Sequences and Time-Series

July 24, 2020

Applied Data Science
MMCi Term 4

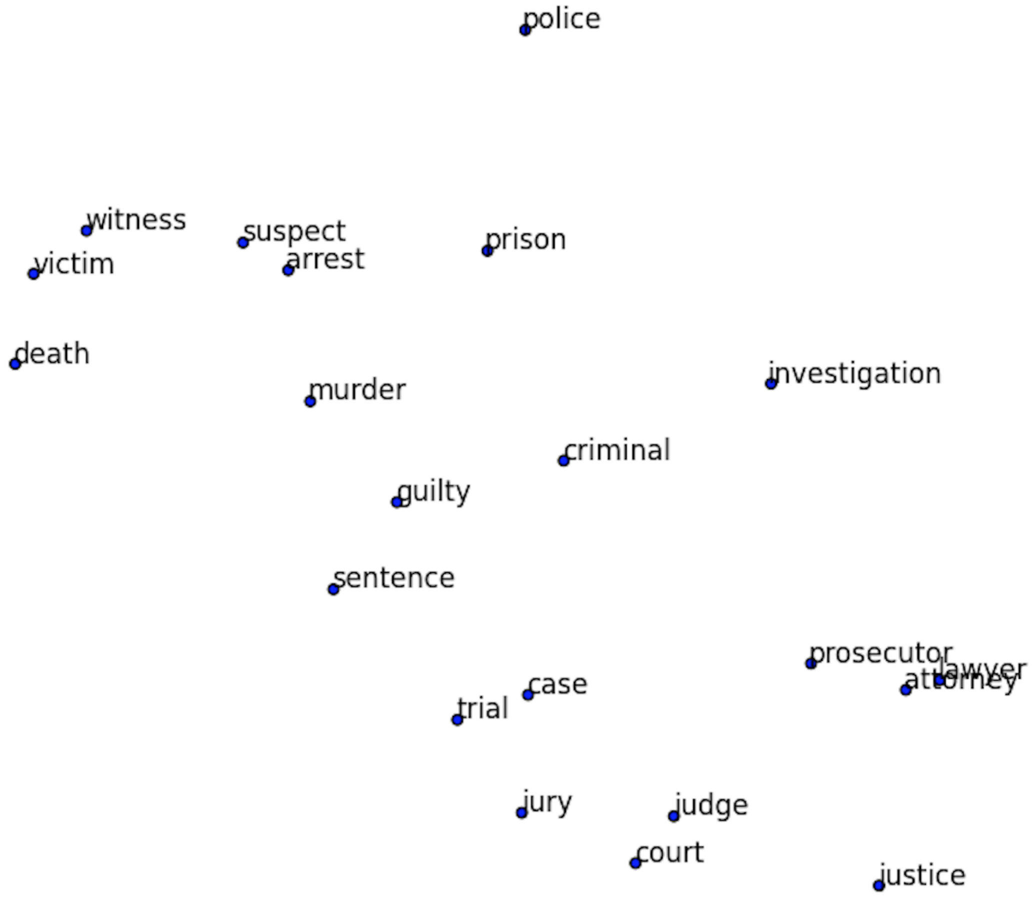
Matthew Engelhard

Recall: Word embeddings allow us to quantify word meaning

If we zoom in on a small region of our word map, it's all related words.

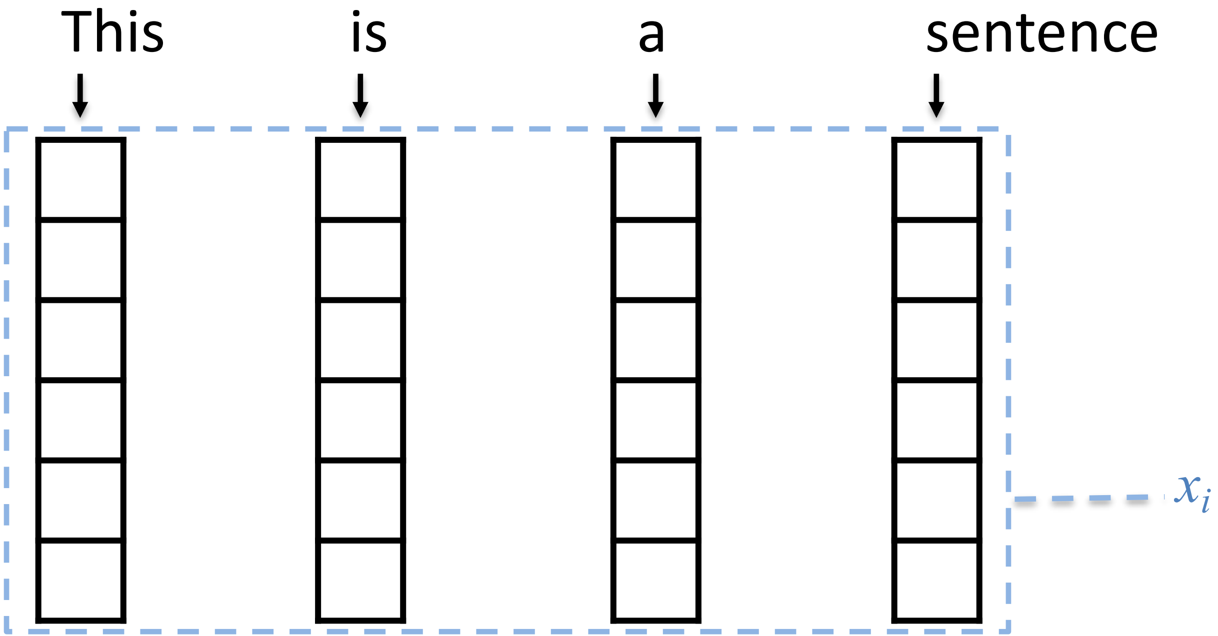
Note the similarity of all the words as a whole, but also of the individual neighbors.

"Lawyer" and "attorney" are nearly identical in space!

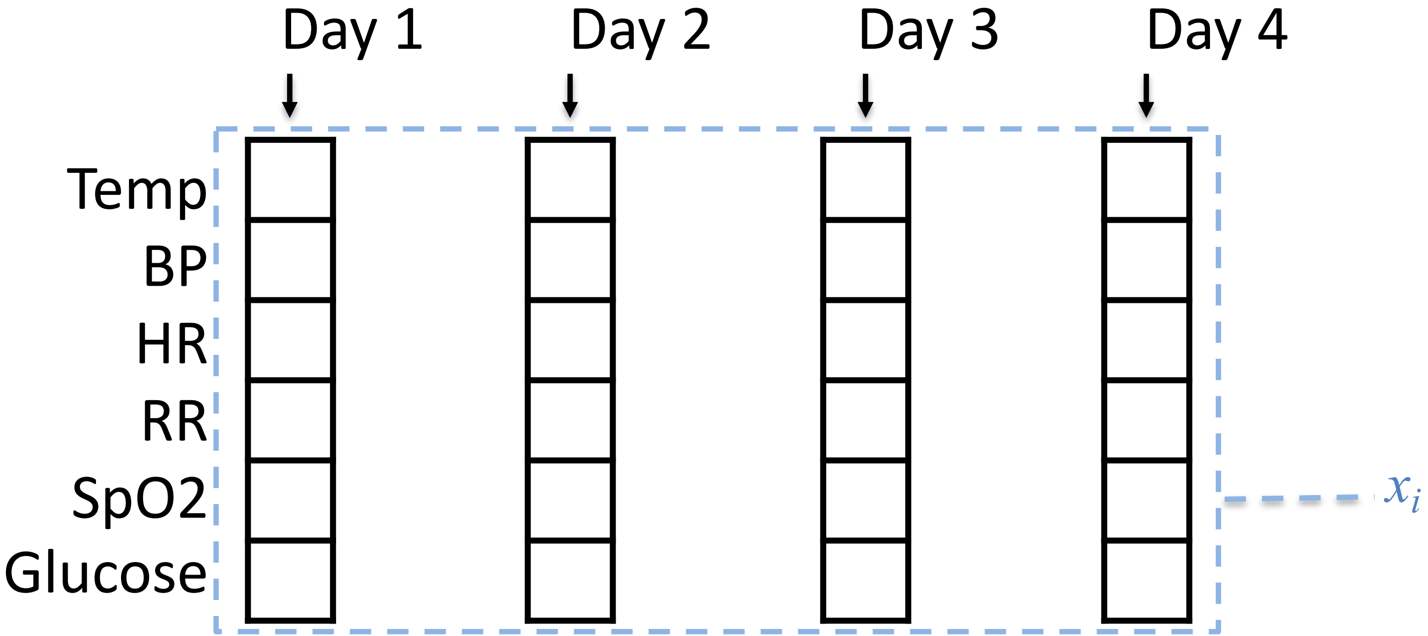


Applying Word Embeddings to a Sentence

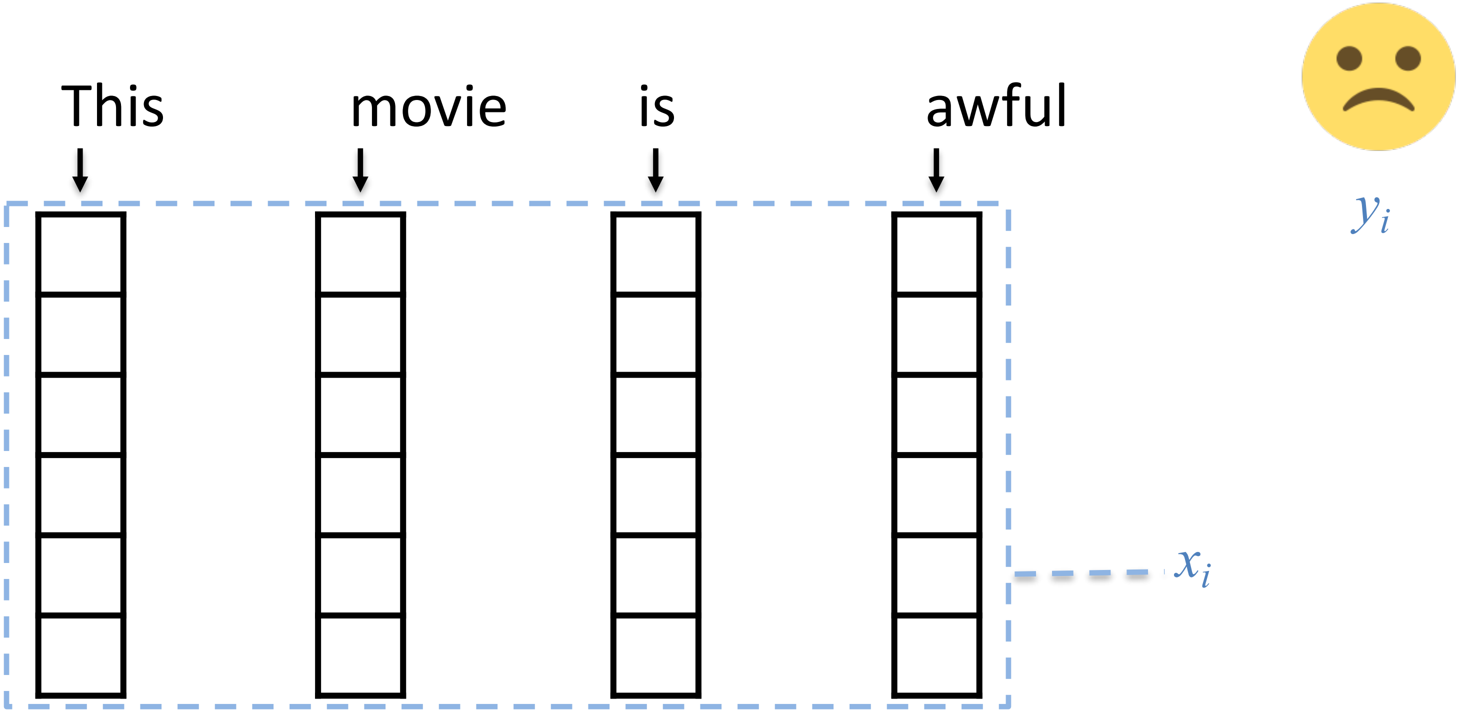
- Look up words individually to obtain their vectors
- Construct a sequence of vectors



Sequences of measurements: *same structure*

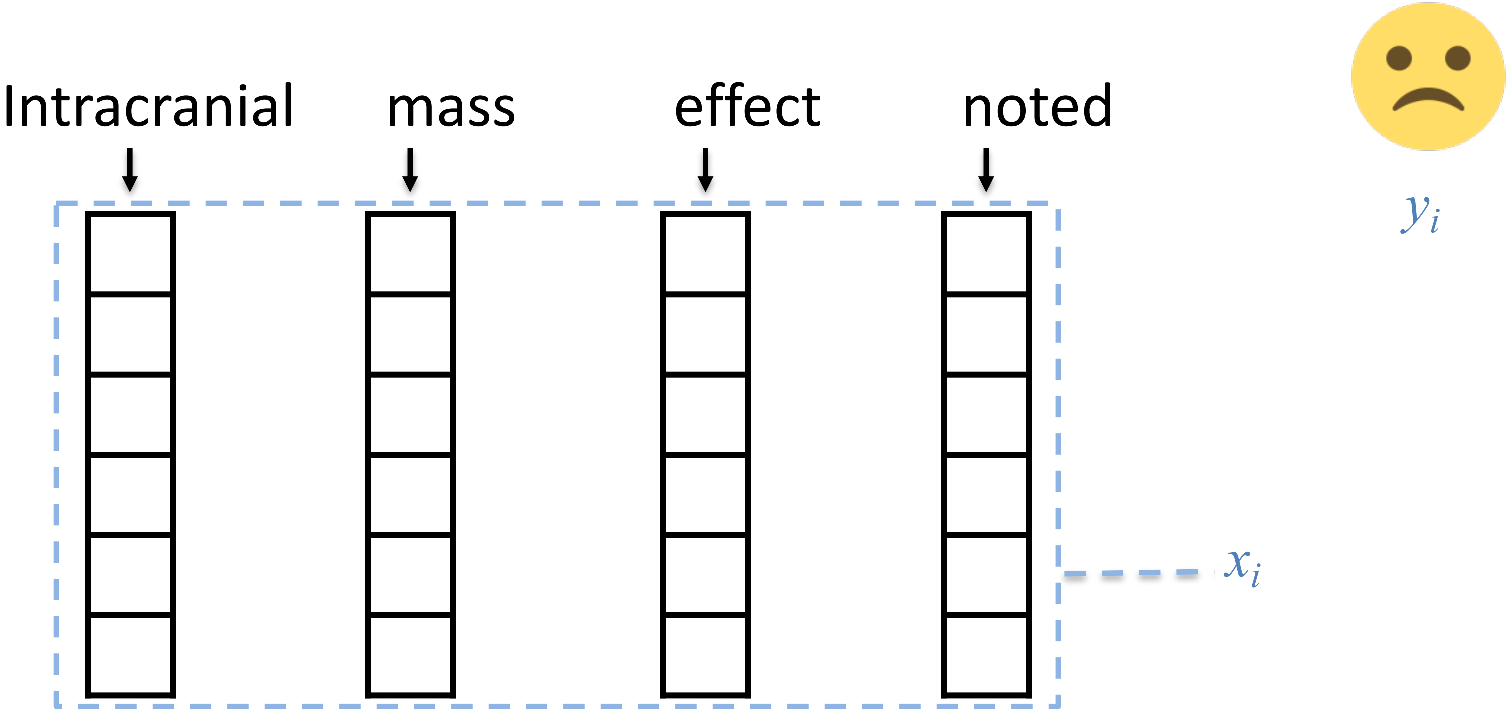


Task 1: Predict a label associated with the sentence



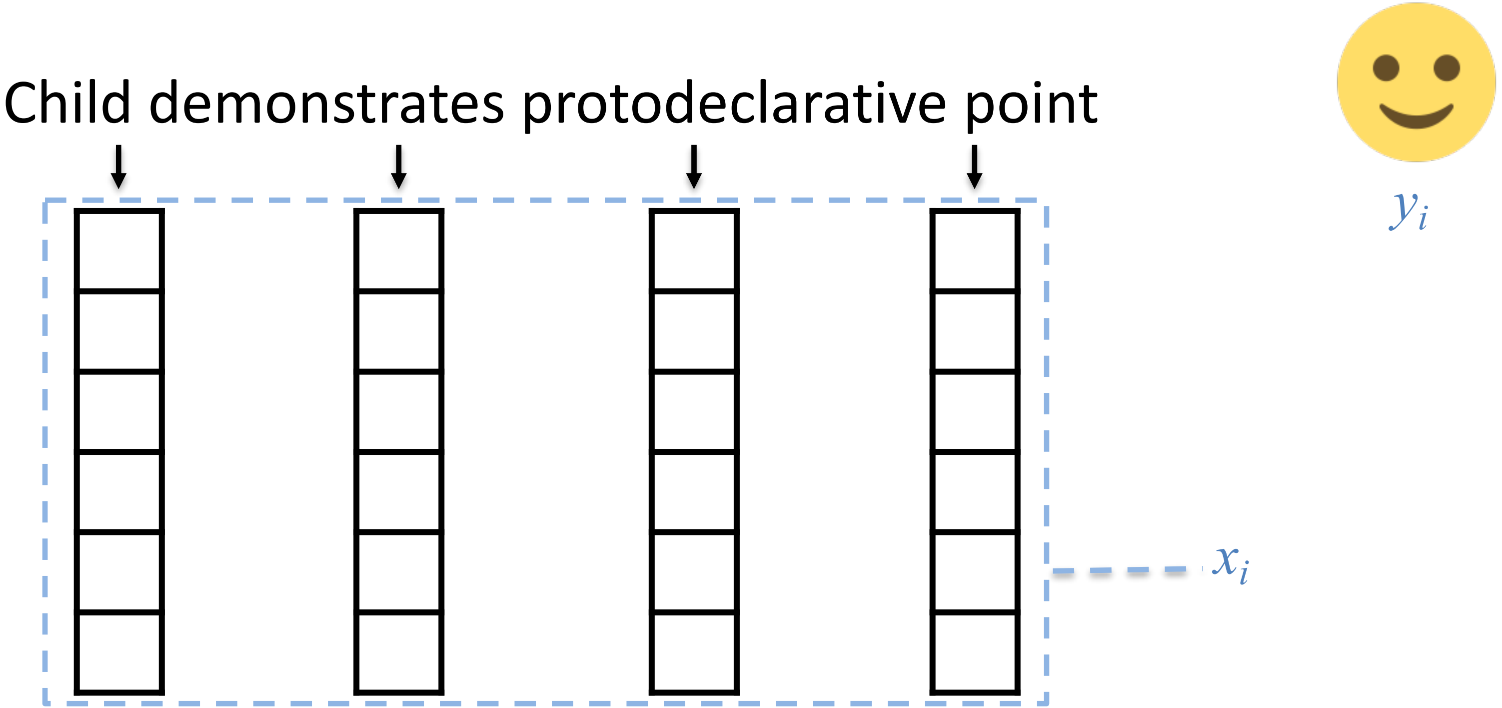
Duke UNIVERSITY

Task 1: Predict a label associated with the report

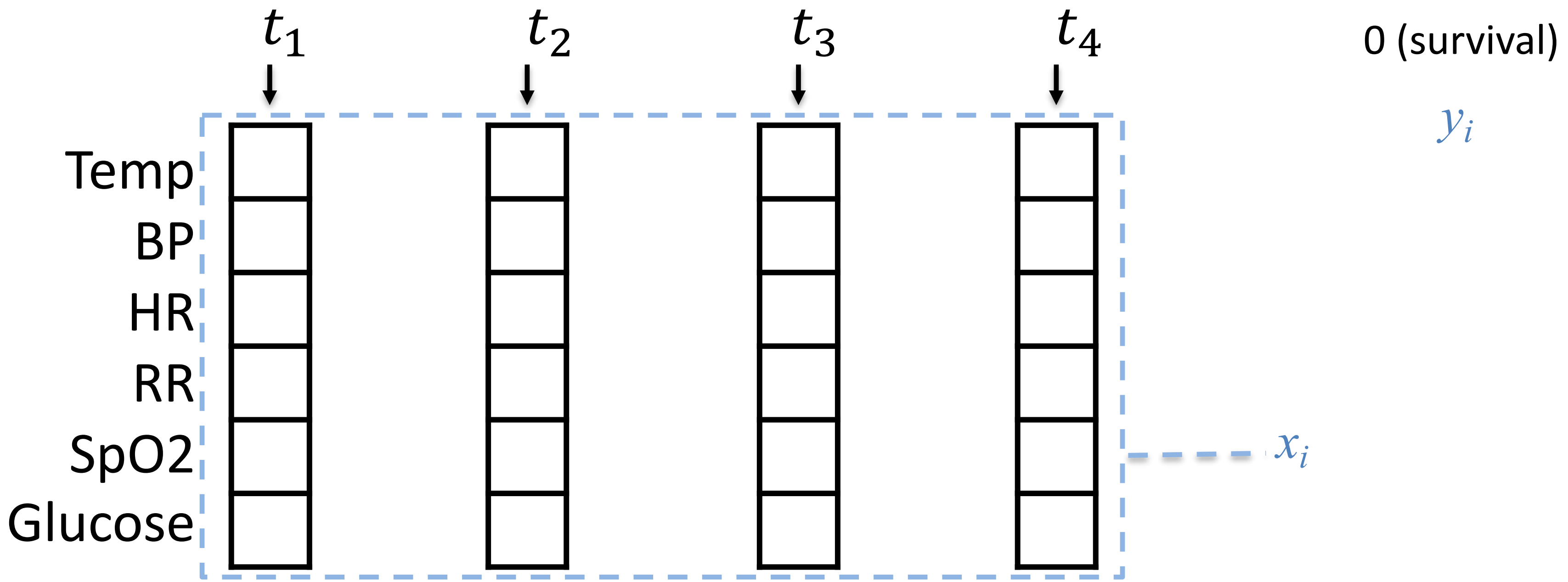


Classification of radiology reports using neural attention models, *IJCNN 2017*

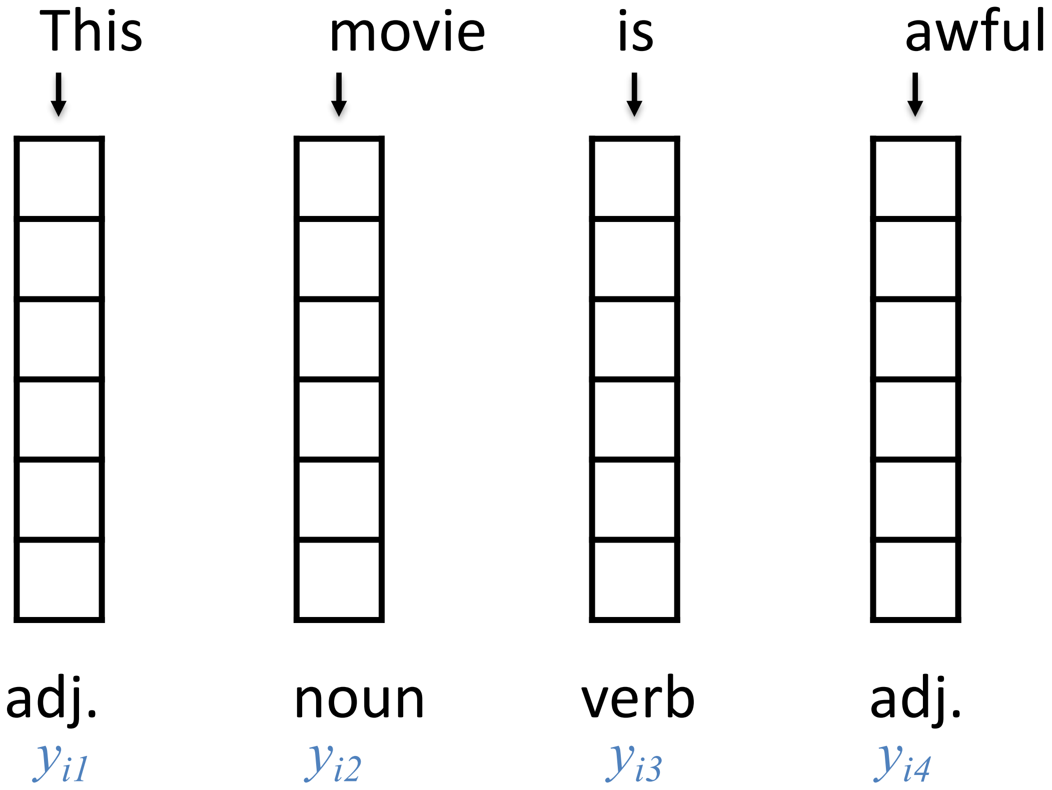
Task 1: Predict a label associated with the note



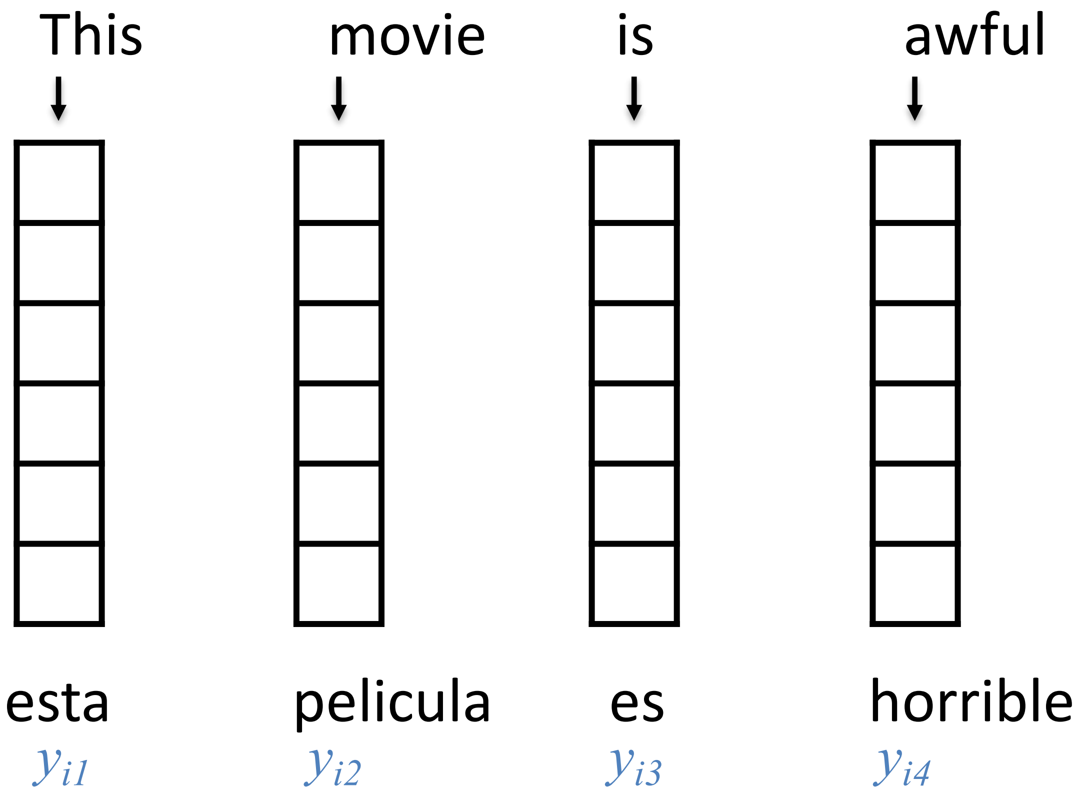
Task 1: Predict label assoc. with all measurements



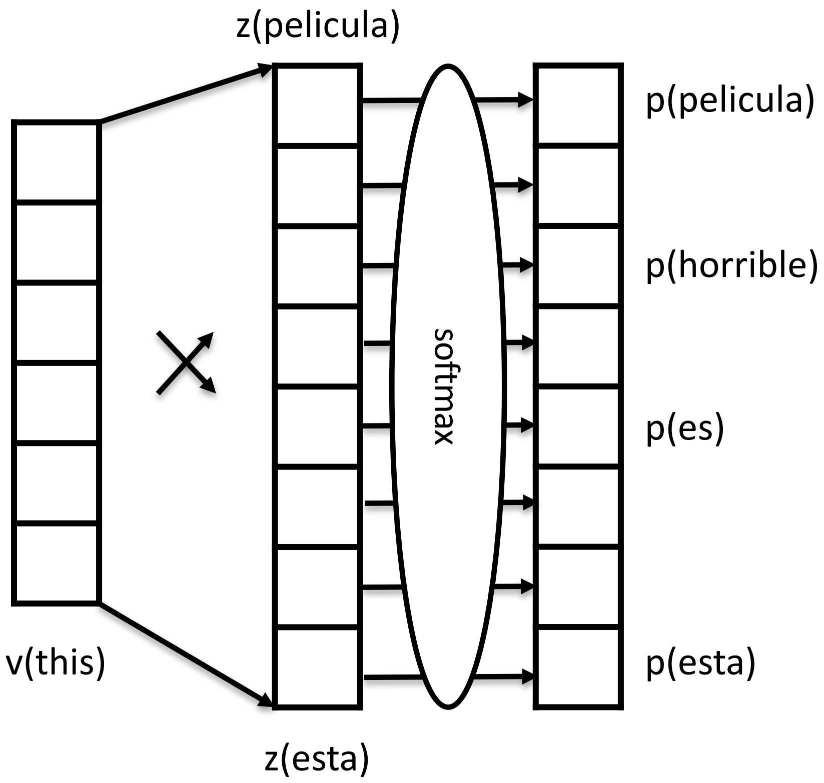
Task 2: Predict a label associated with each word



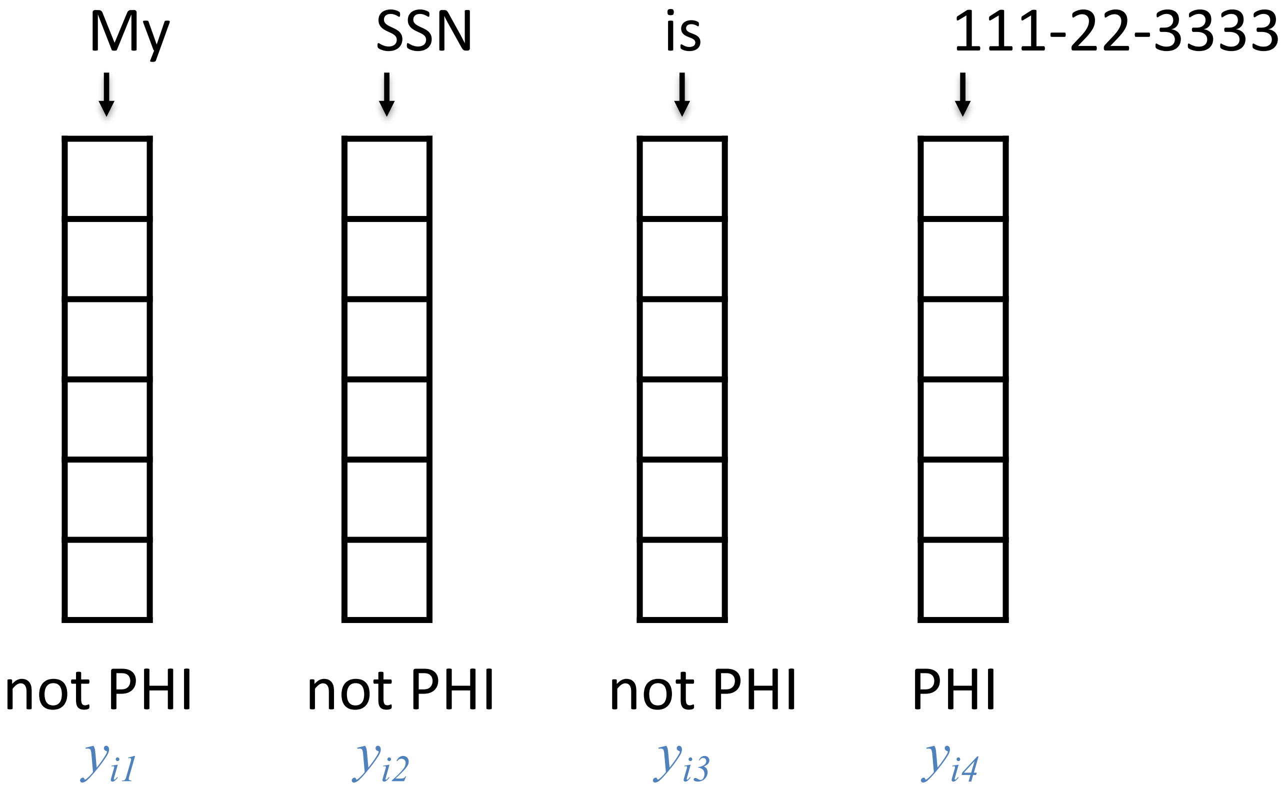
Task 2: Predict a **label** (?) associated with each word



Multi-Class Logistic Regression (many classes)



Task 2: Predict a label associated with each word



Deidentification of Patient Notes

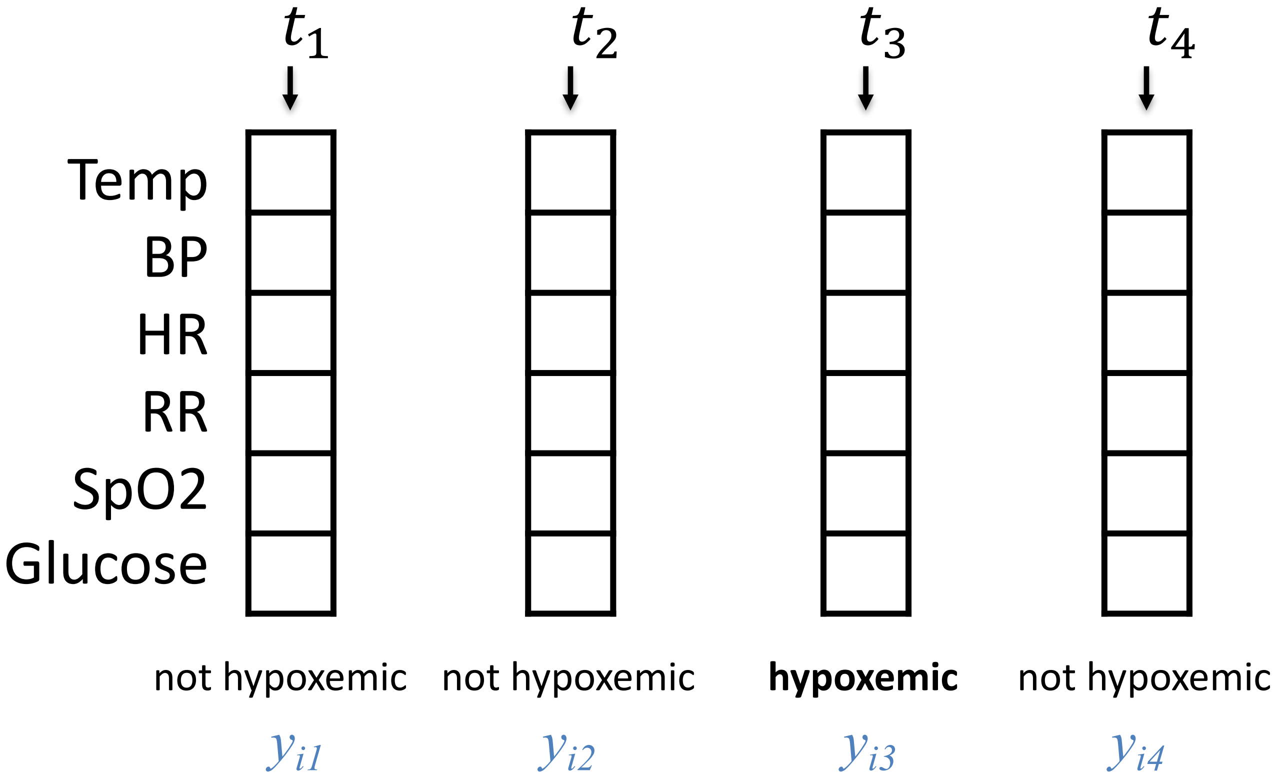
Table 5. Examples of correctly detected PHI instances (in bold) by the ANN

PHI category	ANN
AGE	Father had a stroke at **80** and died of?another stroke at age Personal data and overall health: Now **63**, despite his FH: Father: Died @ **52** from EtOH abuse (unclear exact etiology) Tobacco: smoked from age 7 to **15**, has not smoked since 15.
CONTACT	History of Present Illness **86F** reports worsening b/l leg pain. by phone, Dr. Ivan Guy. Call w/ questions **86383**. Keith Gilbert, H/O paroxysmal afib VNA **171-311-7974** ======= Medications
DATE	During his **May** hospitalization he had dysphagia Social history: divorced, quit smoking in **08**, sober x 10 yrs, She is to see him on the **29th** of this month at 1:00 p.m. He did have a renal biopsy in teh late **60s** adn thus will look for results, Results**02/20/2087** NA 135, K 3.2 (L), CL 96 (L), CO2 30.6, BUN 1 Jose Church, M.D. /ray DD: 01/18/20 DT: **01/19/:0** DV: 01/18/20

De-identification of patient notes with recurrent neural networks
Dernoncourt F, Lee JY, Uzuner O, Szolovits P
JAMIA 24(3), 2017, 596–606

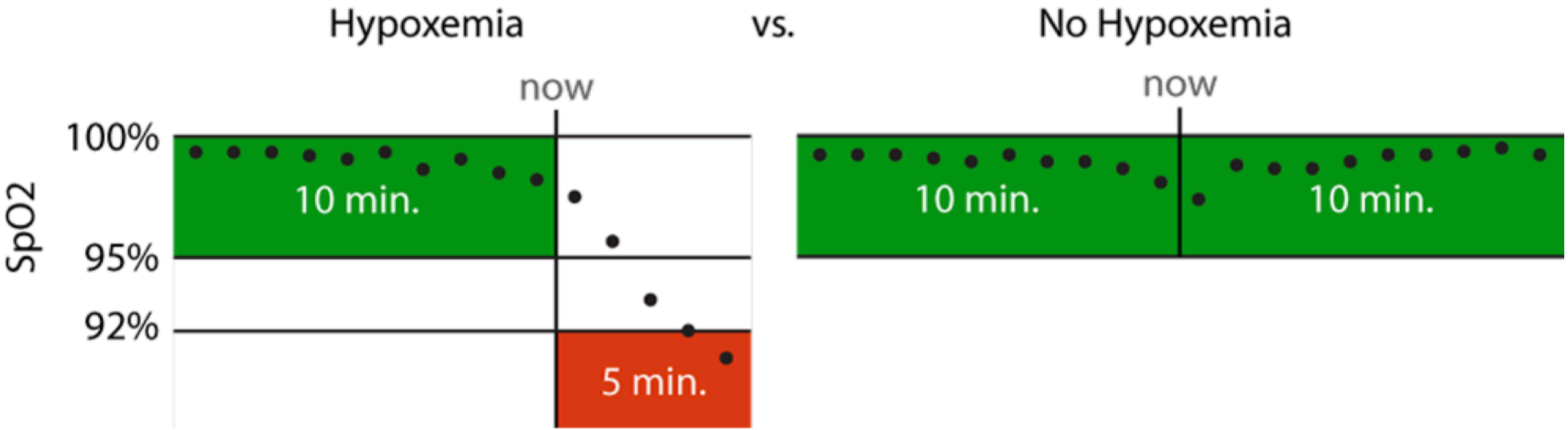
- A bidirectional RNN is used to identify PHI (18 HIPAA fields)
- *i2b2*: 889 discharge summaries, >28k PHI tokens
- *MIMIC*: 1635 discharge summaries, >60k PHI tokens
- State of the art sensitivity and F1 metric on both datasets

Task 1: Predict label assoc. with each time point



Hypoxemia Prediction during Surgery

Real-time Prediction Task:
- hypoxemia (yes/no) in the next 5 minutes
- based on data from the Anesthesia Information Management System
- static features + real-time features collected up to that time point

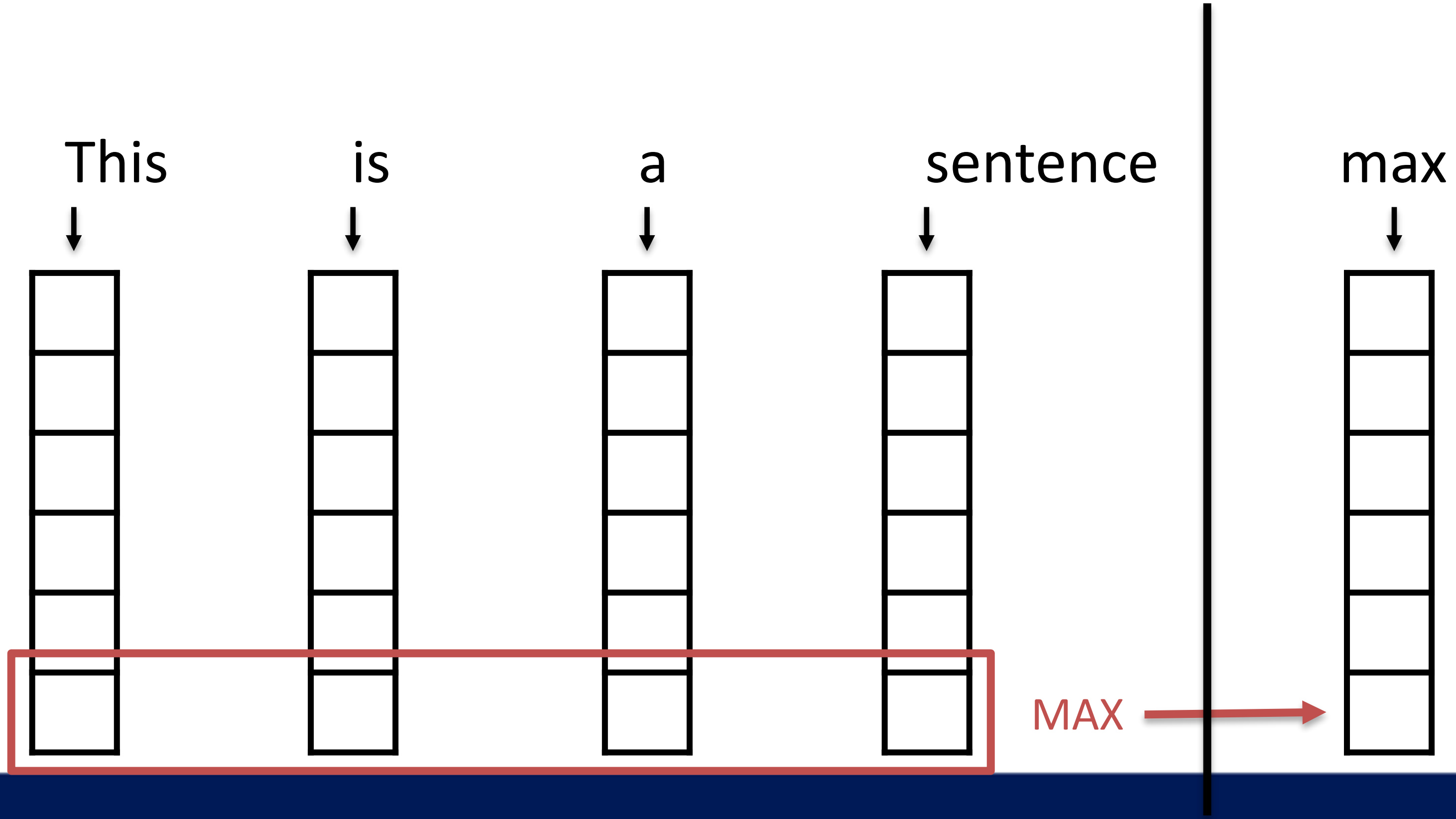


Problem 1: Sequences Vary in Length

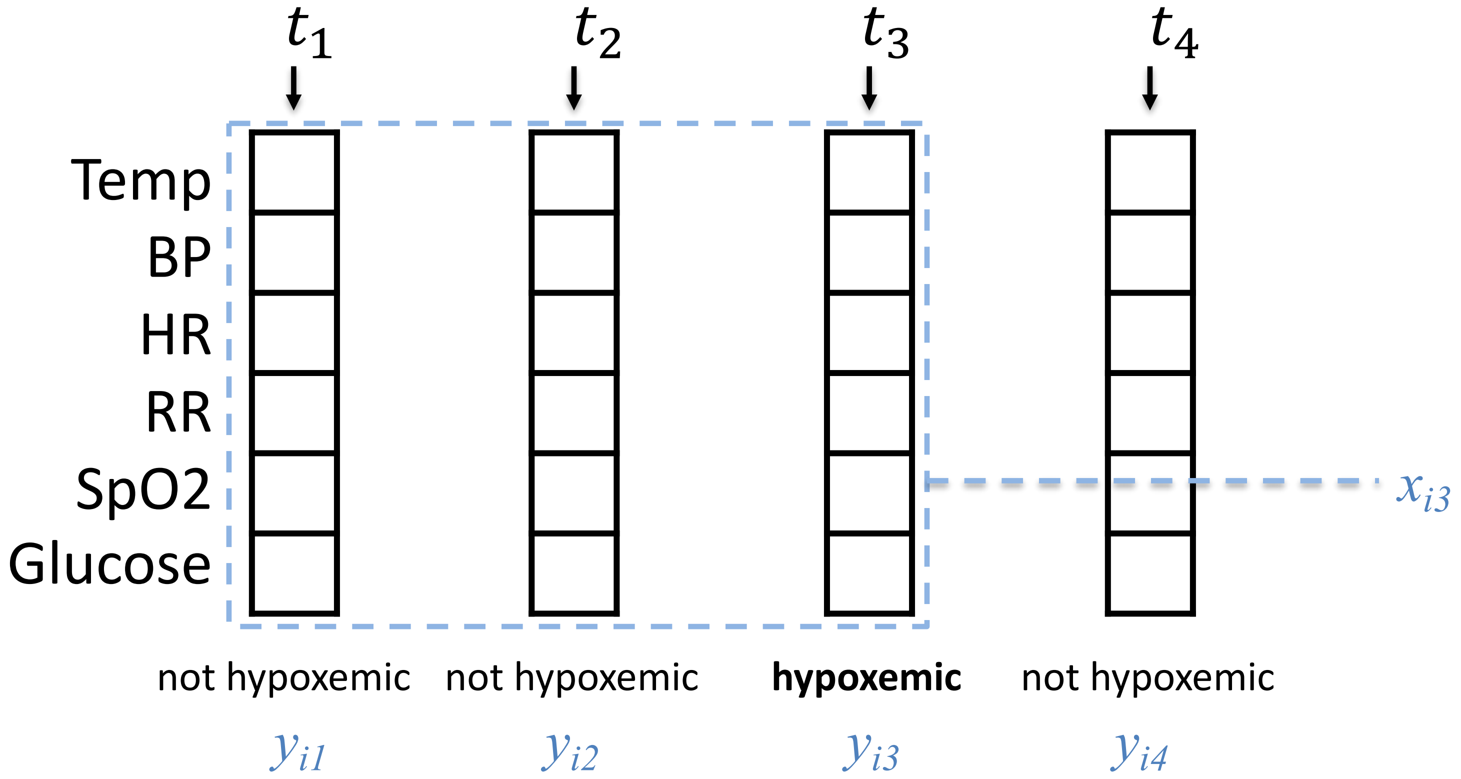
- Sentences/text have different # words
- Time-series have different # measurement times

- <u>Solution 1:</u> aggregate over words/time points

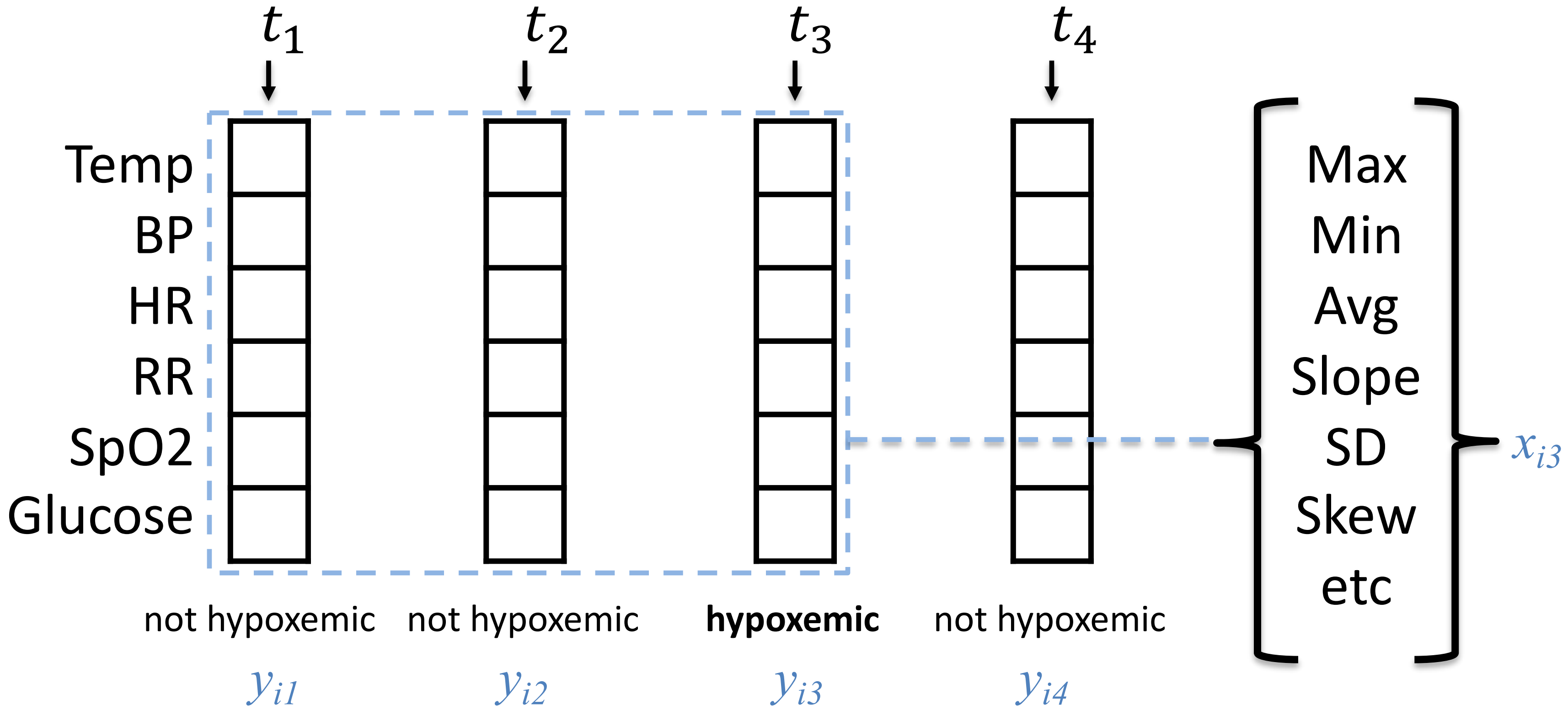
VSWEM allows us to convert a variable-length sentence to a fixed-length feature vector



Similarly, we can aggregate measurements in a time-series



Similarly, we can aggregate measurements in a time-series

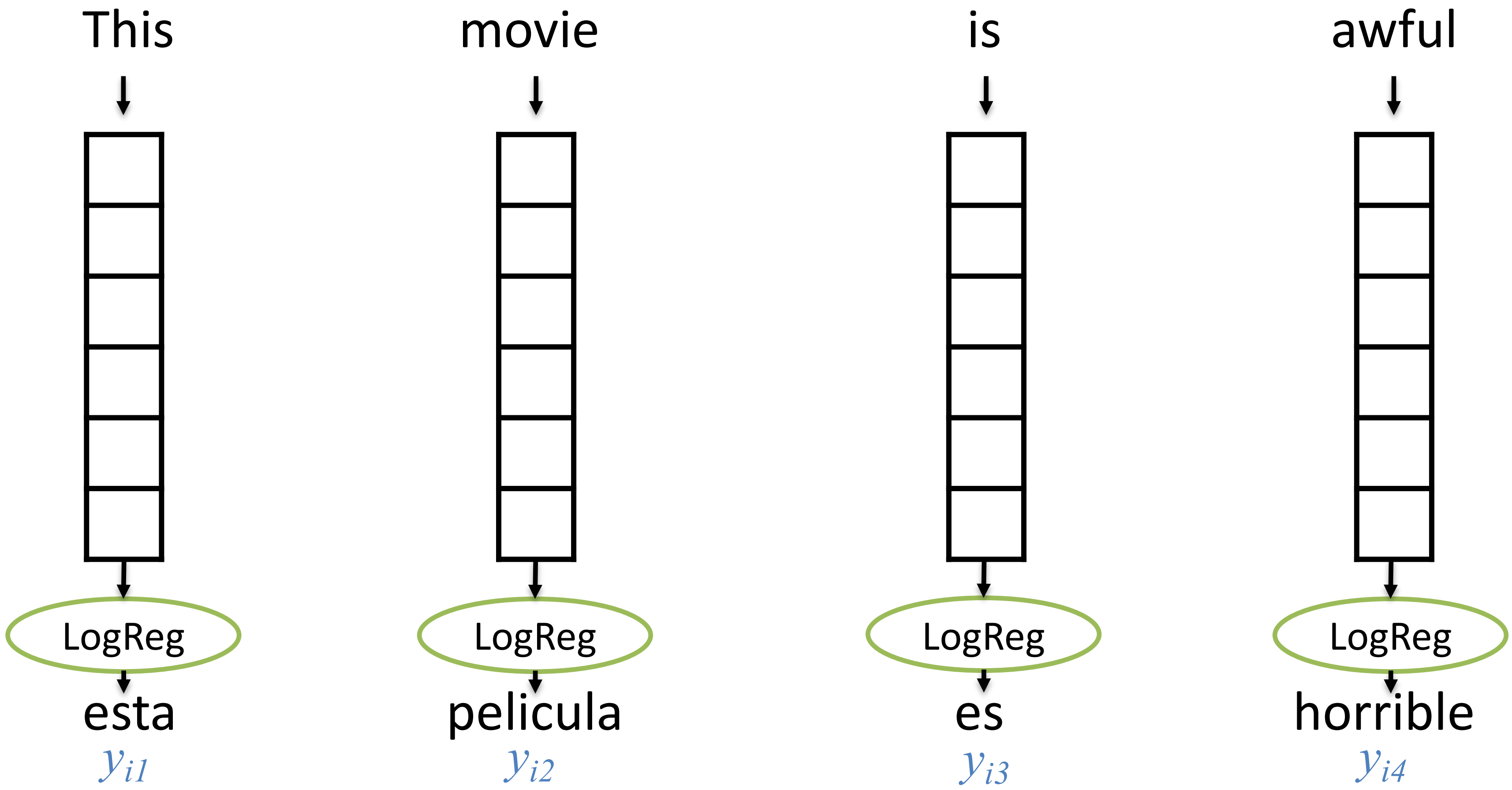


Problem 2: Interpret Words or Measurements *in Context*

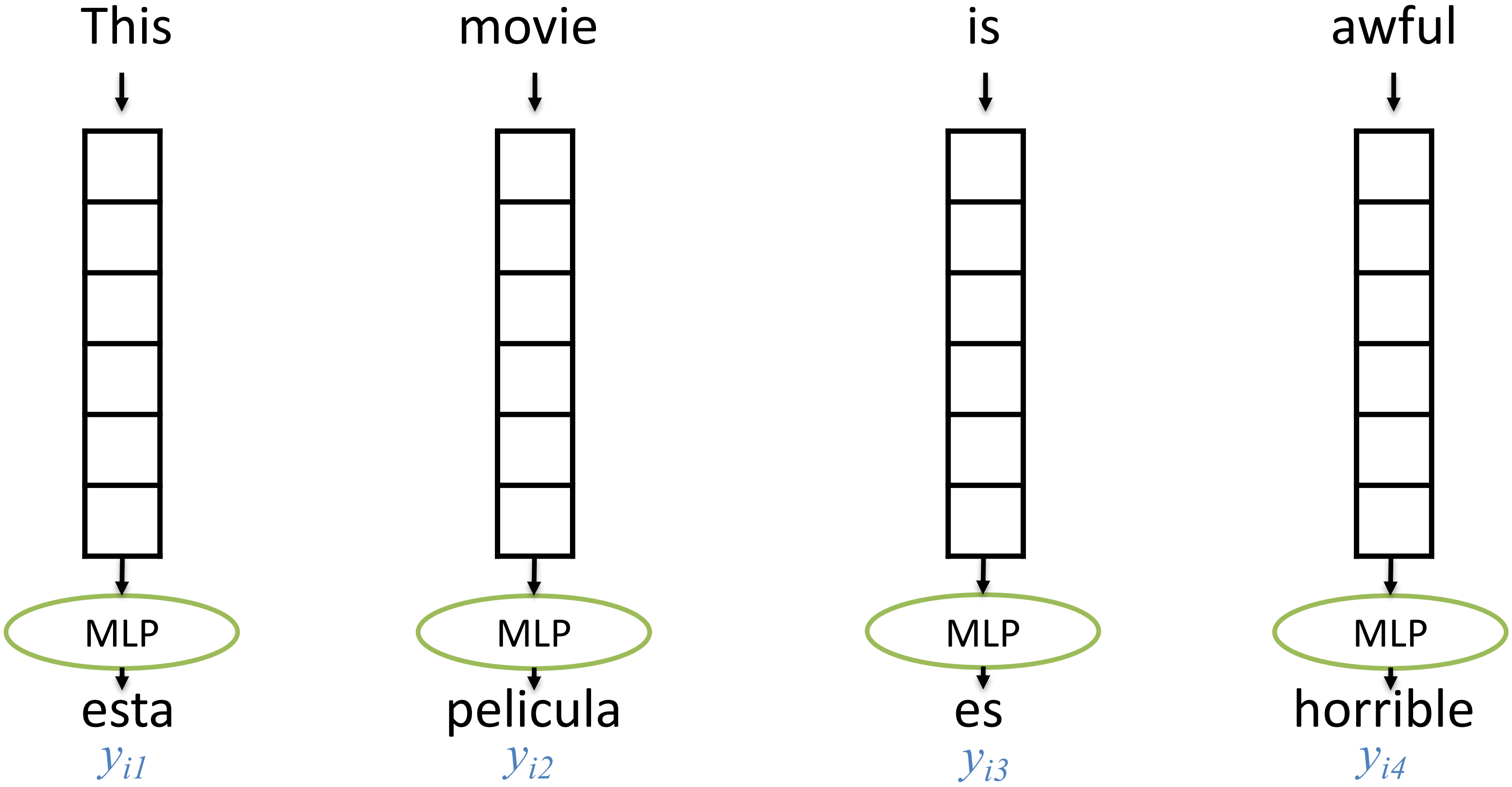
- A sentence is more than the average (or max) of its words
- A time-series is more than the average / min / max / SD of individual measurements
- Deep learning: we *learn* what's important about the sequence rather than choosing features or summary stats

RECURRENT NEURAL NETWORKS

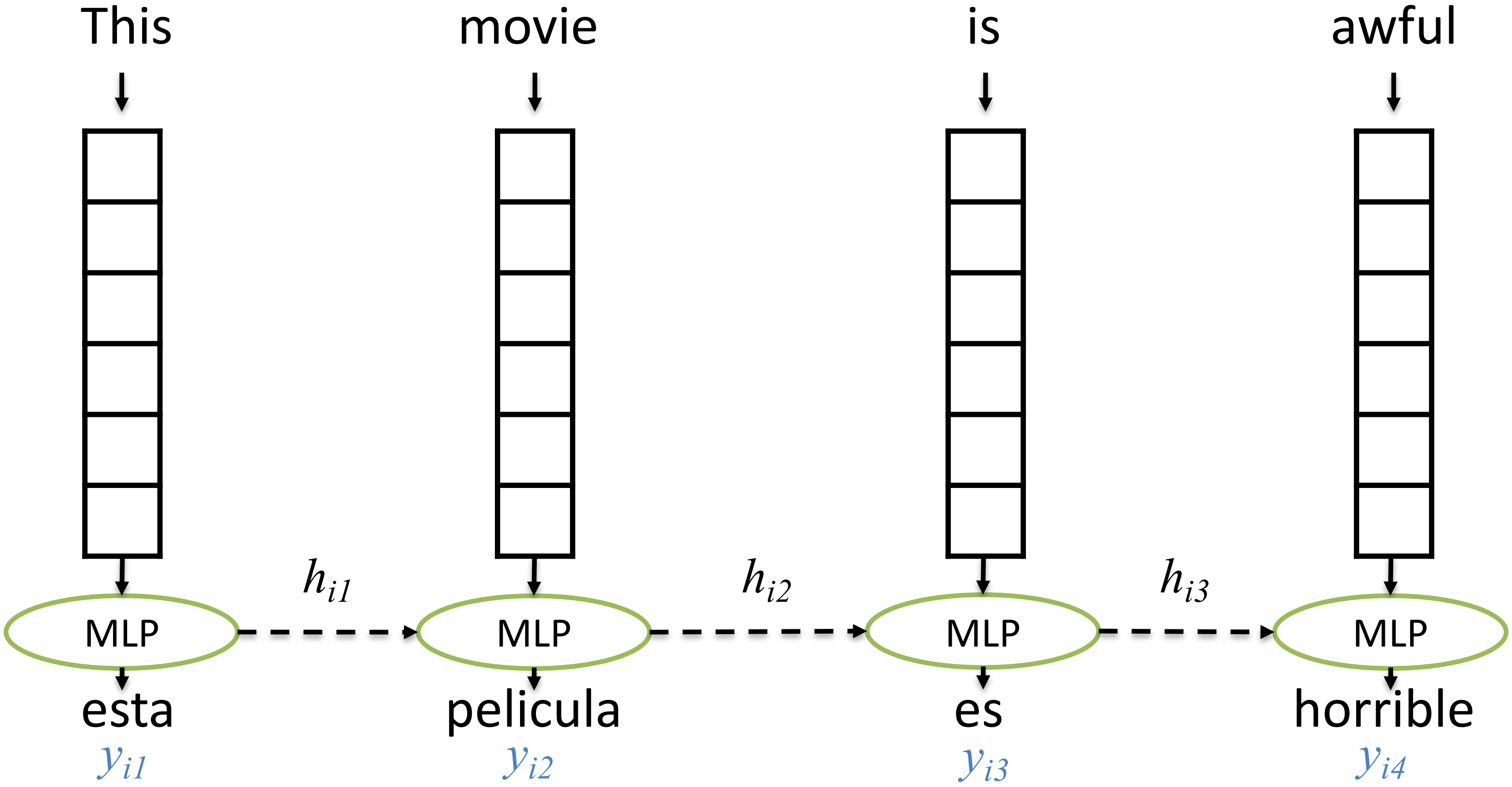
Predict a label associated with each word



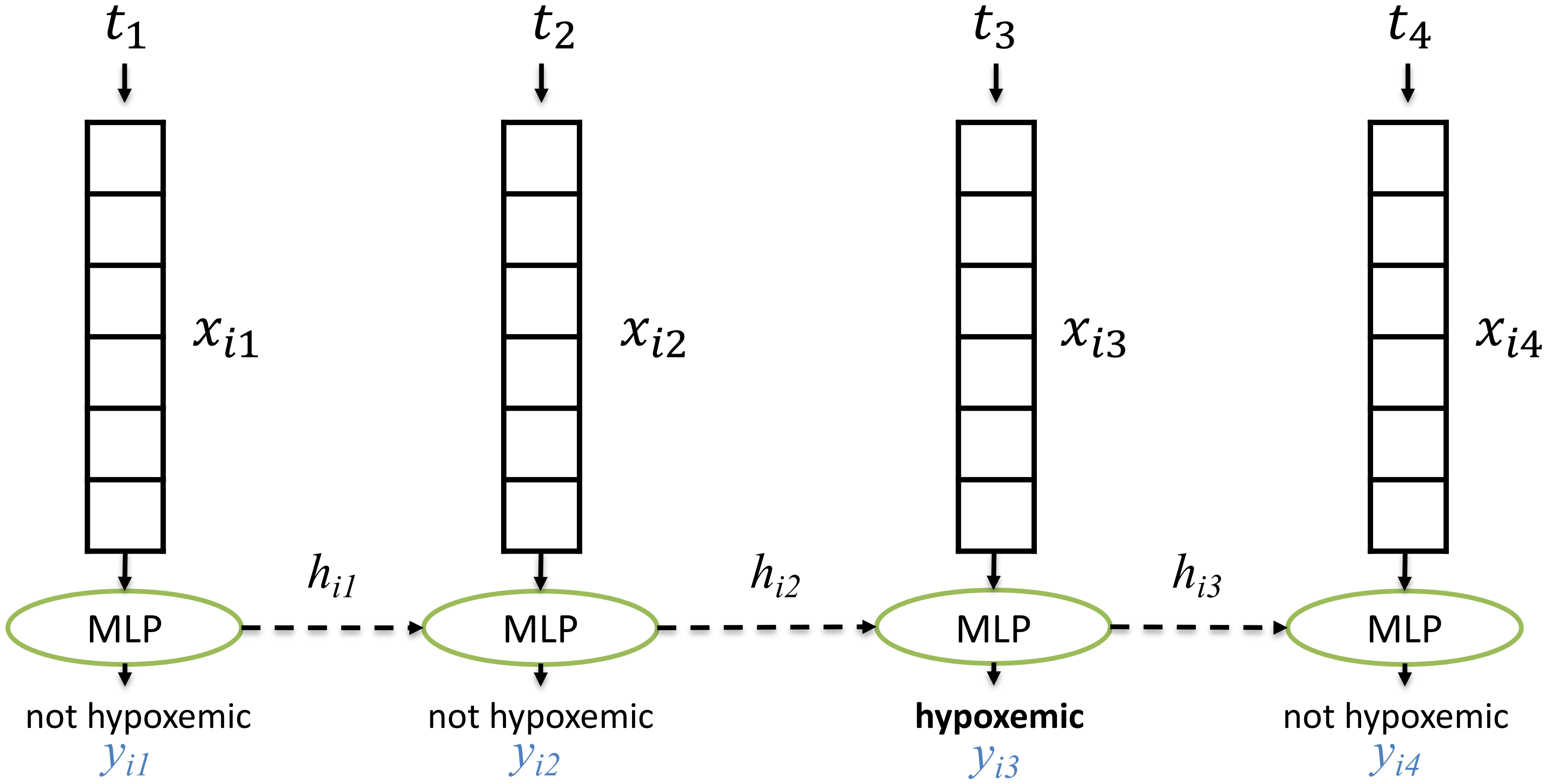
Predict a label associated with each word



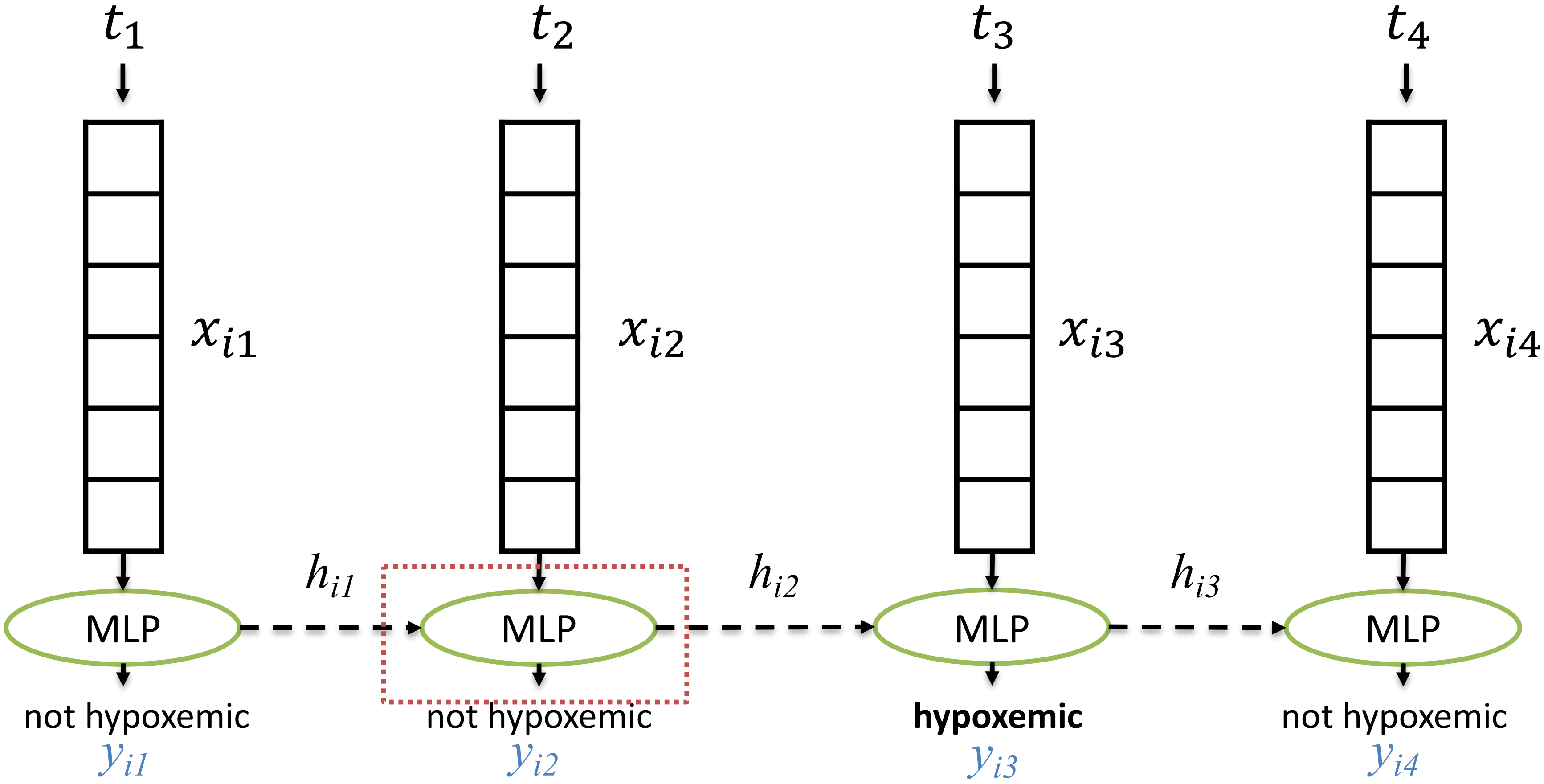
Transfer *relevant* information about earlier words



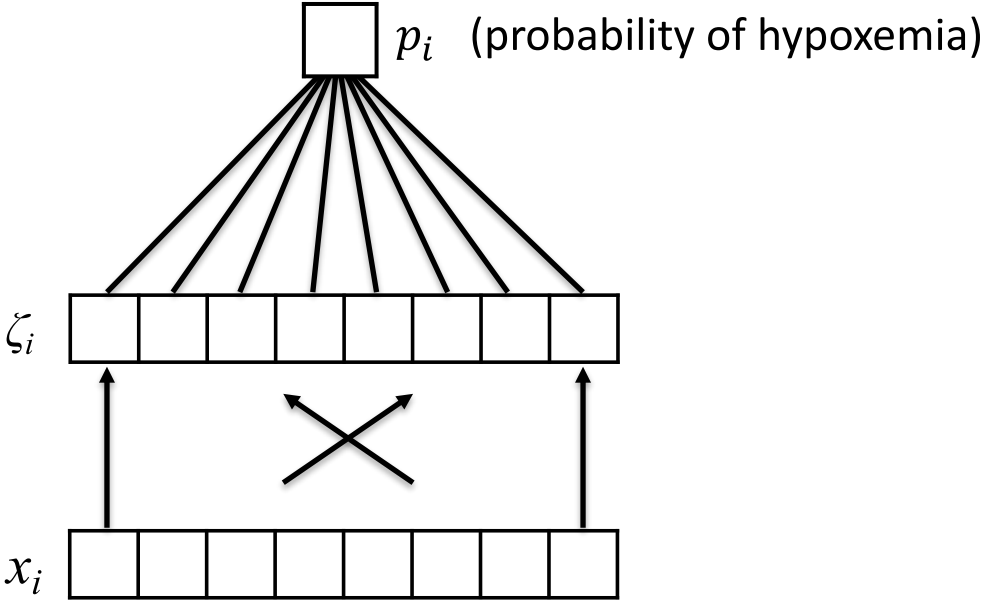
Transfer *relevant* information about earlier values



Transfer *relevant* information about earlier values



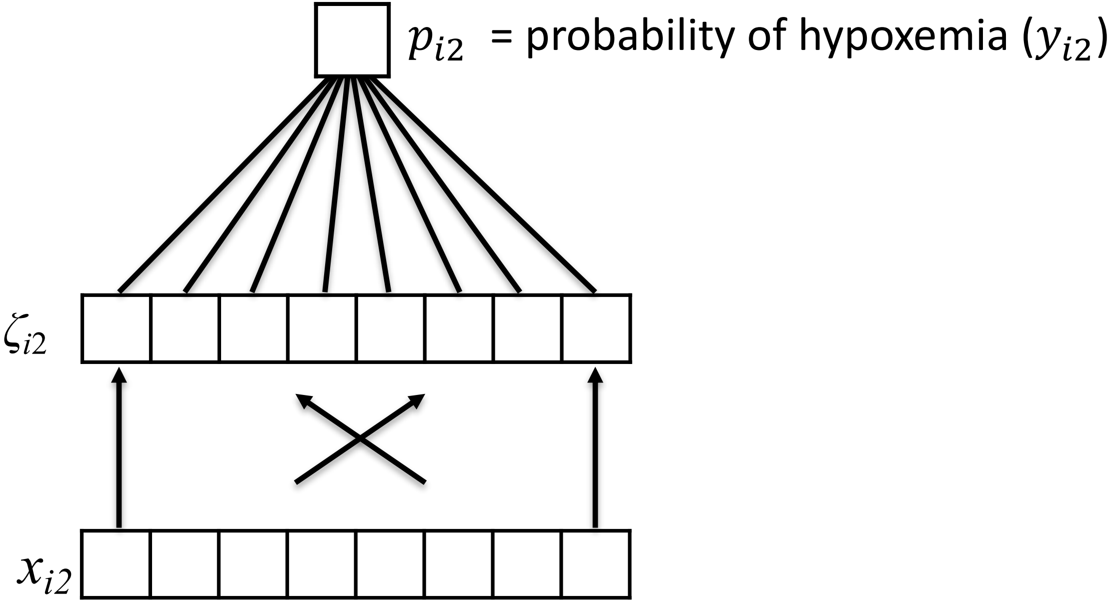
Back to Lecture 1…



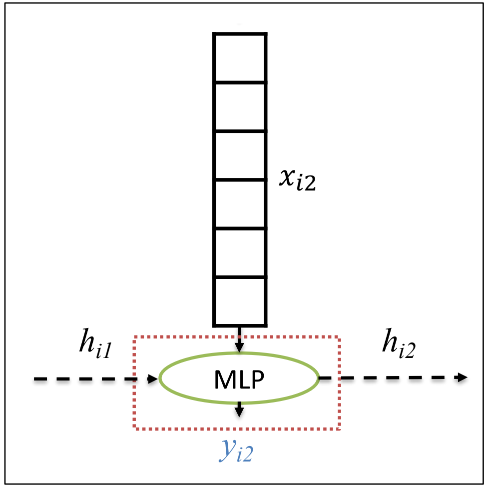
Since they are neither an input nor an output, the features ζ are said to be a “hidden” layer

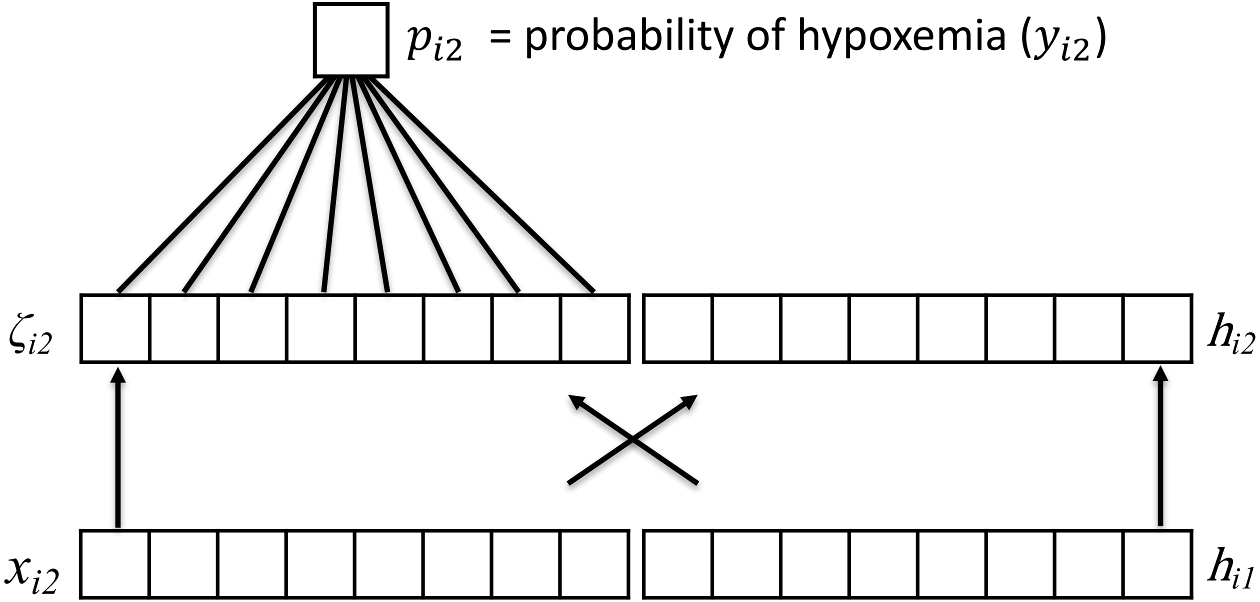
- Instead of predicting p_i directly from our feature vector x, introduce a vector of **“latent” features** ζ (zeta) that we will use to predict p_i

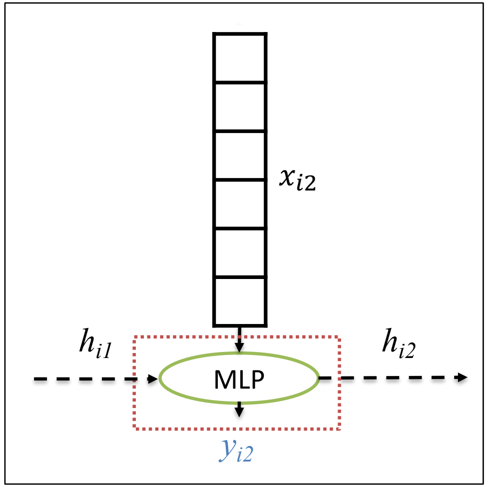
- Think of ζ as a <u>learned representation</u> that is useful for predicting *p*

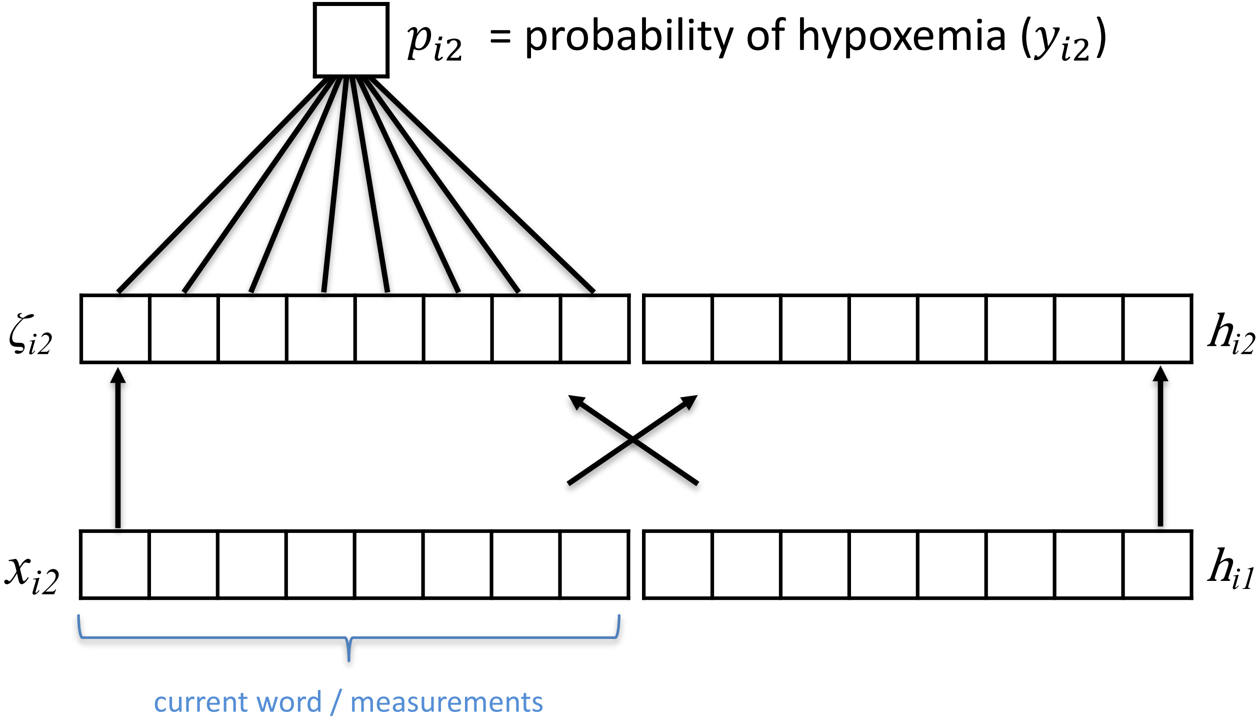


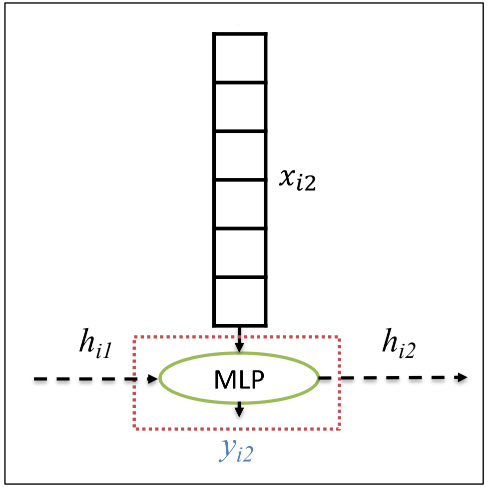
Since they are neither an input nor an output, the features ζ are said to be a "hidden" layer

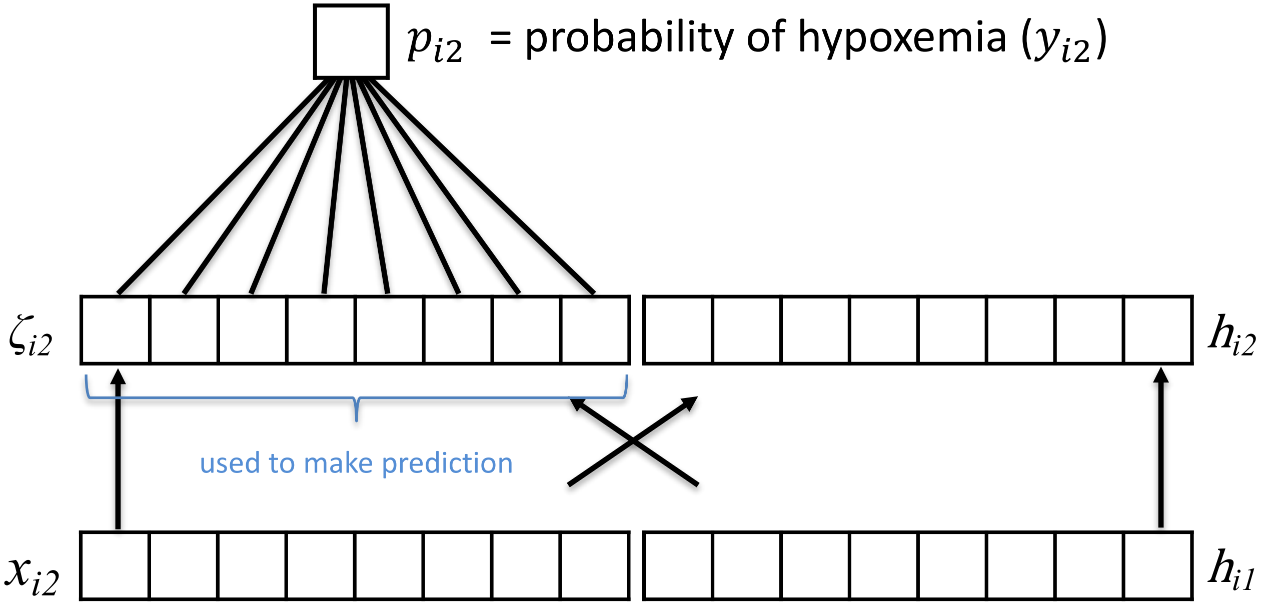


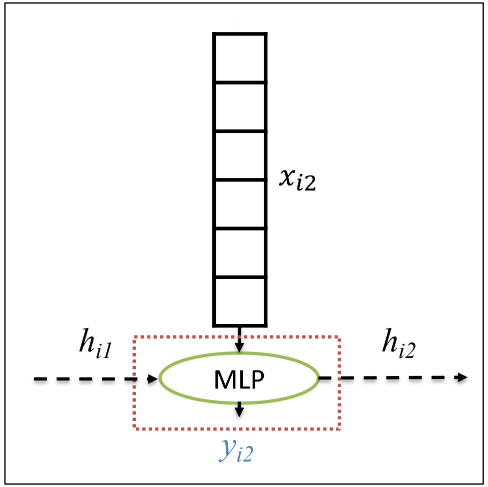

p_{i2} = probability of hypoxemia (y_{i2})
ζ_{i2}
h_{i2}
x_{i2}
h_{i1}

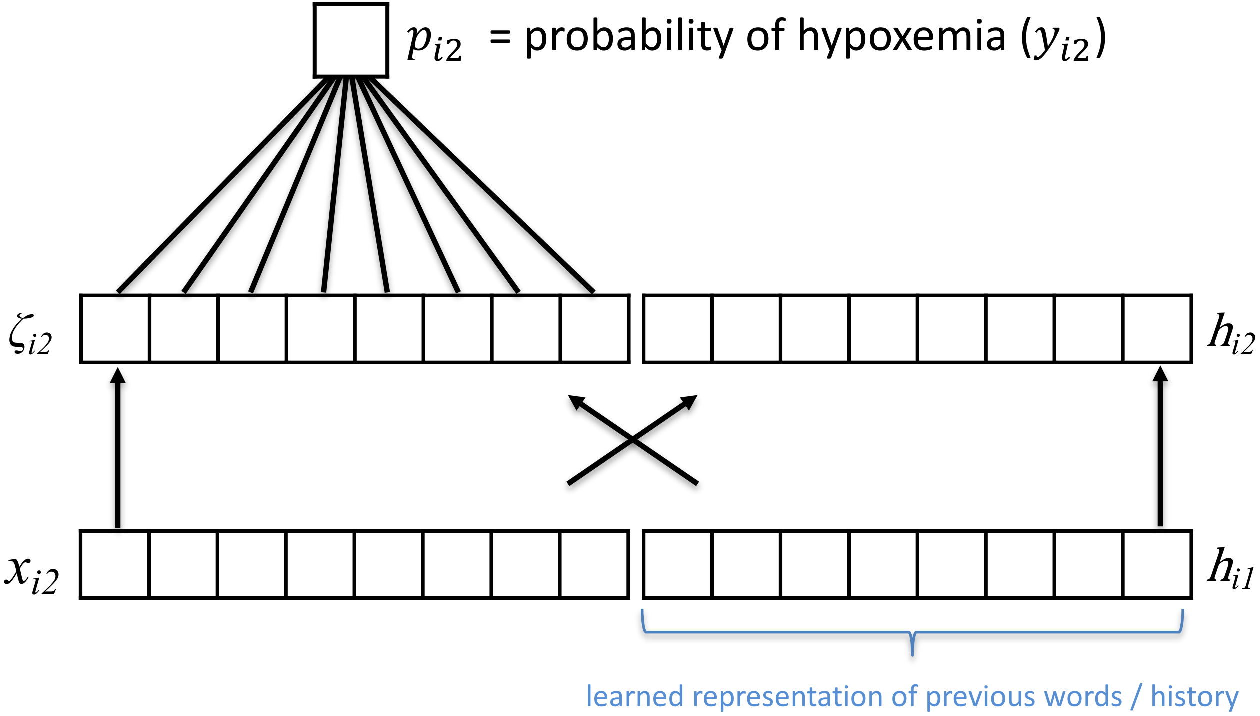

x_{i2}
h_{i1}
h_{i2}
MLP
y_{i2}

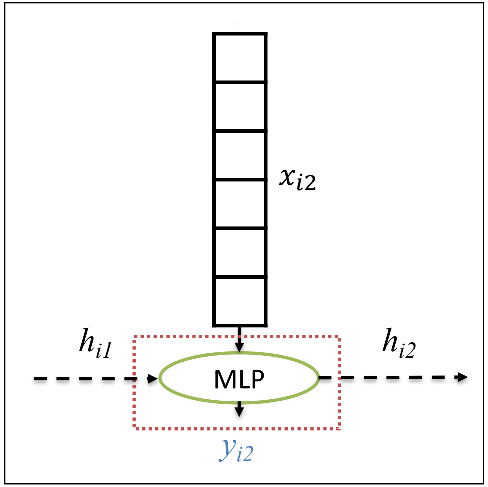

p_{i2} = probability of hypoxemia (y_{i2})
ζ_{i2}
h_{i2}
x_{i2}
h_{i1}
current word / measurements

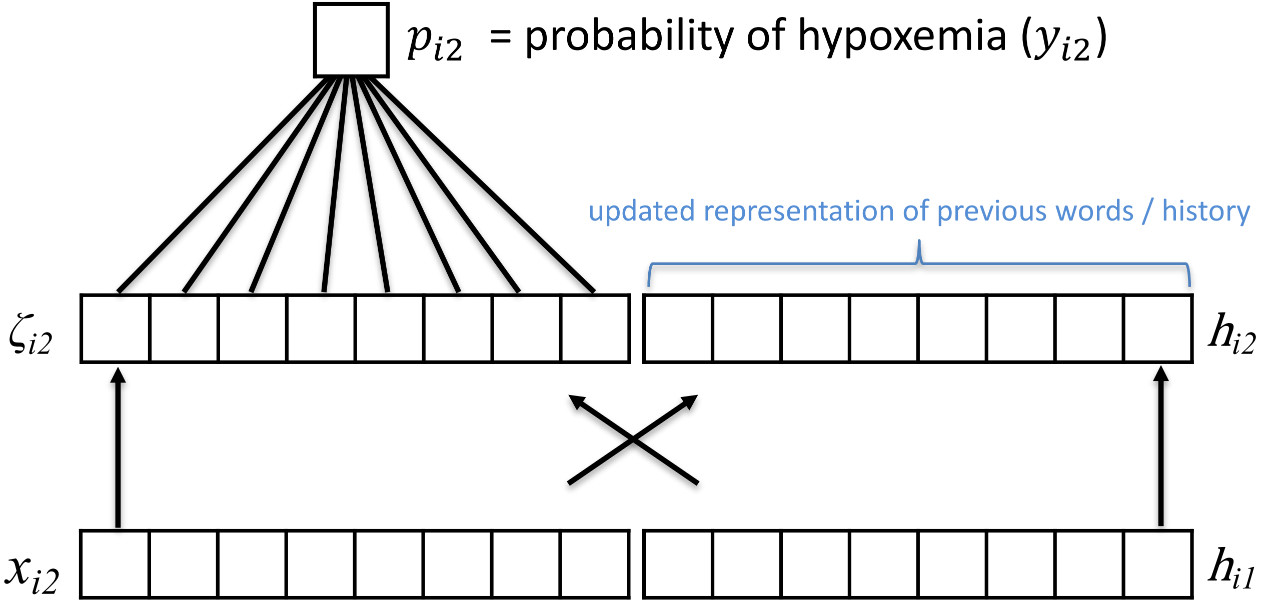

x_{i2}
h_{i1}
h_{i2}
MLP
y_{i2}

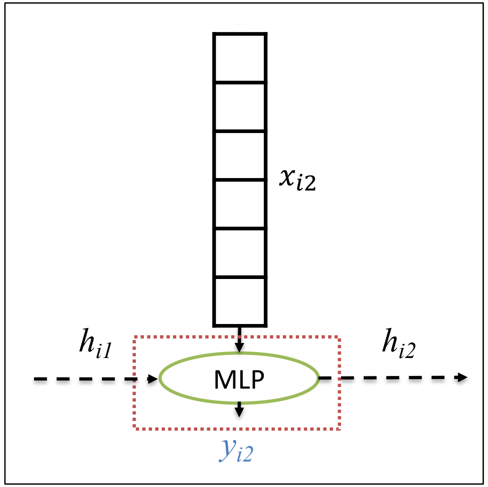

p_{i2} = probability of hypoxemia (y_{i2})
ζ_{i2}
h_{i2}
used to make prediction
x_{i2}
h_{i1}


x_{i2}
h_{i1}
MLP
h_{i2}
y_{i2}

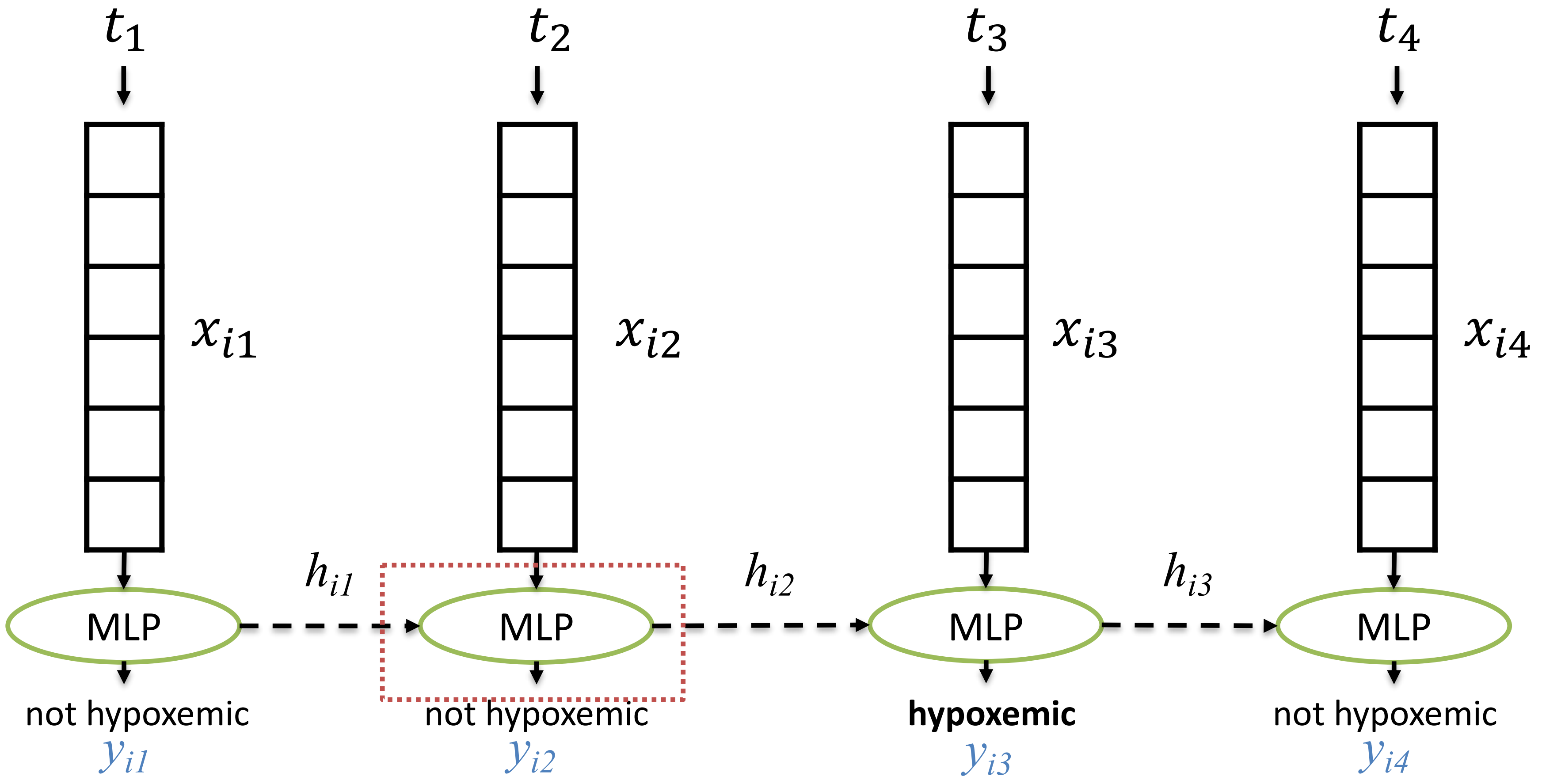

p_{i2} = probability of hypoxemia (y_{i2})
ζ_{i2}
h_{i2}
x_{i2}
h_{i1}
learned representation of previous words / history


x_{i2}
h_{i1}
h_{i2}
MLP
y_{i2}

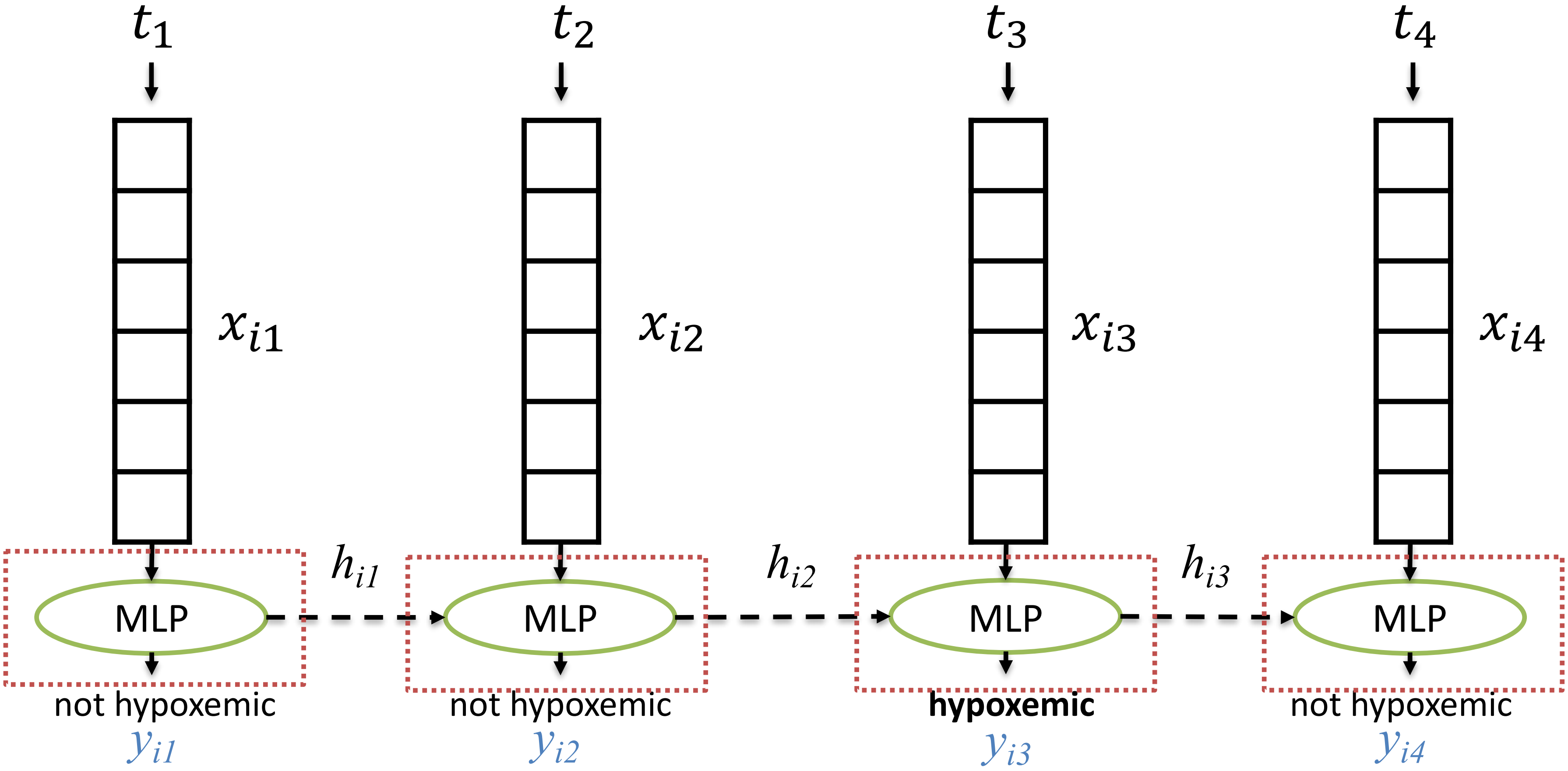

p_{i2} = probability of hypoxemia (y_{i2})
updated representation of previous words / history
ζ_{i2}
h_{i2}
x_{i2}
h_{i1}


x_{i2}
h_{i1}
h_{i2}
MLP
y_{i2}

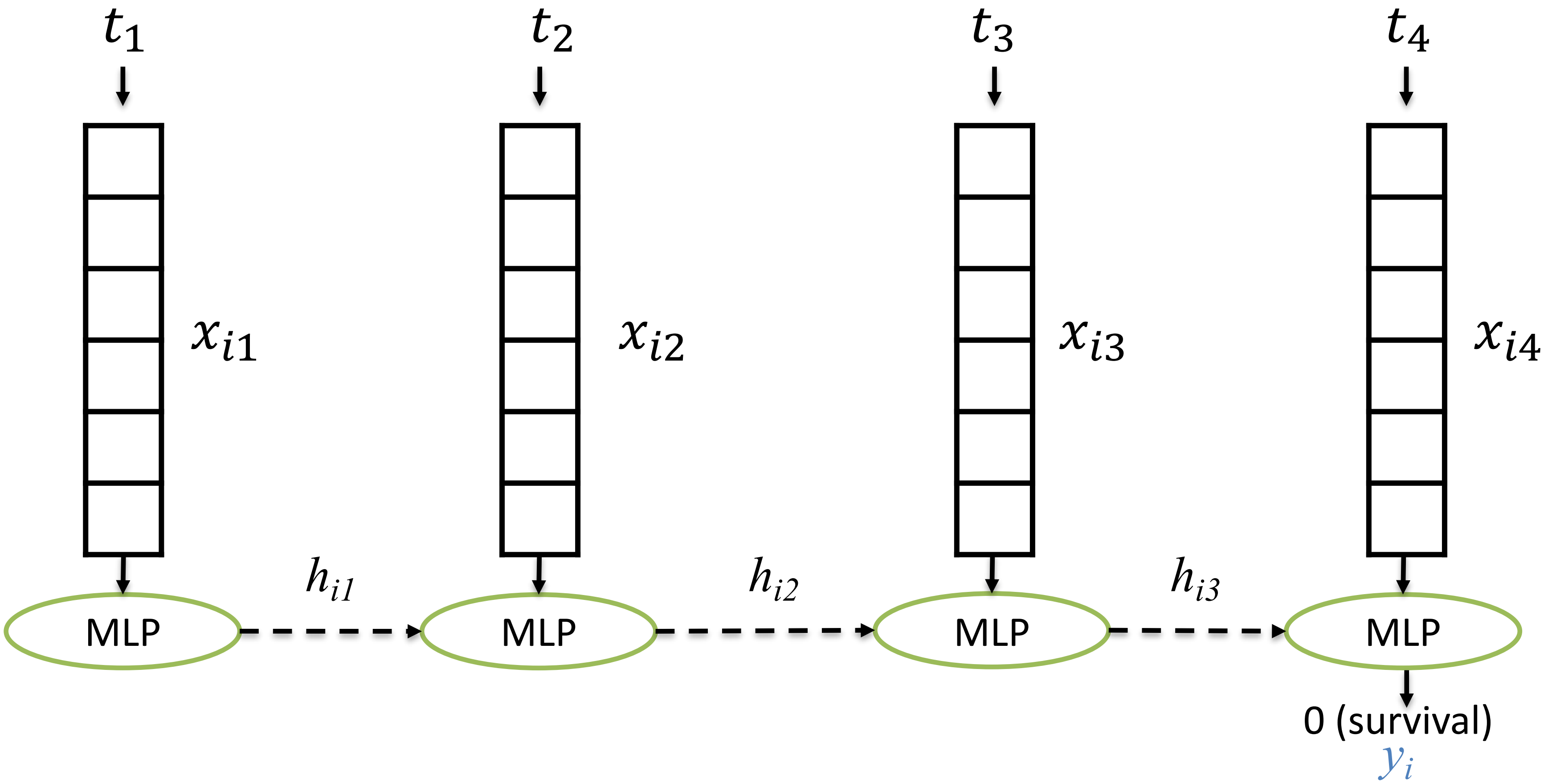
We *learn* what's important about previous values



Recurrent MLP (NN): these are all the same / have same weights

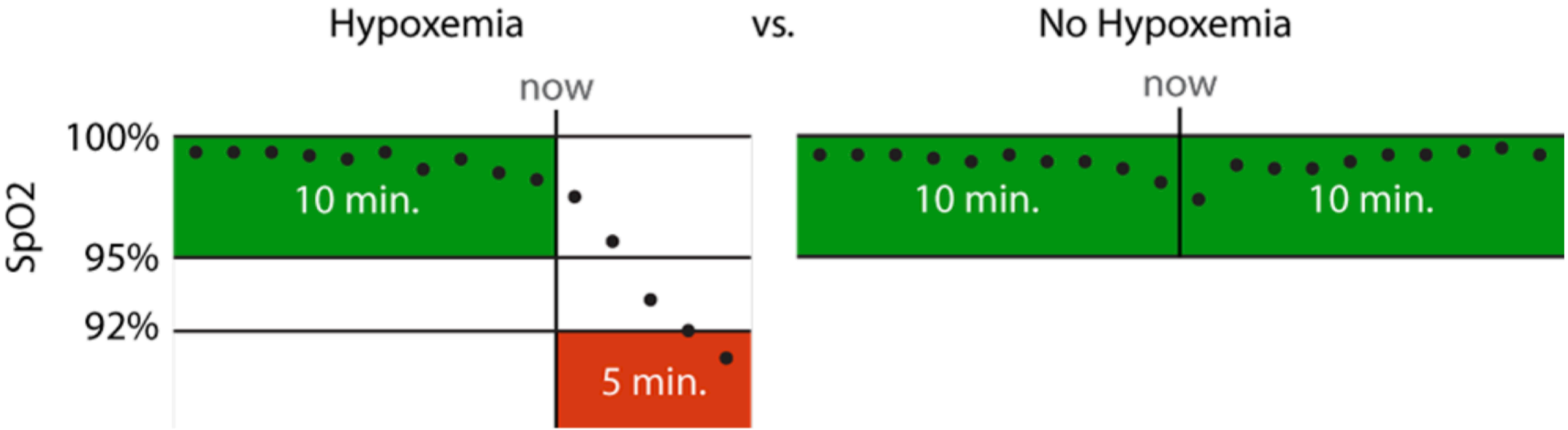


Task 1: Predict a label associated with the sequence



Hypoxemia Prediction: Use learned representation of previous measurements

Real-time Prediction Task:
- hypoxemia (yes/no) in the next 5 minutes
- based on data from the Anesthesia Information Management System
- static features + real-time features collected up to that time point



Common RNN Variants

- Gated Recurrent Unit (GRU)
- Long Short Term Memory (LSTM)
- Bidirectional RNNs
  - Look at previous words and upcoming words
  - Usually not appropriate for time-series

Deidentification of Patient Notes

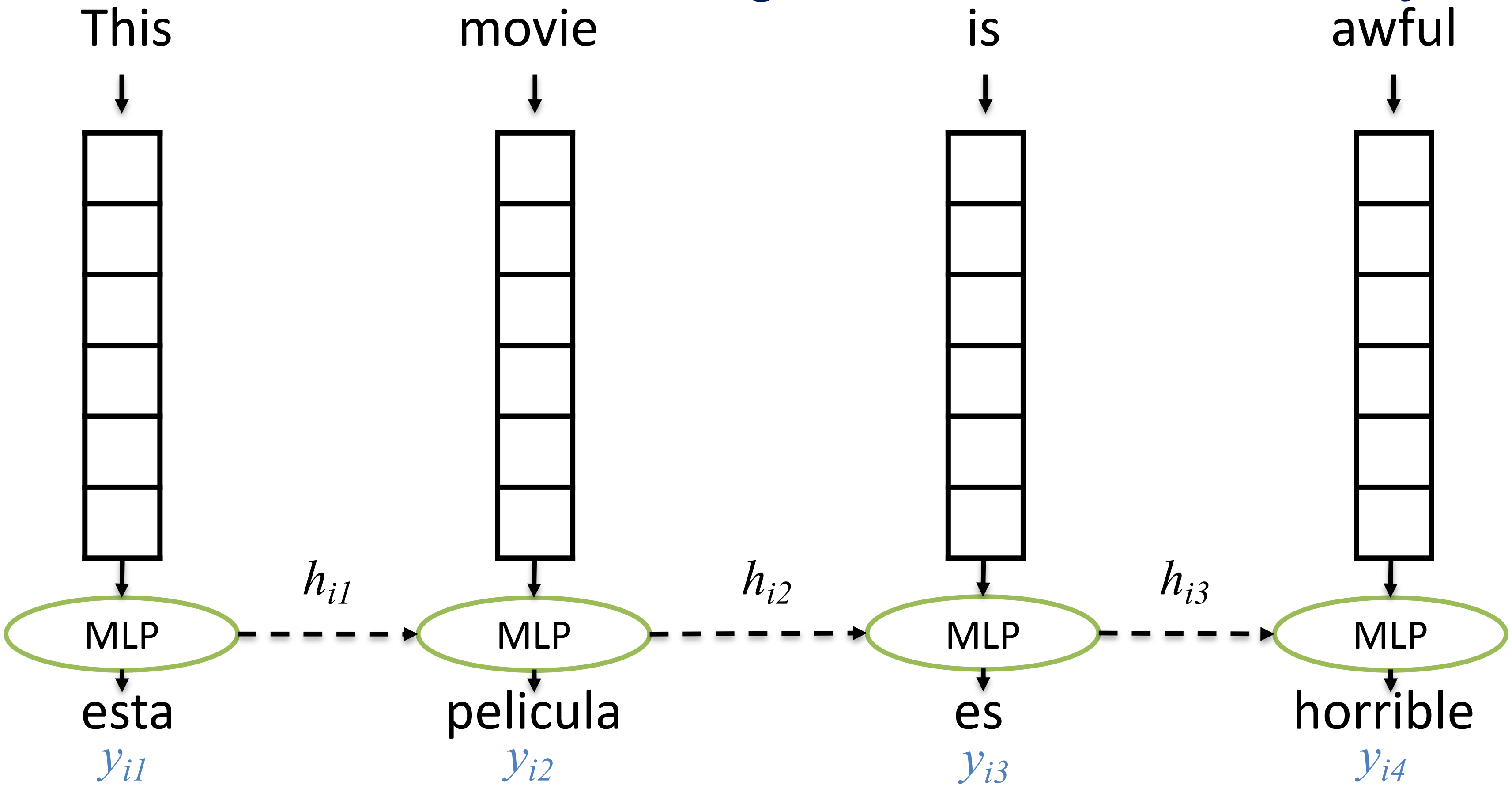
Table 5. Examples of correctly detected PHI instances (in bold) by the ANN

PHI category	ANN
AGE	Father had a stroke at 80 and died of?another stroke at age Personal data and overall health: Now 63, despite his FH: Father: Died @ 52 from EtOH abuse (unclear exact etiology) Tobacco: smoked from age 7 to 15, has not smoked since 15.
CONTACT	History of Present Illness 86F reports worsening b/l leg pain. by phone, Dr. Ivan Guy. Call w/ questions 86383. Keith Gilbert, H/O paroxysmal afib VNA 171-311-7974 ======= Medications
DATE	During his May hospitalization he had dysphagia Social history: divorced, quit smoking in 08, sober x 10 yrs, She is to see him on the 29th of this month at 1:00 p.m. He did have a renal biopsy in teh late 60s adn thus will look for results, Results02/20/2087 NA 135, K 3.2 (L), CL 96 (L), CO2 30.6, BUN 1 Jose Church, M.D. /ray DD: 01/18/20 DT: 01/19/:0 DV: 01/18/20

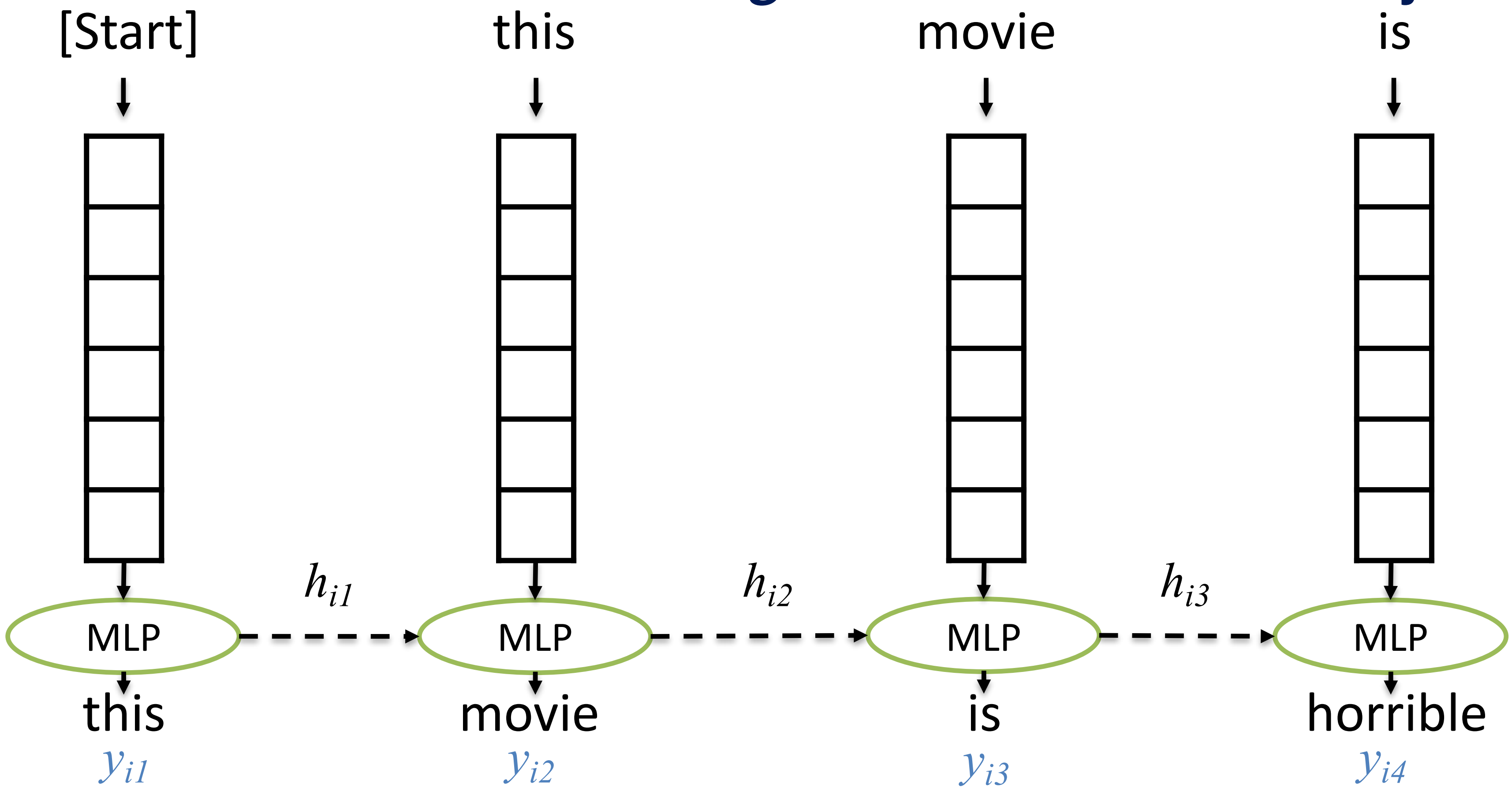
De-identification of patient notes with recurrent neural networks
Dernoncourt F, Lee JY, Uzuner O, Szolovits P
JAMIA 24(3), 2017, 596–606

- A bidirectional RNN is used to identify PHI (18 HIPAA fields)
- *i2b2*: 889 discharge summaries, >28k PHI tokens
- *MIMIC*: 1635 discharge summaries, >60k PHI tokens
- State of the art sensitivity and F1 metric on both datasets

Note: we can also *generate* text this way.



Note: we can also *generate* text this way.

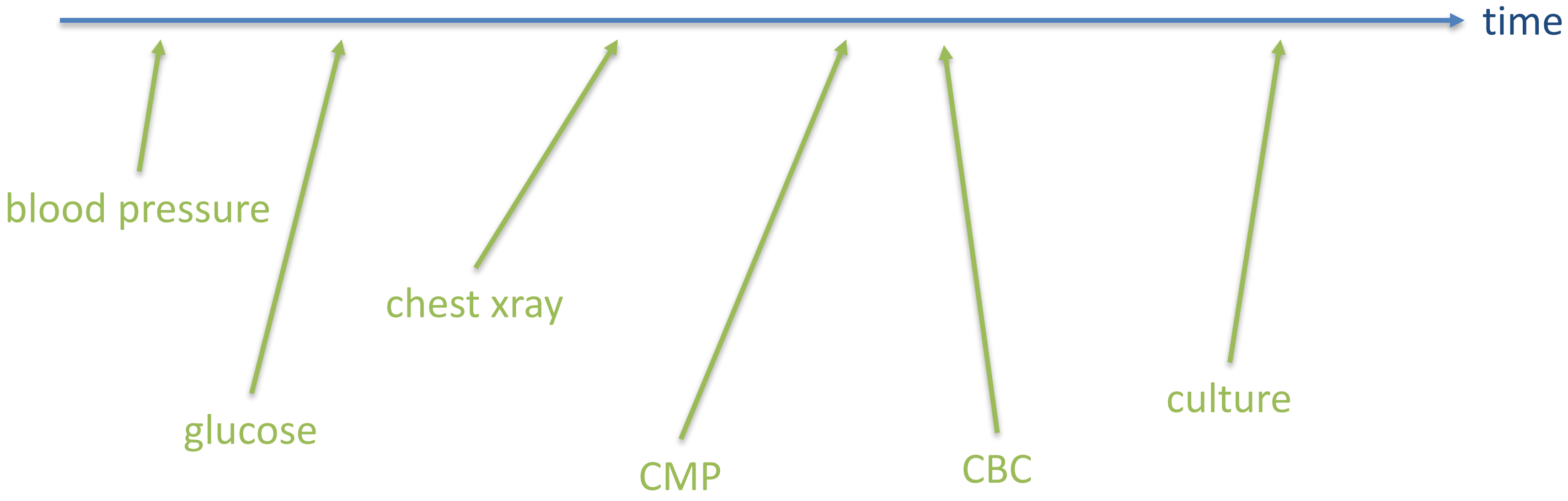


WORKING WITH IRREGULARLY SPACED MEASUREMENTS

Still Another Problem…

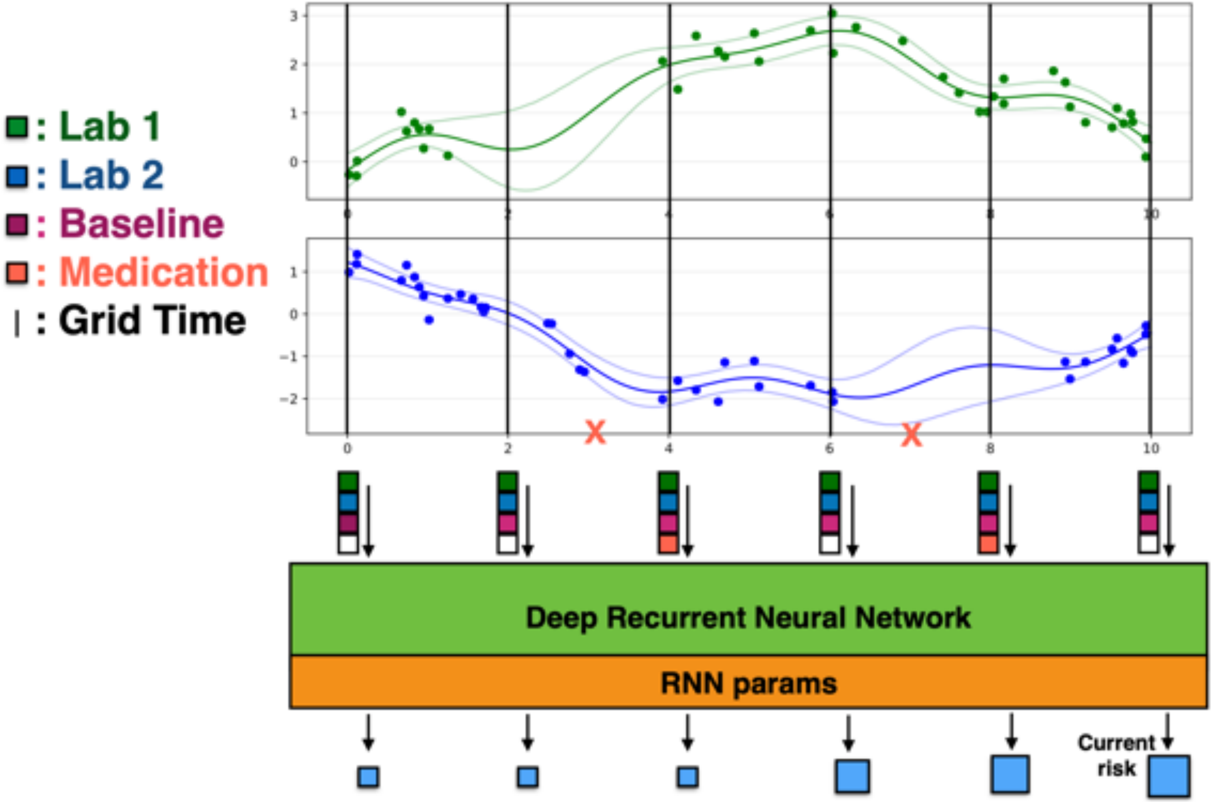
- All of this supposes we have a nice grid of complete measurements
- For text, we do have this.
- But in real-world time-series data – and particularly in healthcare – we usually have incomplete sets of measurements at irregular intervals
- How do we use an RNN?

Measurements on the Wards…



This is a major difficulty!

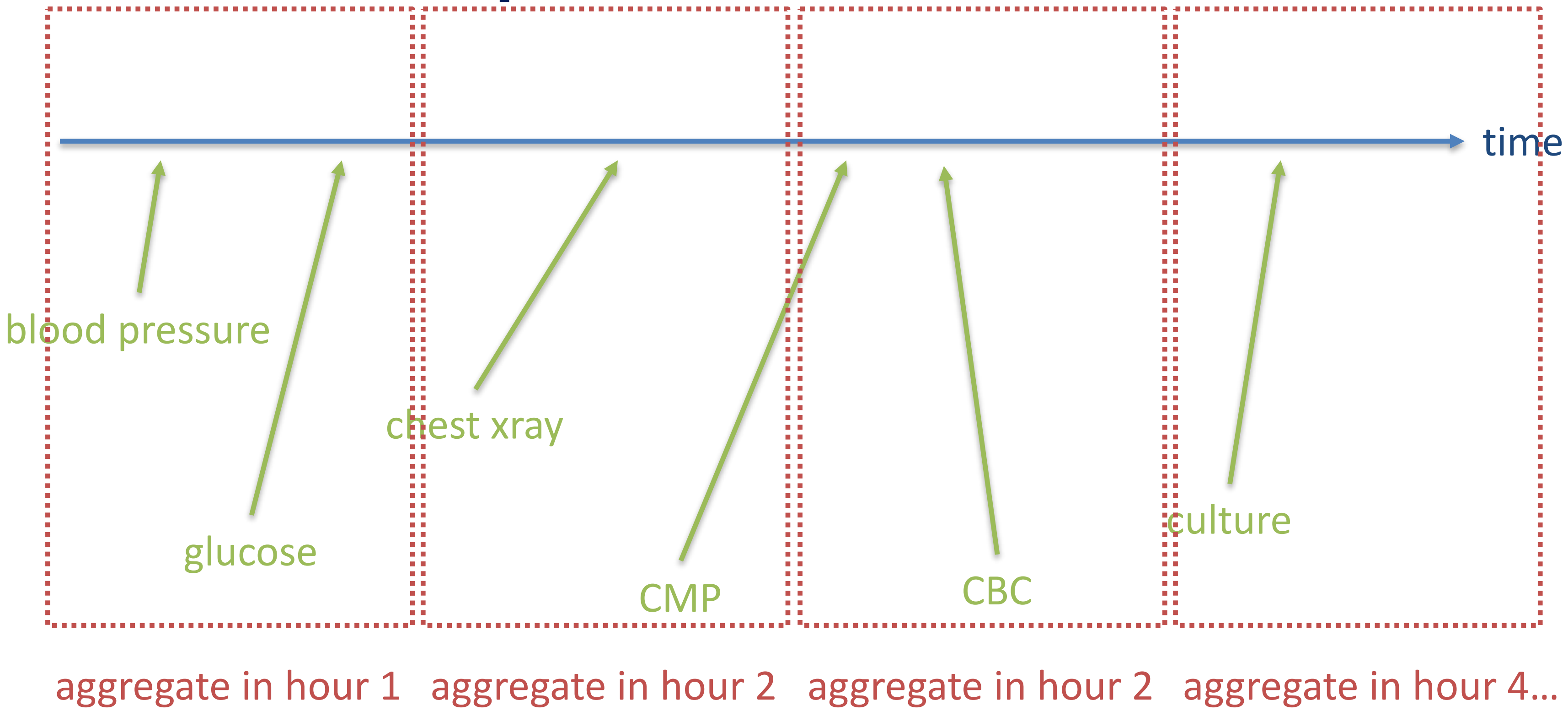
DIHI Sepsis Watch



<- Use GP regression to predict measurements at regular intervals

<- Predict sepsis risk using an RNN

Simplest Method…



In the EHR, measurements are highly “sparse”

- Many more missing measurements than non-missing

- Consider diagnosis codes, procedure codes, uncommon labs, etc

- We want to learn from these measurements, but most patients don’t have them

Conclusions

- Often, aggregating measurements/features is sufficient
- RNNs allow us to learn a representation of earlier measurements (or words) that helps us make predictions. But, in can be time and memory intensive to train.
- The RNN is just that: a recurrent / repeating MLP block
- Sparse data (most values are missing) is common in health applications and remains very challenging