4) \\ \alpha_3(2) a_{22} b_2(y_4) \end{cases}$$



Viterbi Decoding – example



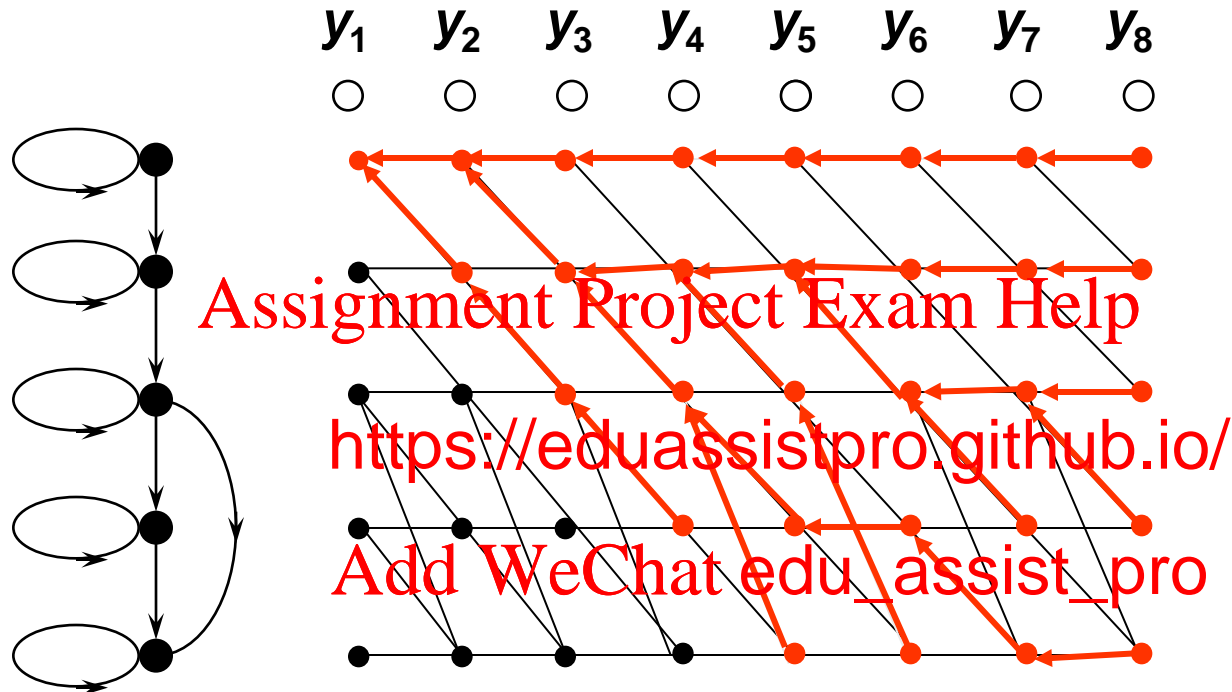
Viterbi Decoding – example



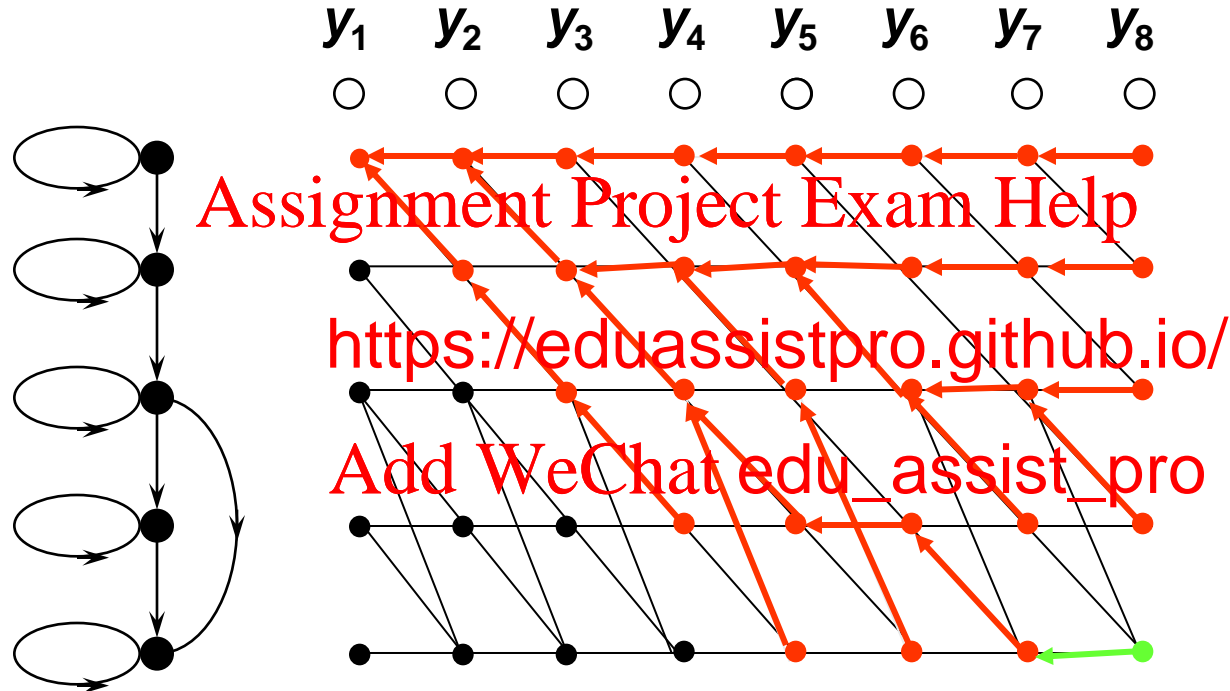
$$\alpha_6(5) = \max \begin{cases} \alpha_5(5) a_{55} b_6(y_6) \\ \alpha_5(4) a_{45} b_6(y_6) \\ \alpha_5(3) a_{35} b_6(y_6) \end{cases}$$



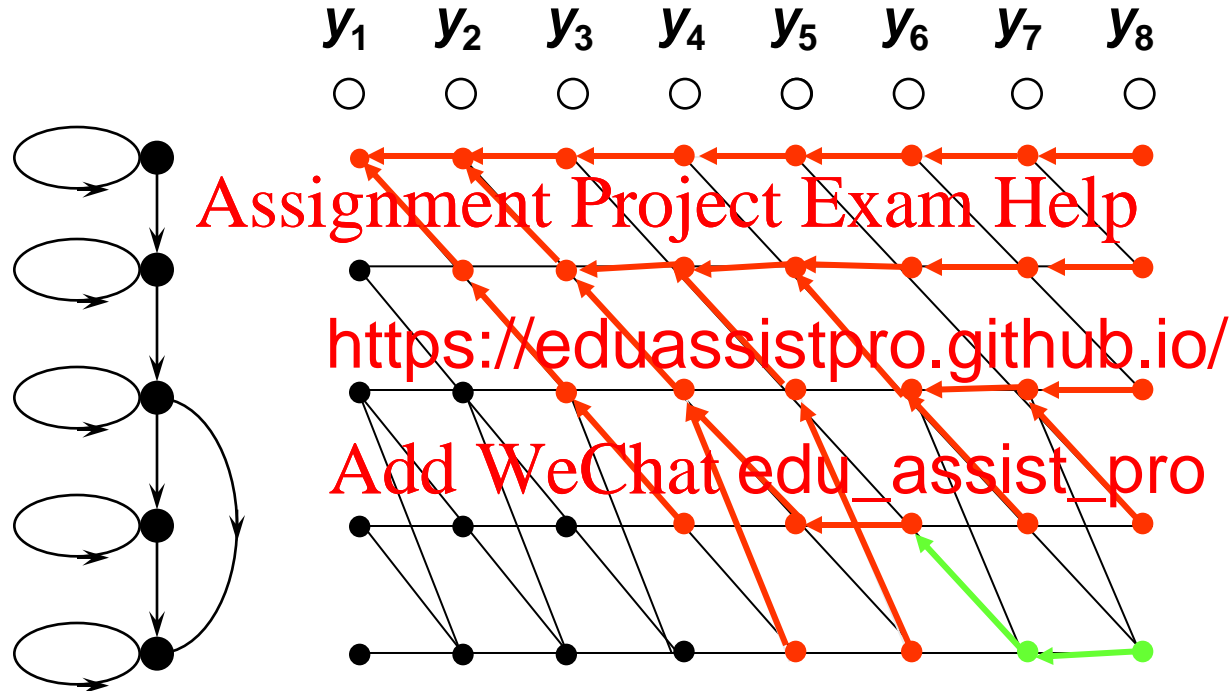
Viterbi Decoding – example



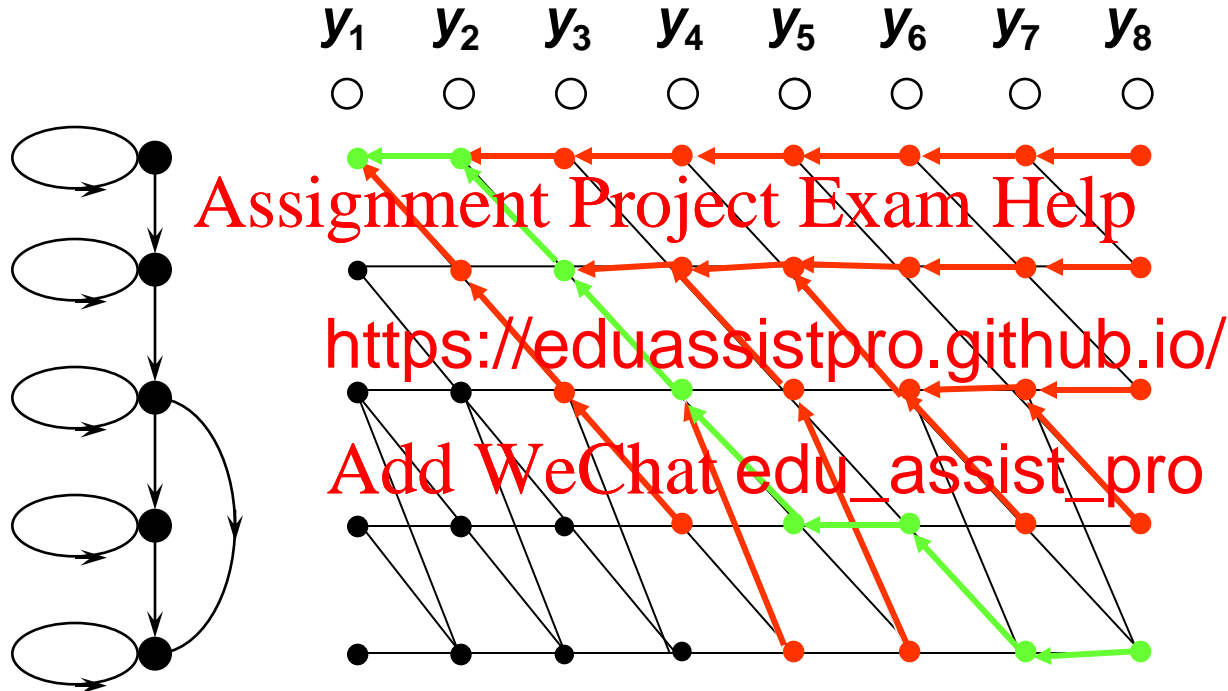
Viterbi Decoding – Trace-back



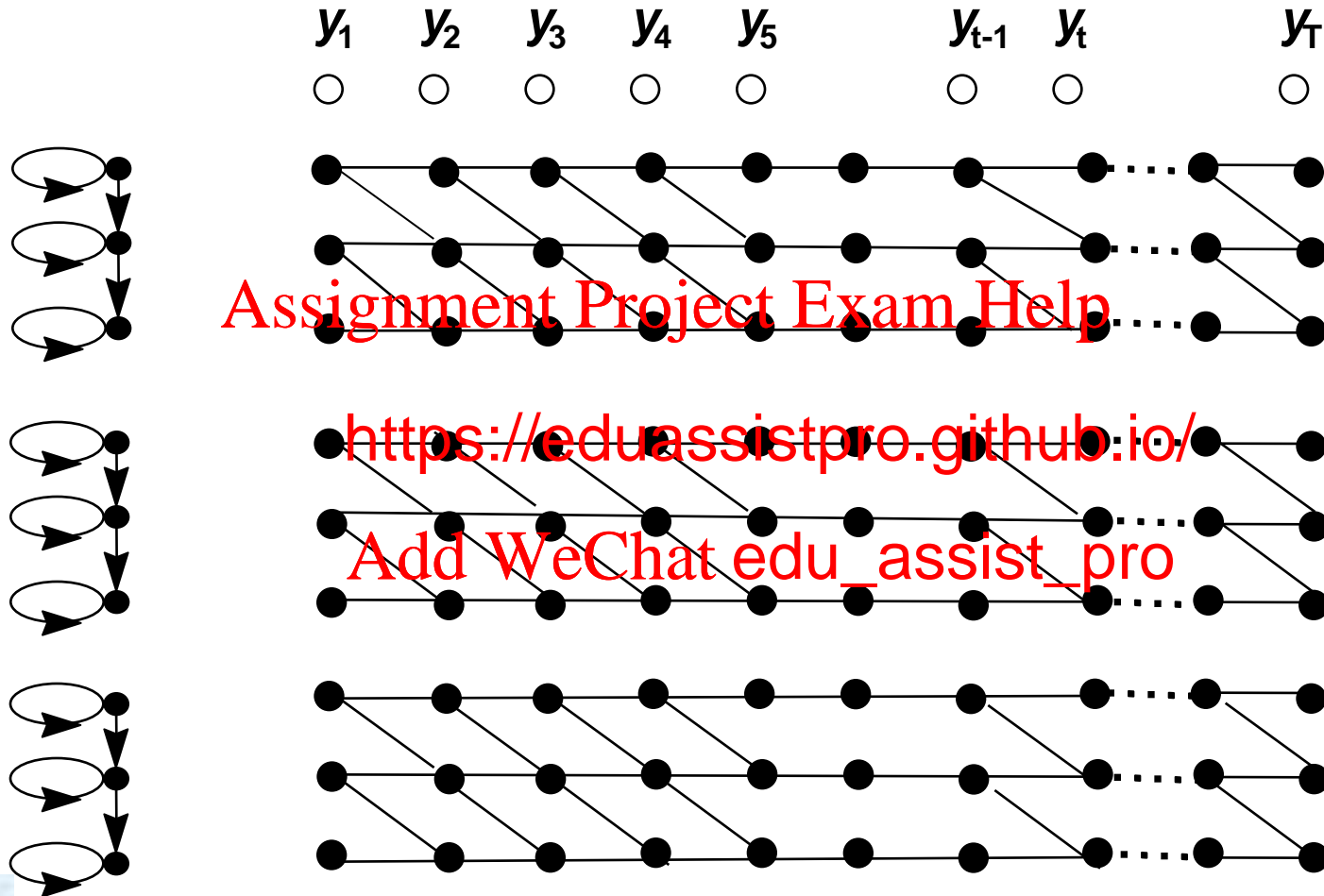
Viterbi Decoding – Trace-back



Viterbi Decoding – Trace-back

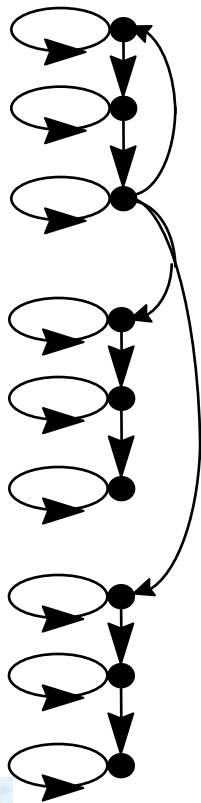


Isolated Speech Recognition



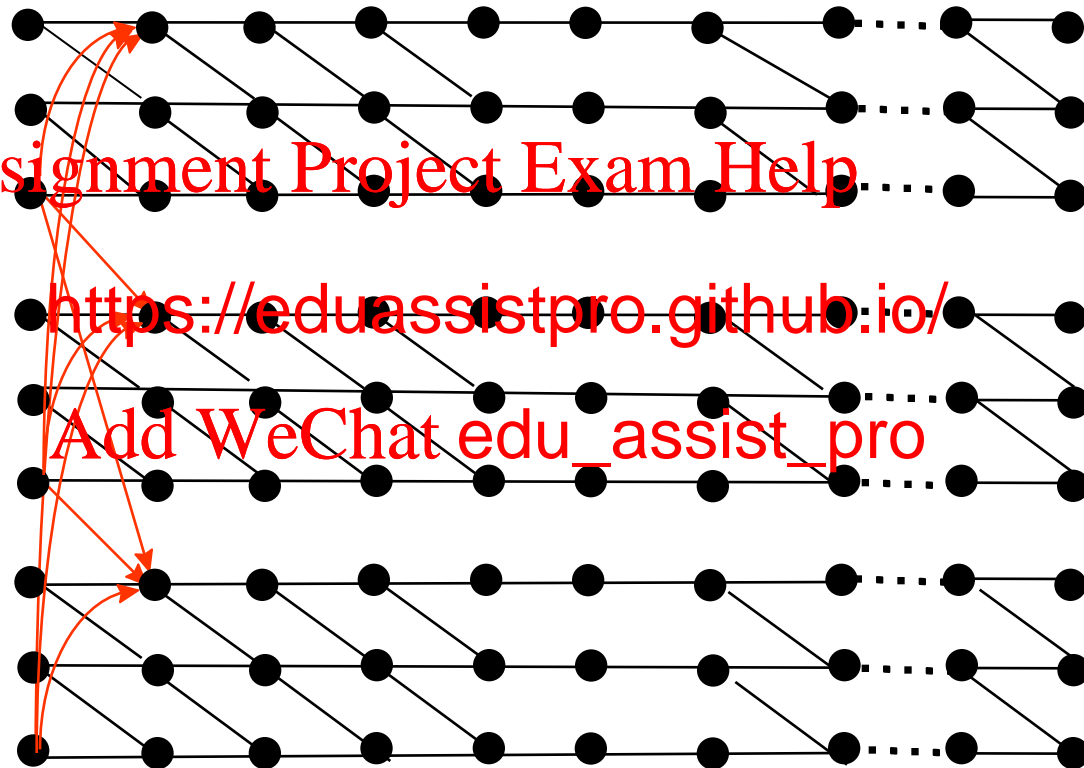
Connected Speech Recognition

New transitions connect
end of every model to
start of every model



y_1 y_2 y_3 y_4 y_5 y_{t-1} y_t y_T

○ ○ ○ ○ ○ ○ ○ ○ ○



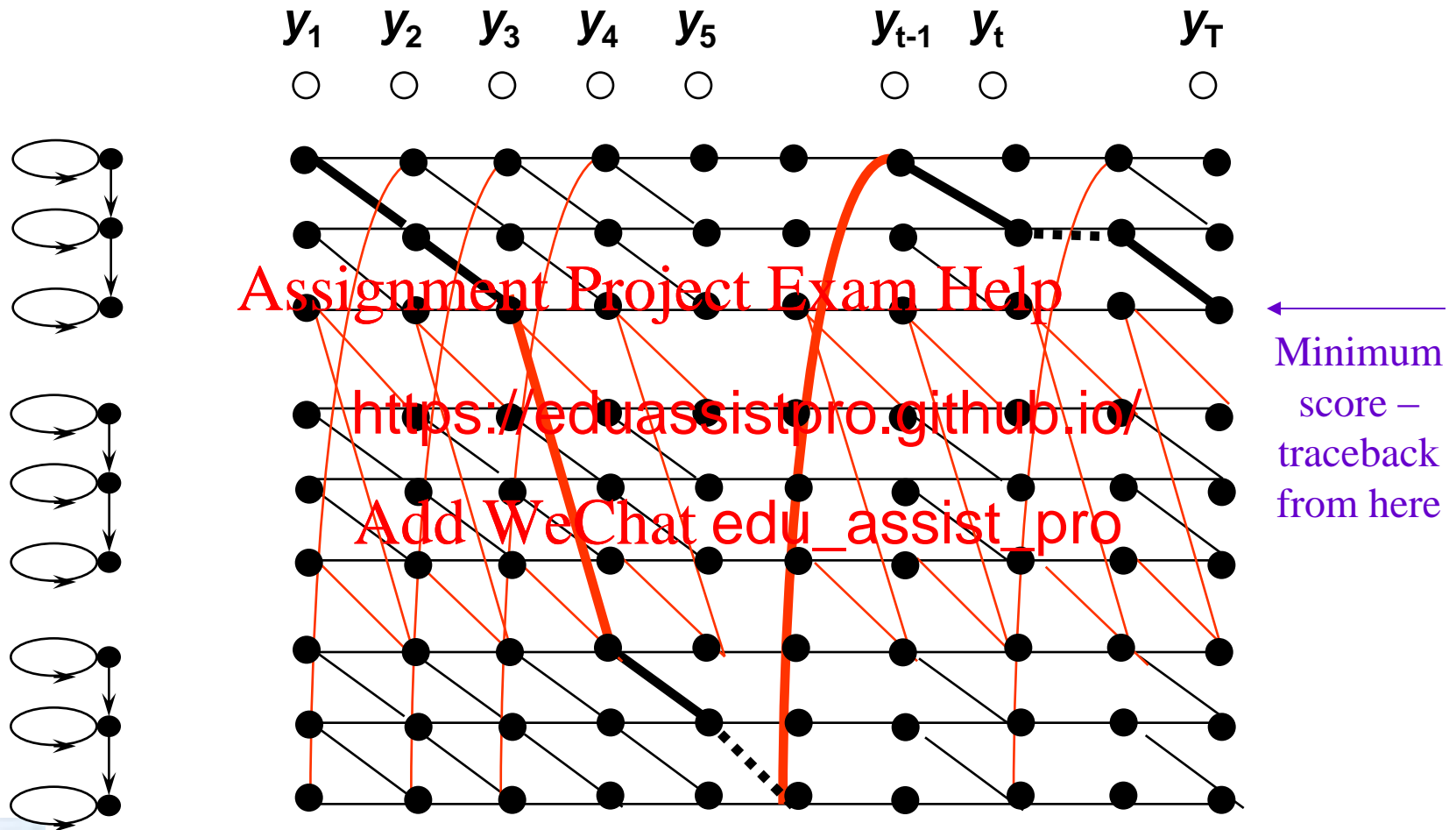
Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro



Connected Speech Recognition

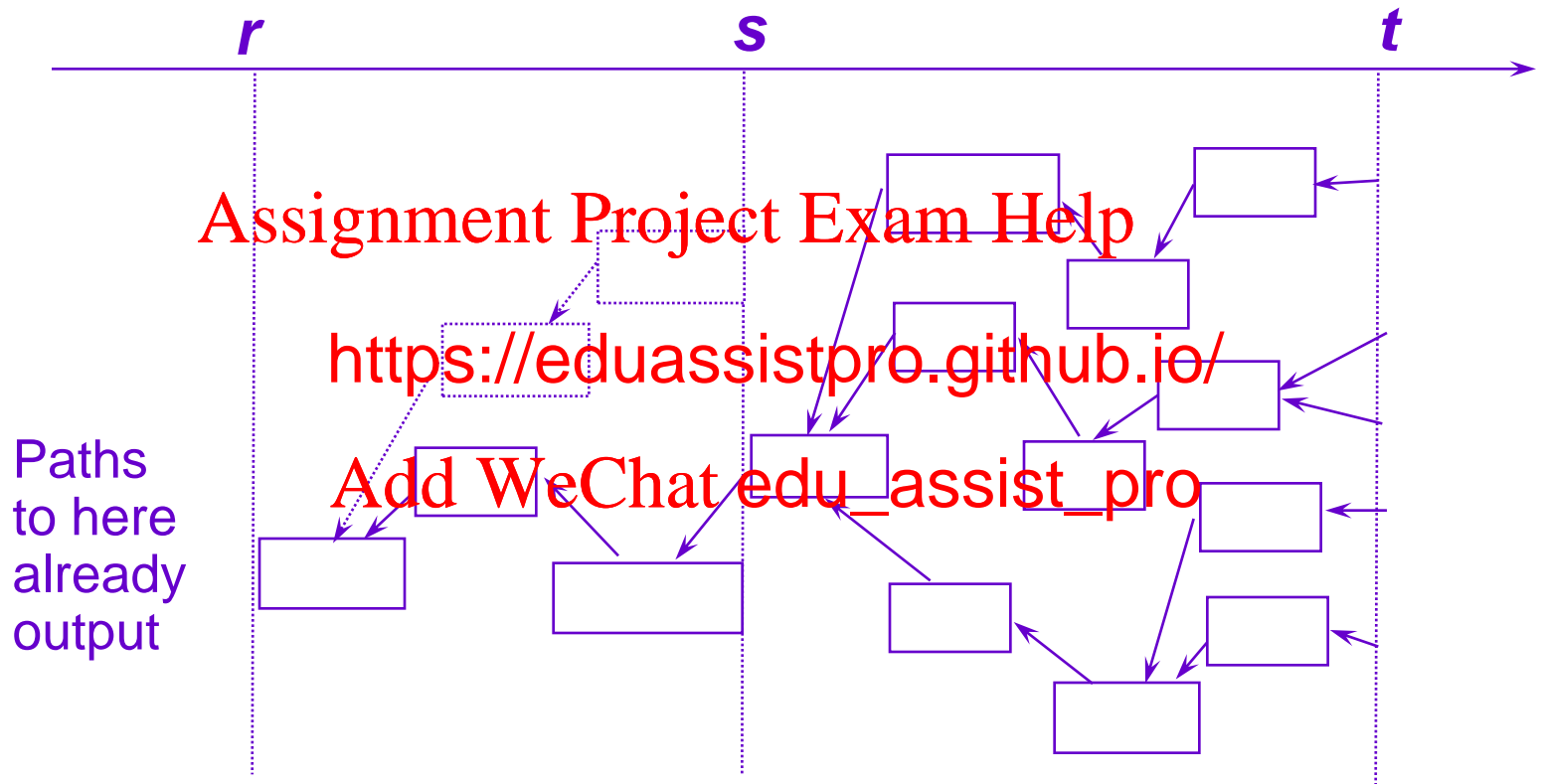


Partial Traceback (1)

- In continuous speech recognition, cannot trace-back from the end of the utterance (there is no end!)
- Instead **partial traceback** operates as follows:
 - For each time t and state i a *word link record* describes the best path to (t, i) .
 - At regular intervals all active paths are traced back to see if they converge at some time s in the past
 - If so, the best path up to time s cannot change, and the sequence of words up to s can be output

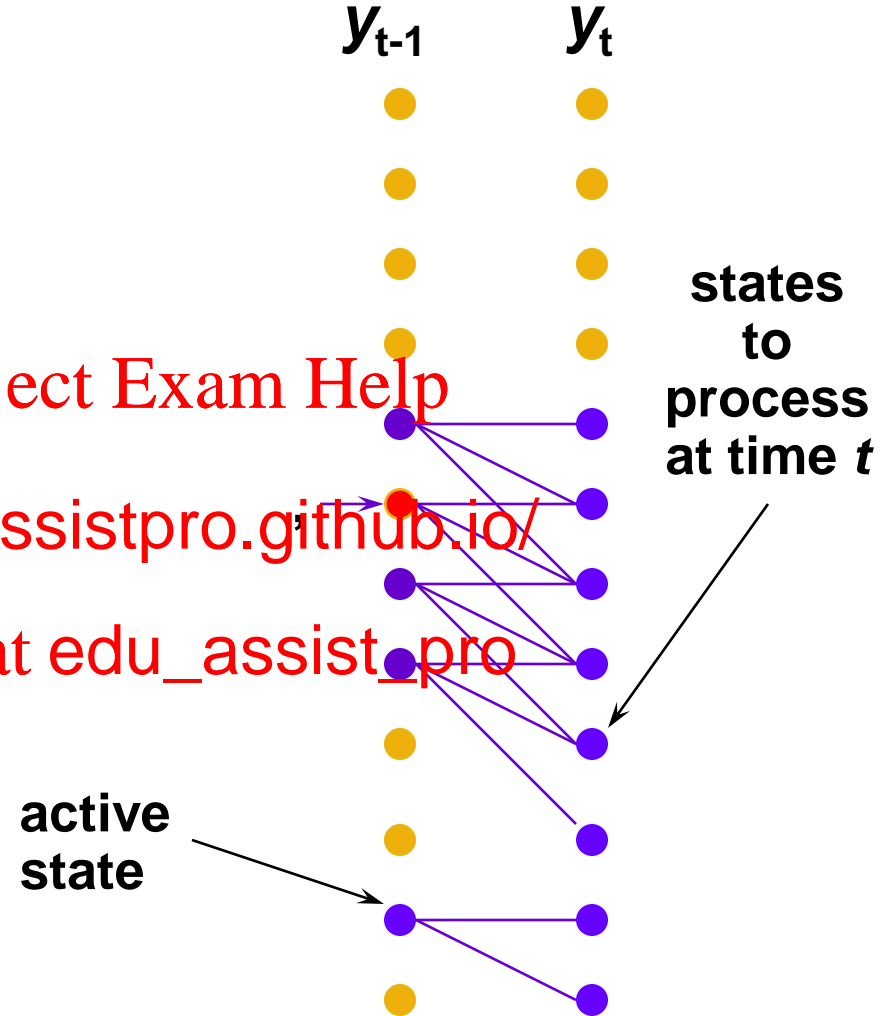


Partial Traceback (2)



Beam Pruning

- Choose threshold T
- The active states at time $t-1$ are those whose partial path scores are the best path scores at time $t-1$
- At time t , only process those states which link back to an active state



Partial Traceback, Beam Pruning & Recognition ‘Speed’

- Partial traceback introduces a ‘lag’ into recognition process - **not** error speed
- Lag worse w <https://eduassistpro.github.io/>
- Beam Pruning **less effective** w [edu_assist_pro](https://eduassistpro.github.io/) input
- Severe Beam Pruning will degrade performance
- Proper management of Partial Traceback and Beam Pruning is essential for optimal performance



Summary

- Speech recognition using HMMs
 - Viterbi decoding
 - Isolated & Connected/Continuous speech recognition
 - Beam pruning
 - Partial tracing

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro

