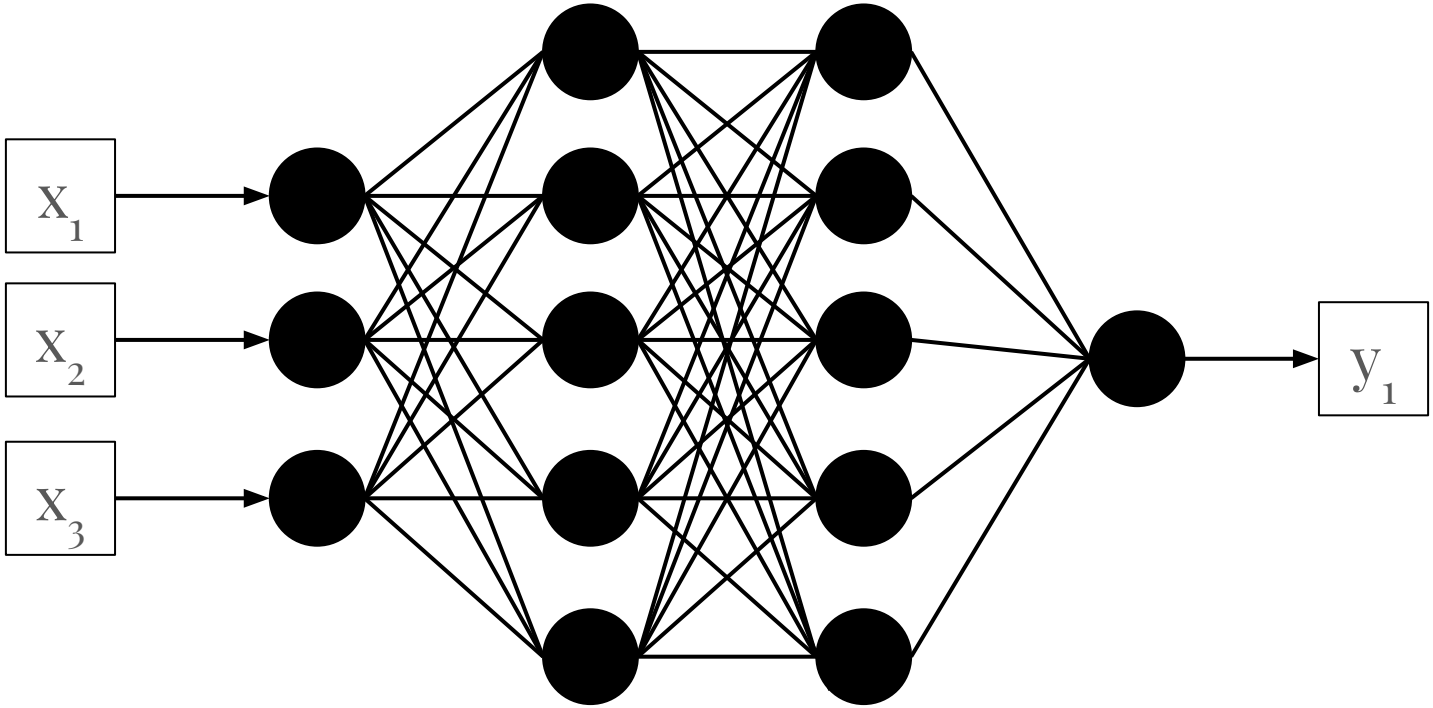
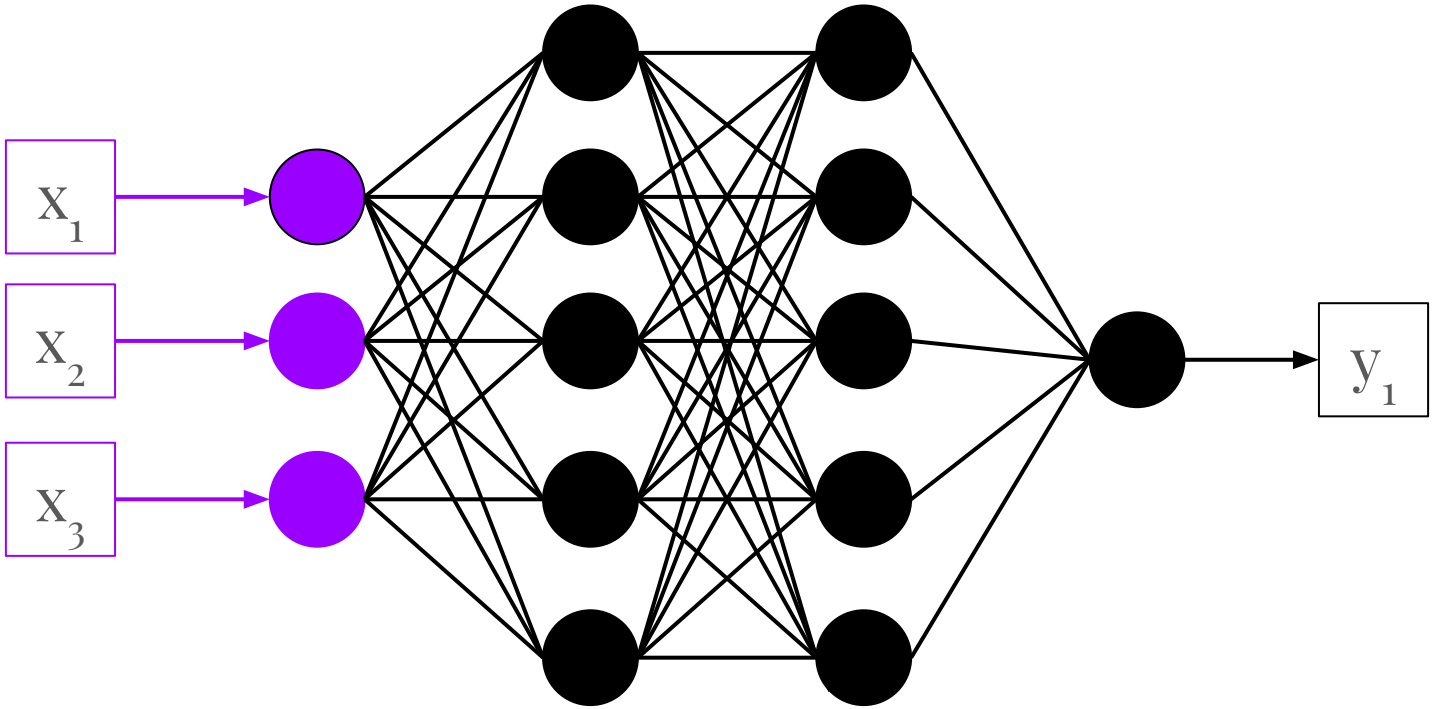
Transformers and Large Language Models (LLMs)

Dr. Julie Butler
DSC 340 Week 13 Slides

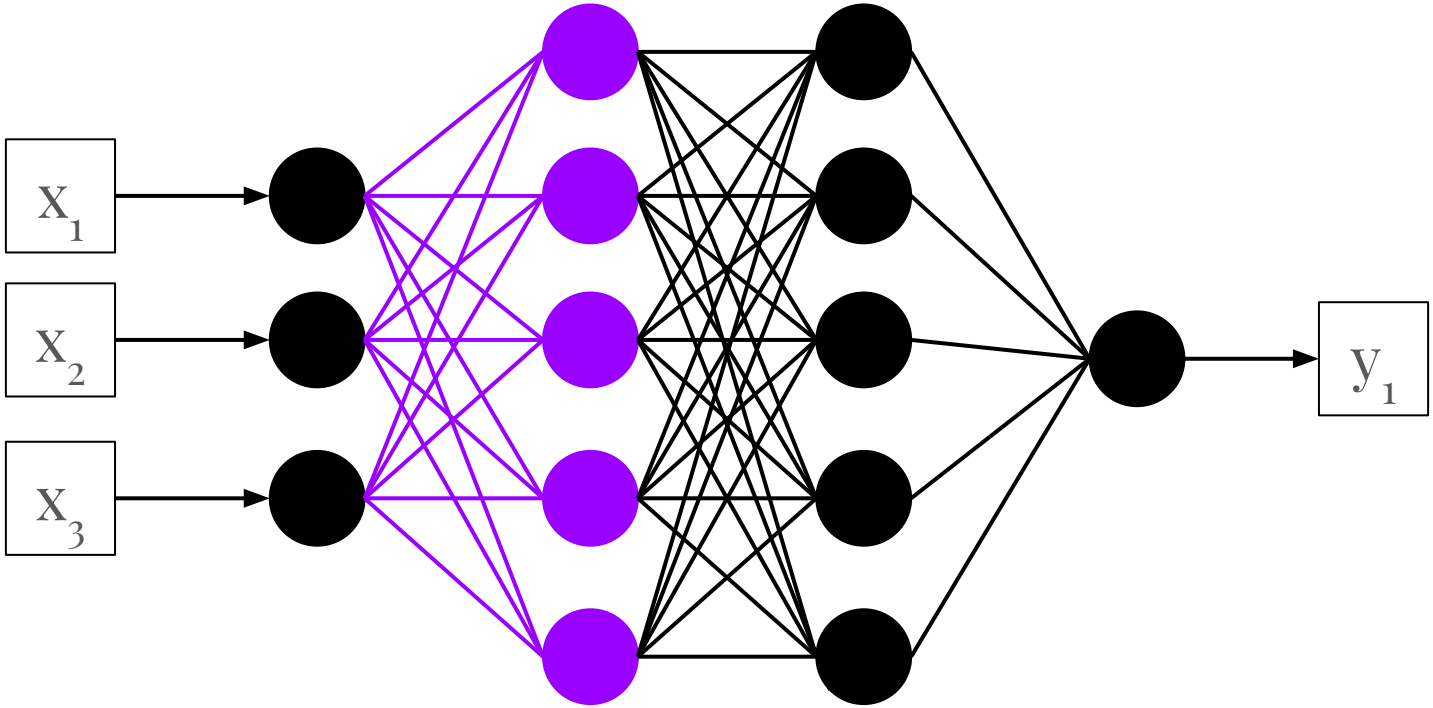
Neural Networks



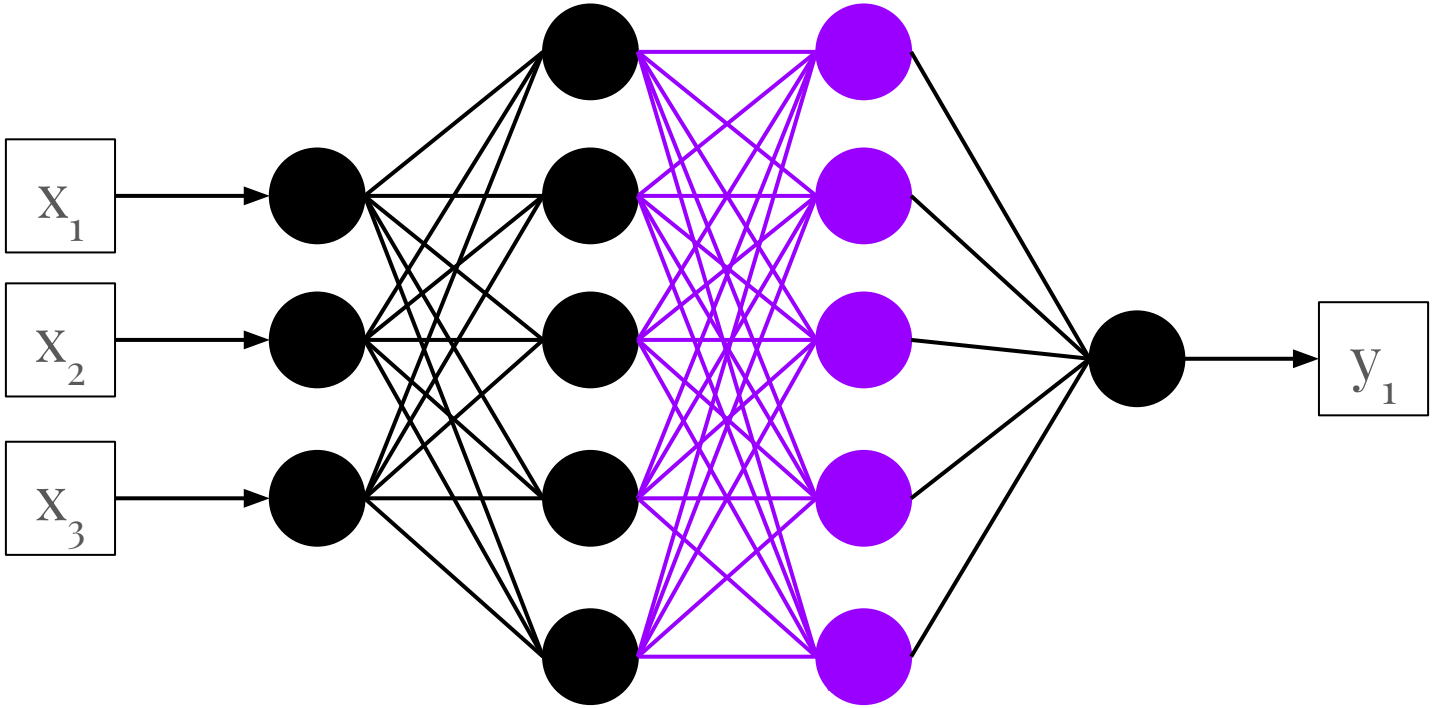
Feedforward Neural Network



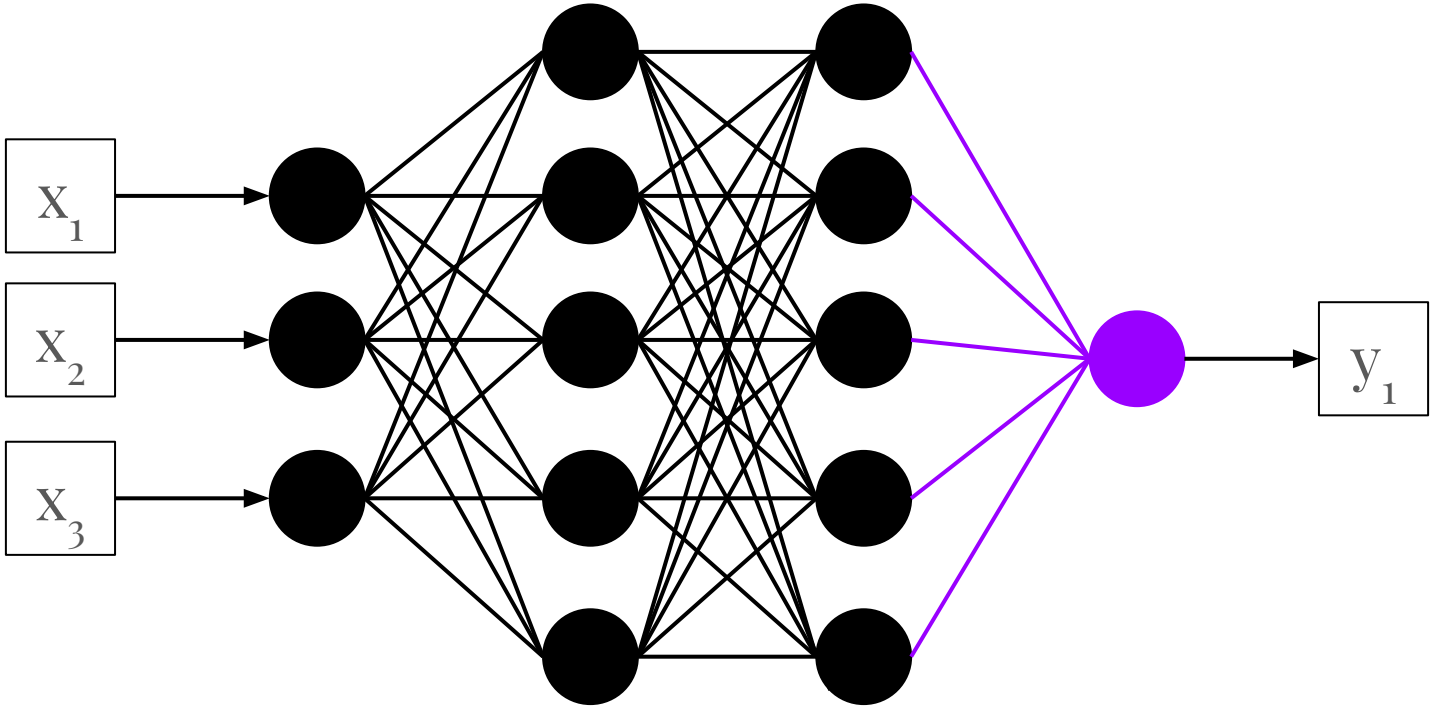
Feedforward Neural Network



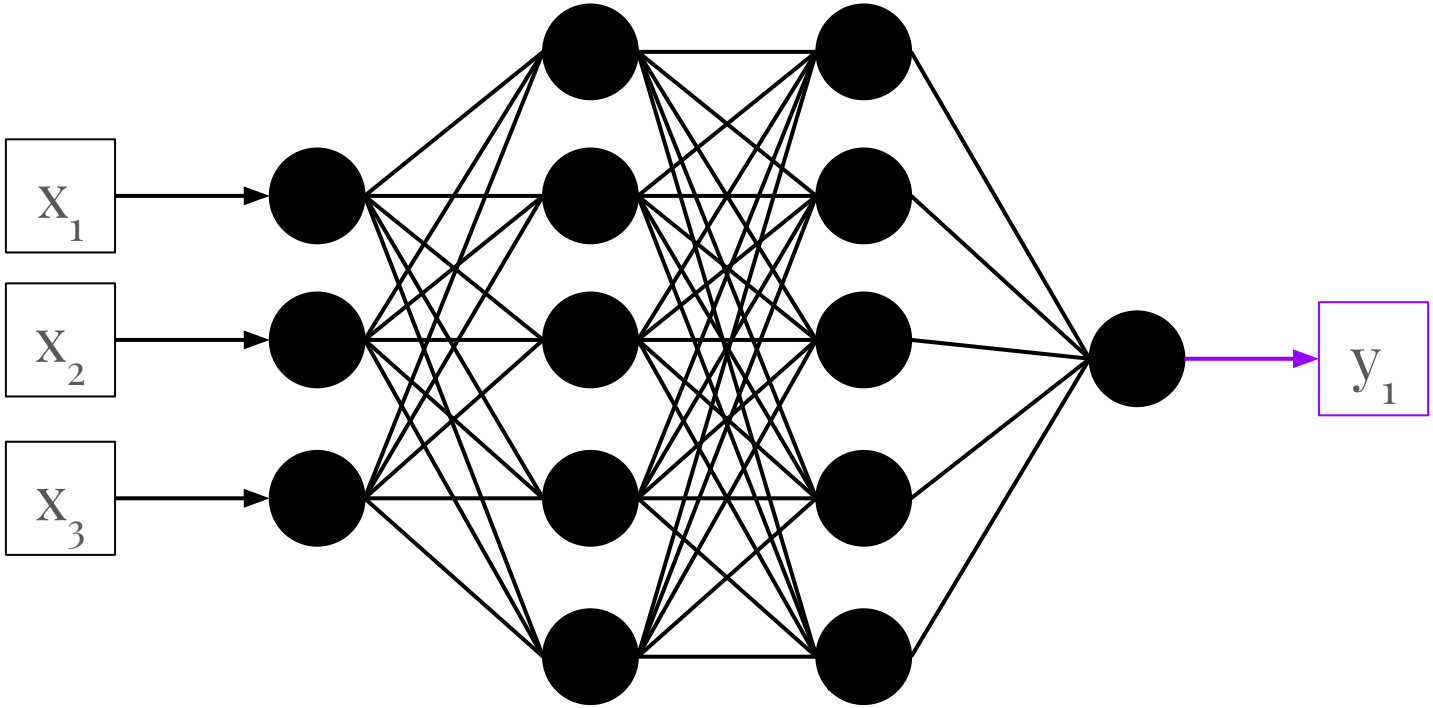
Feedforward Neural Network



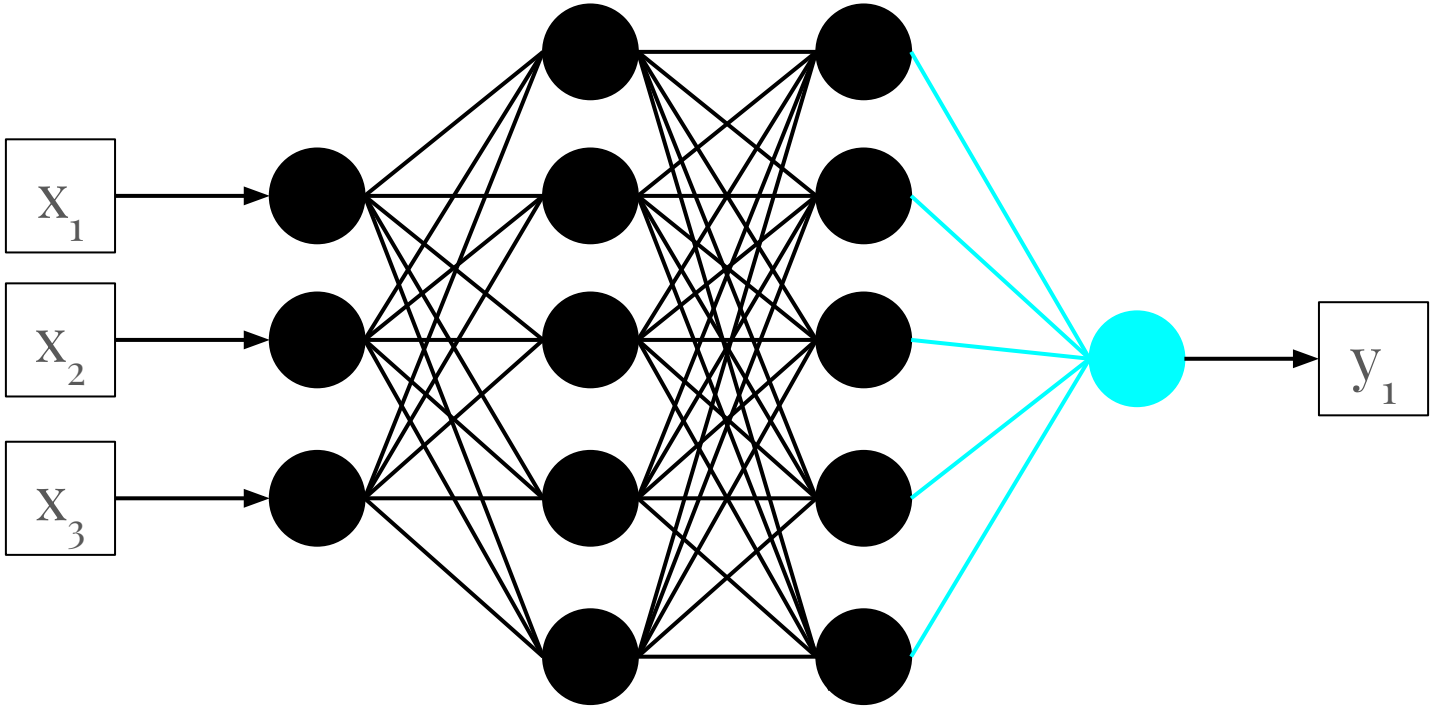
Feedforward Neural Network



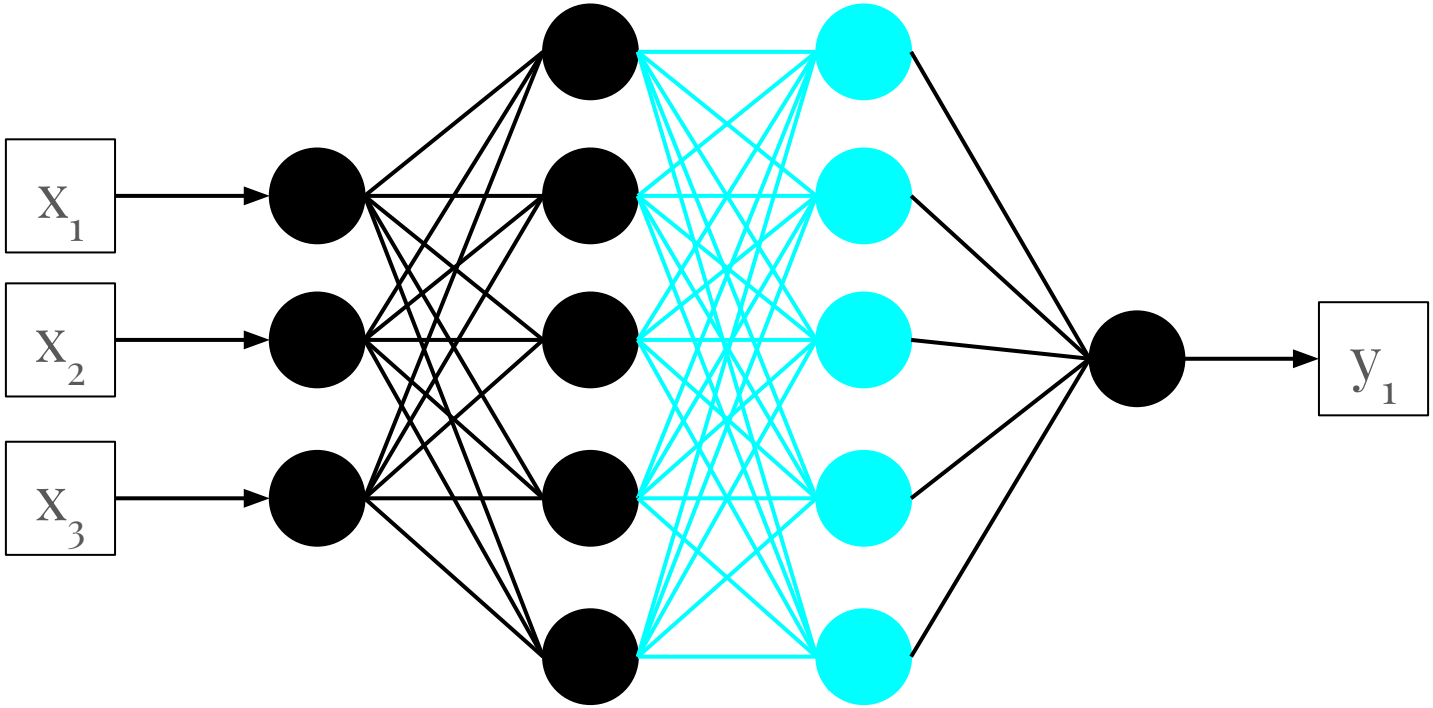
Feedforward Neural Network



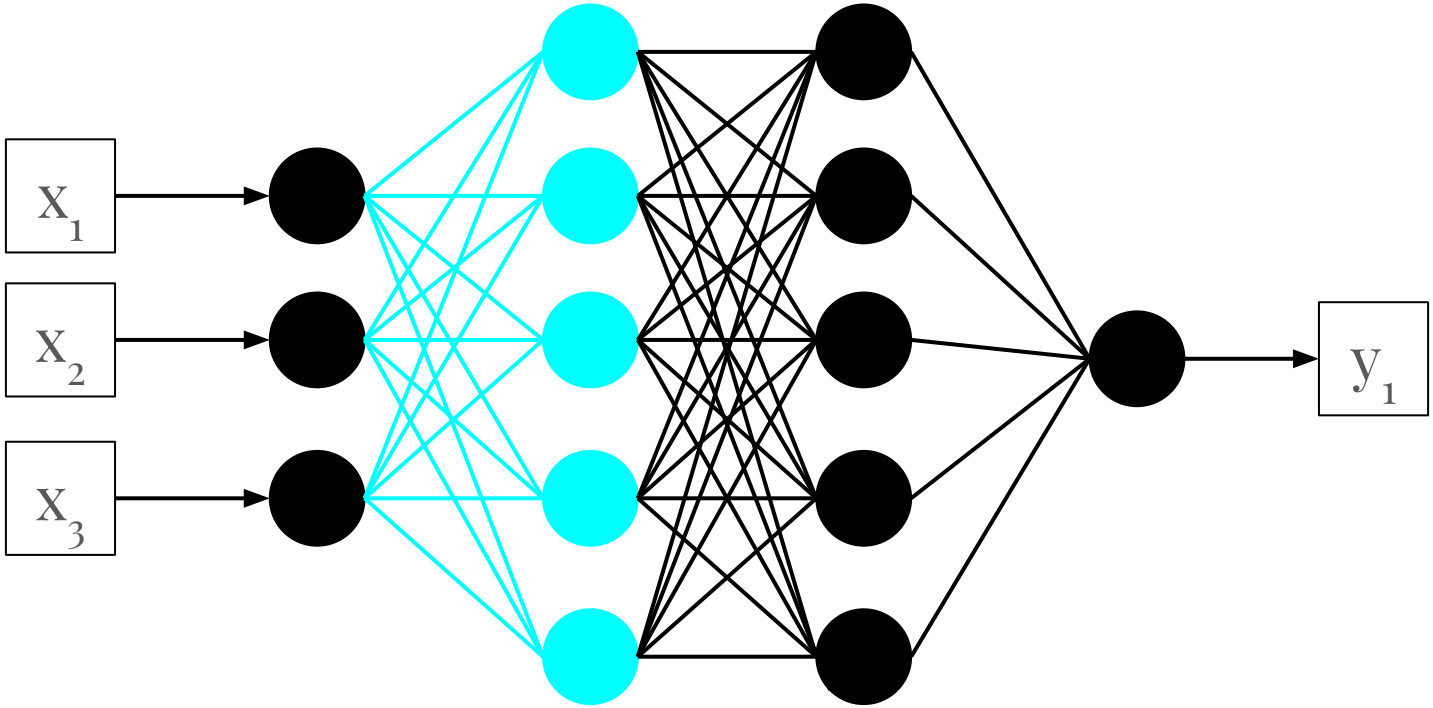
Feedforward Neural Network



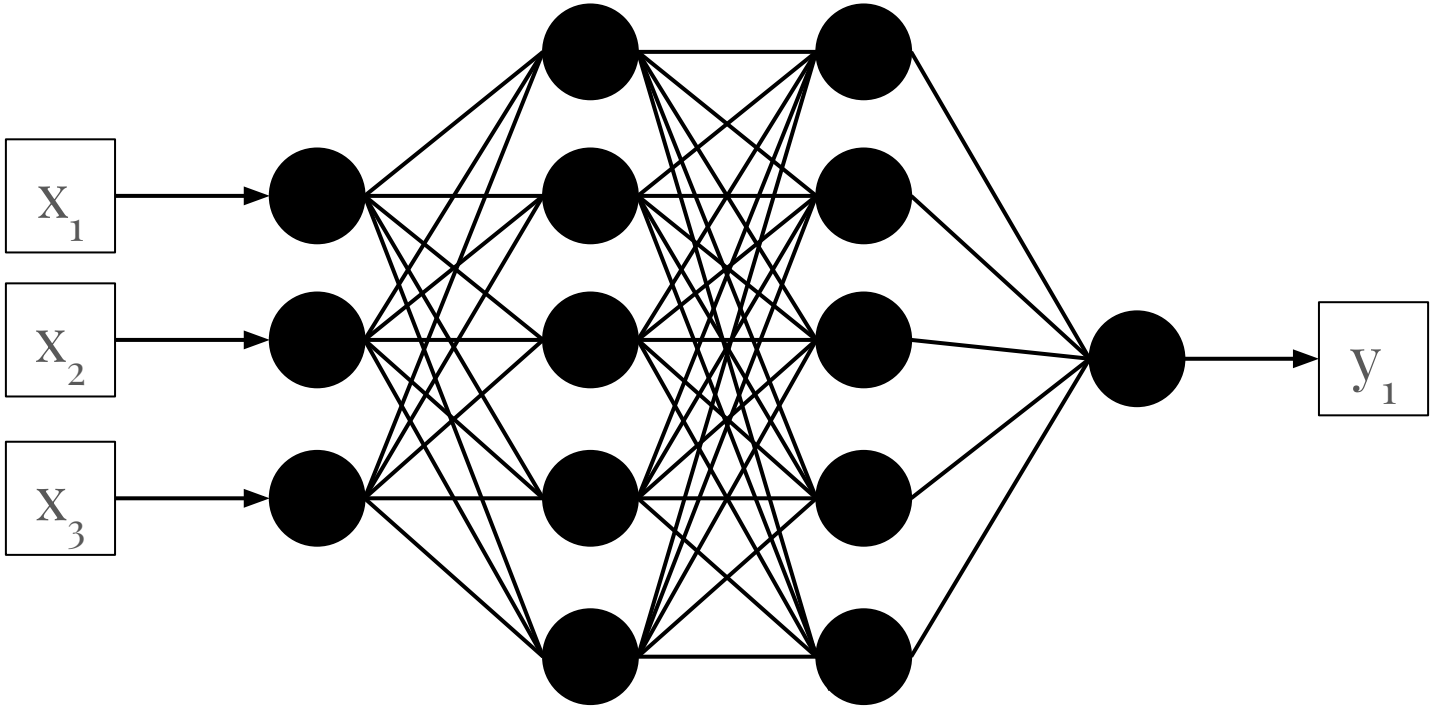
Feedforward Neural Network



Feedforward Neural Network

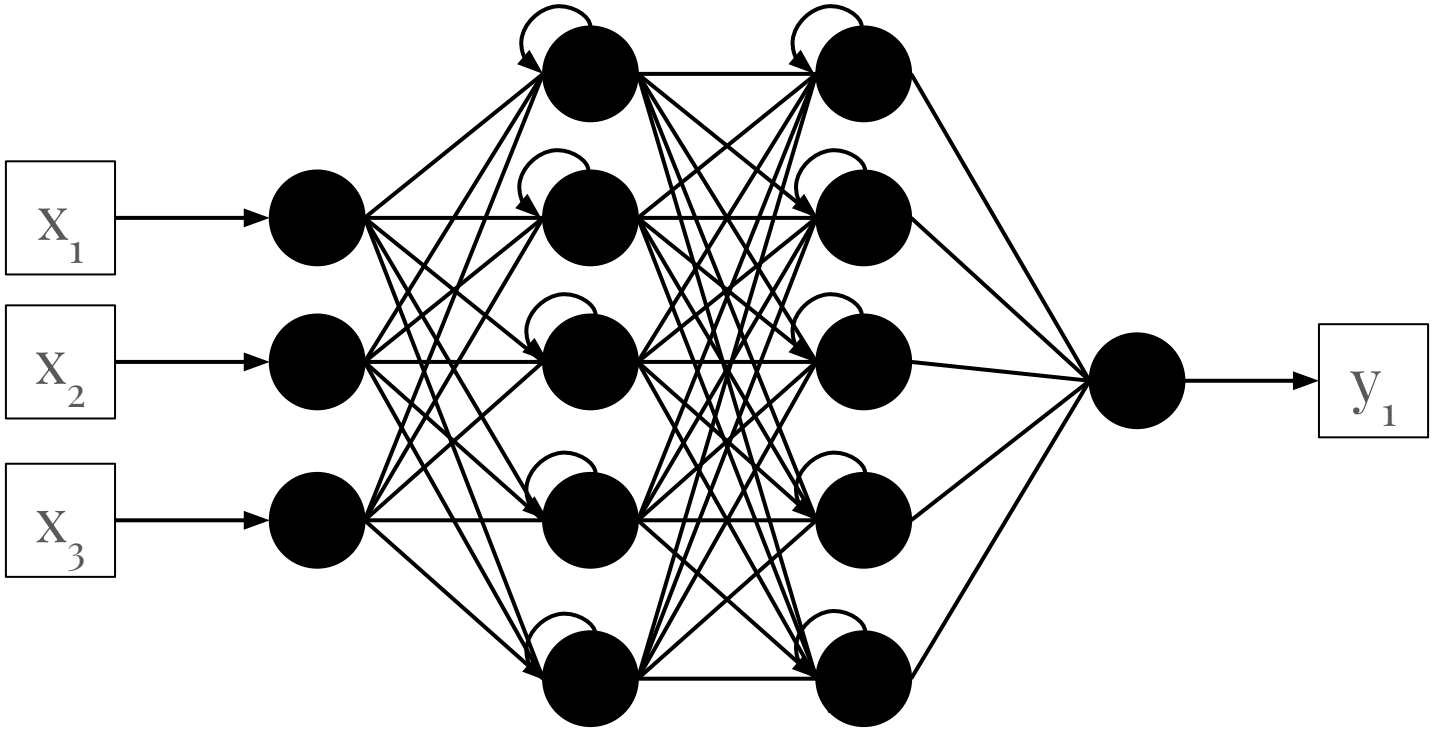


Feedforward Neural Network

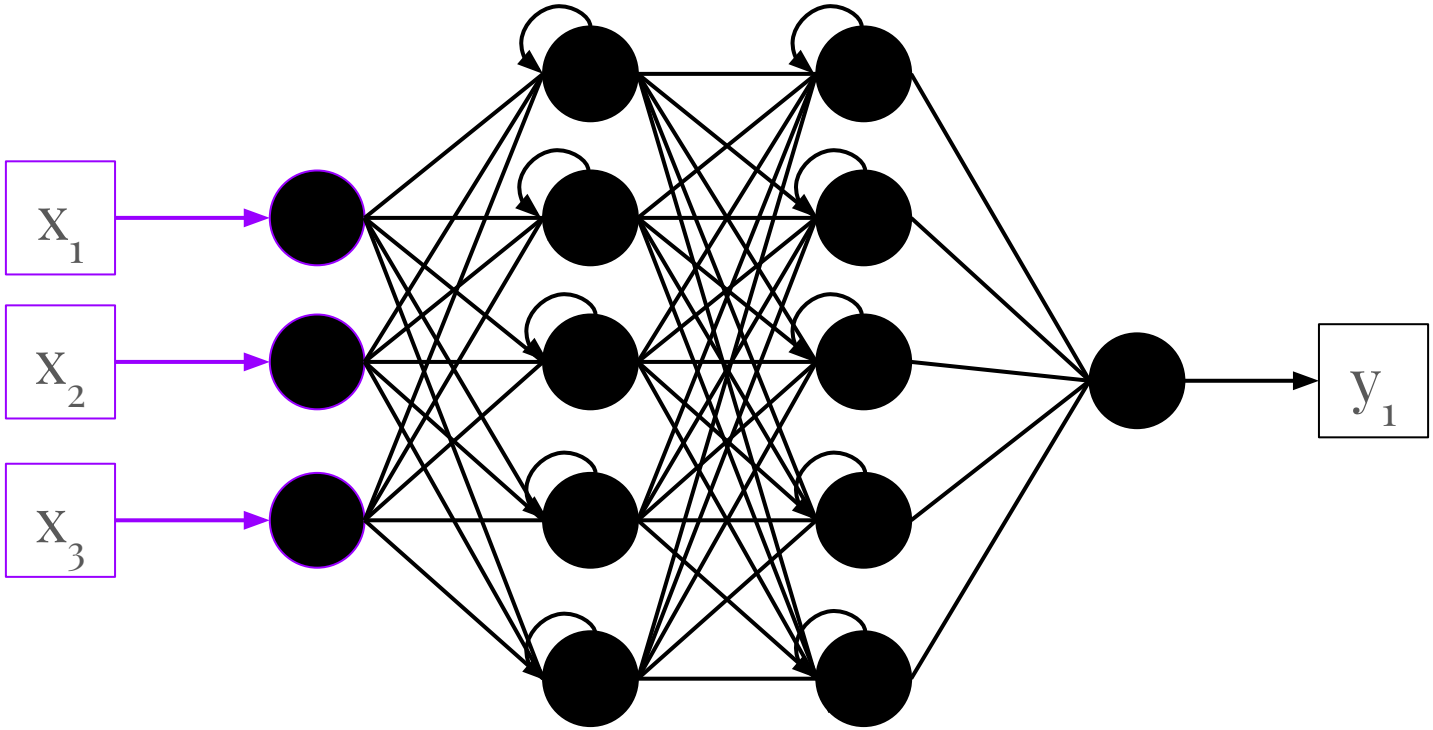


Feedforward Neural Network

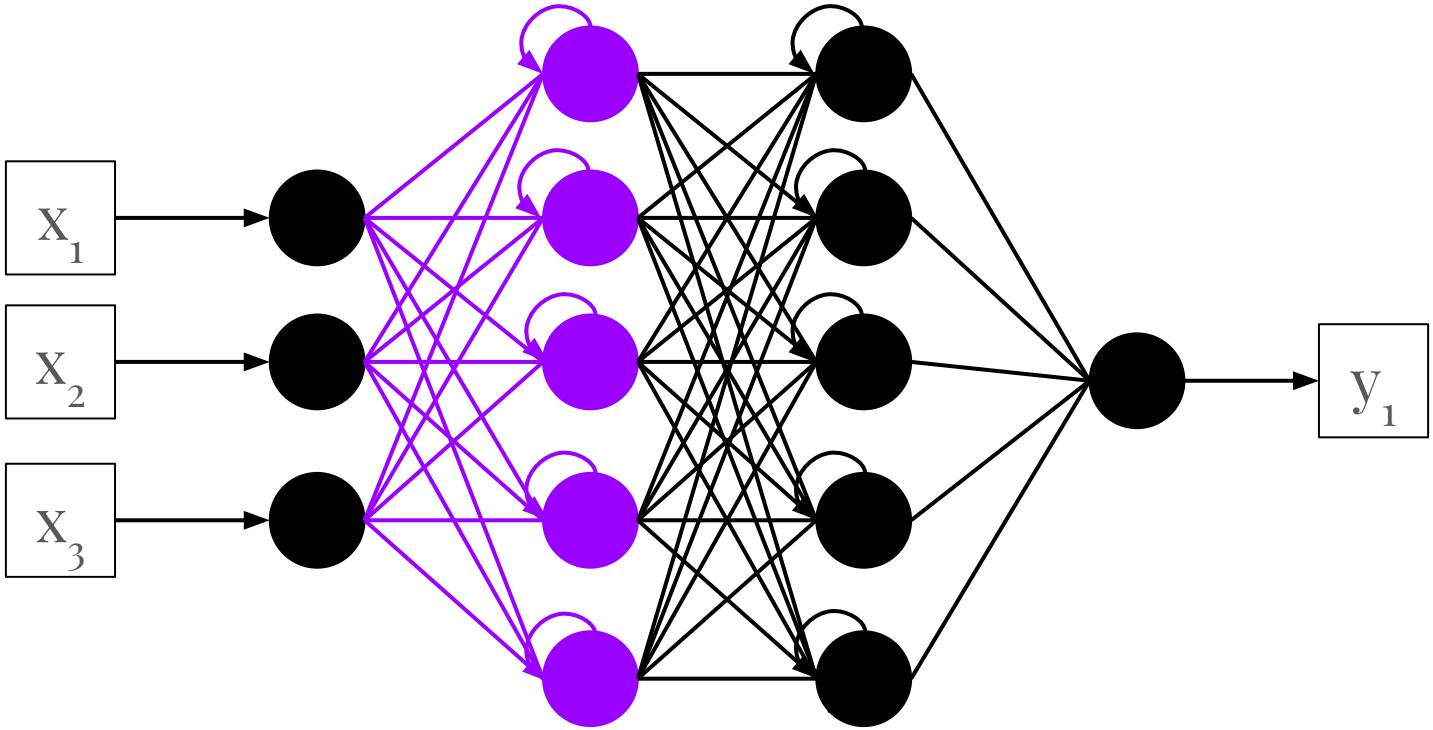
Recurrent Neural Networks



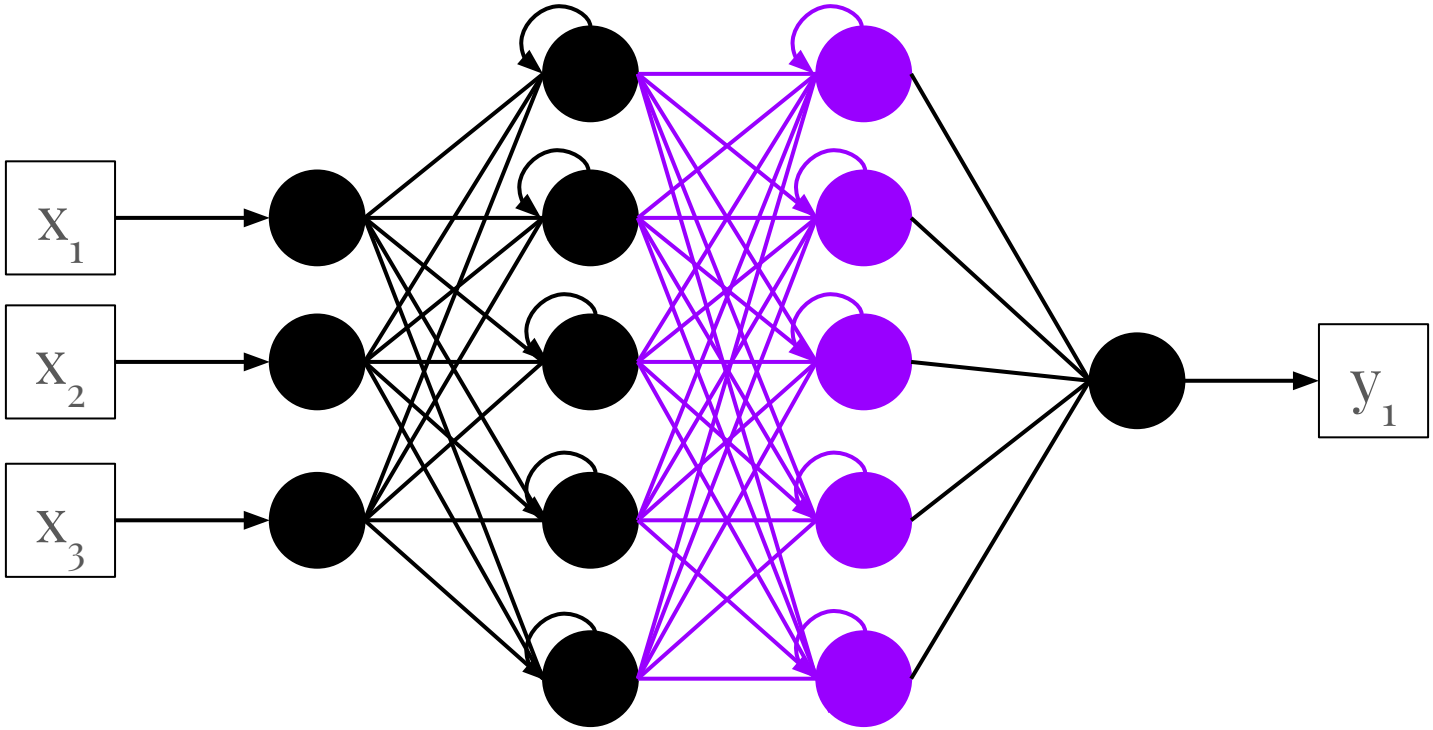
Simple Recurrent Neural Network



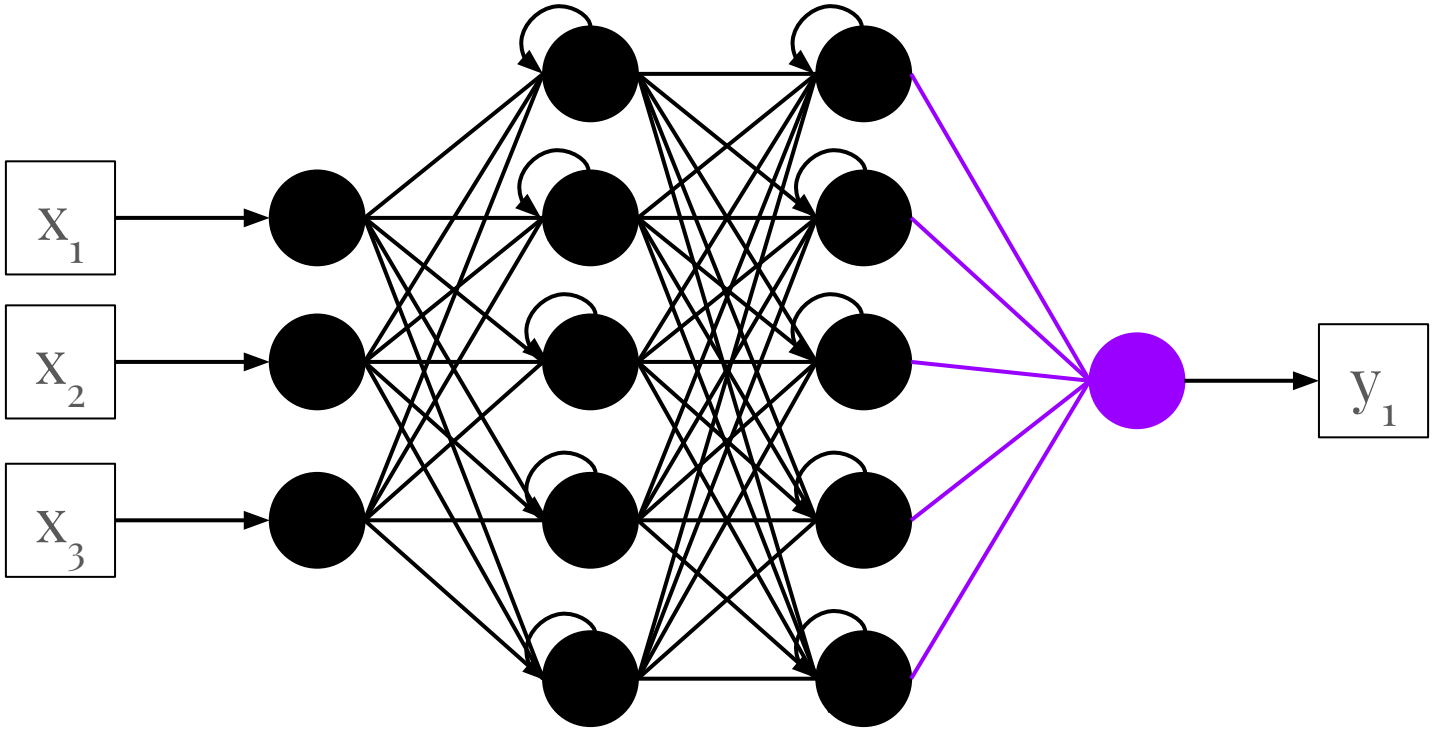
Simple Recurrent Neural Network



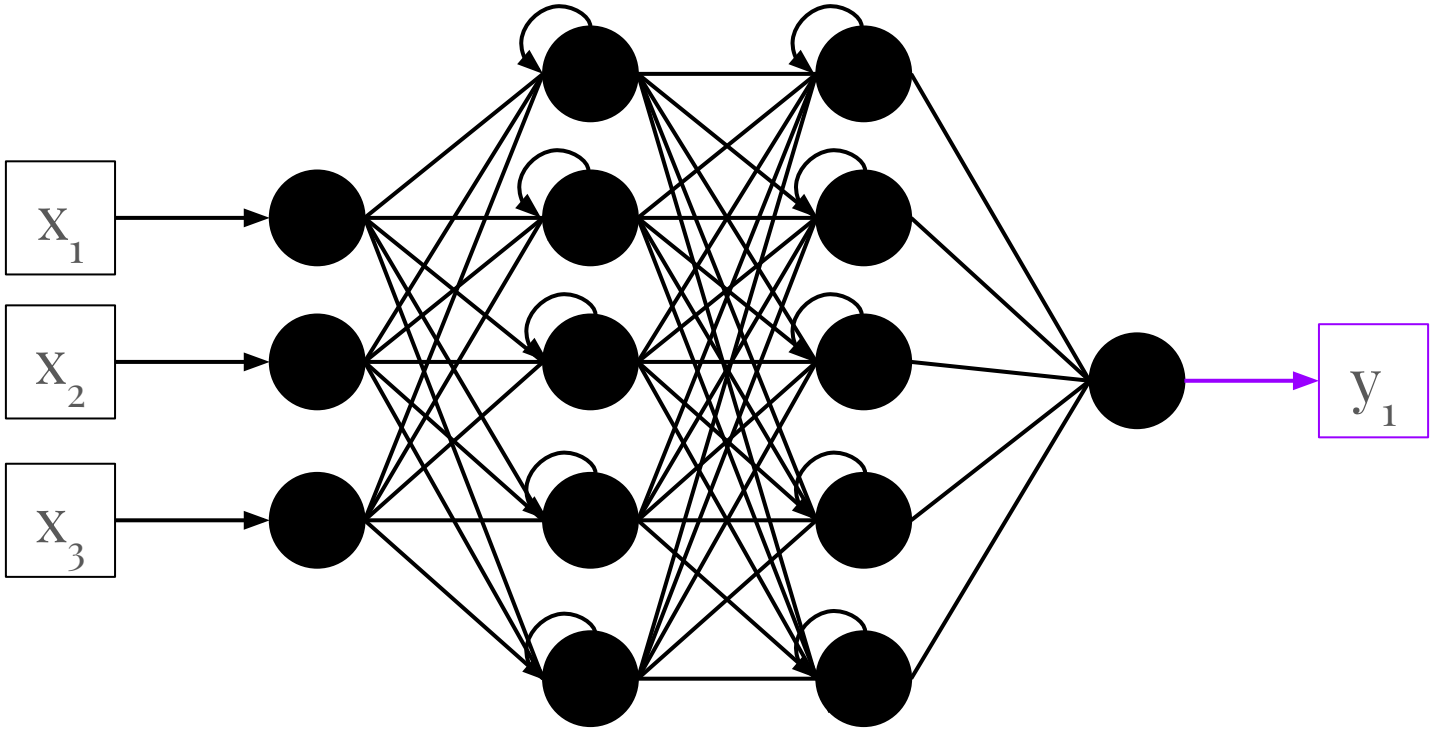
Simple Recurrent Neural Network



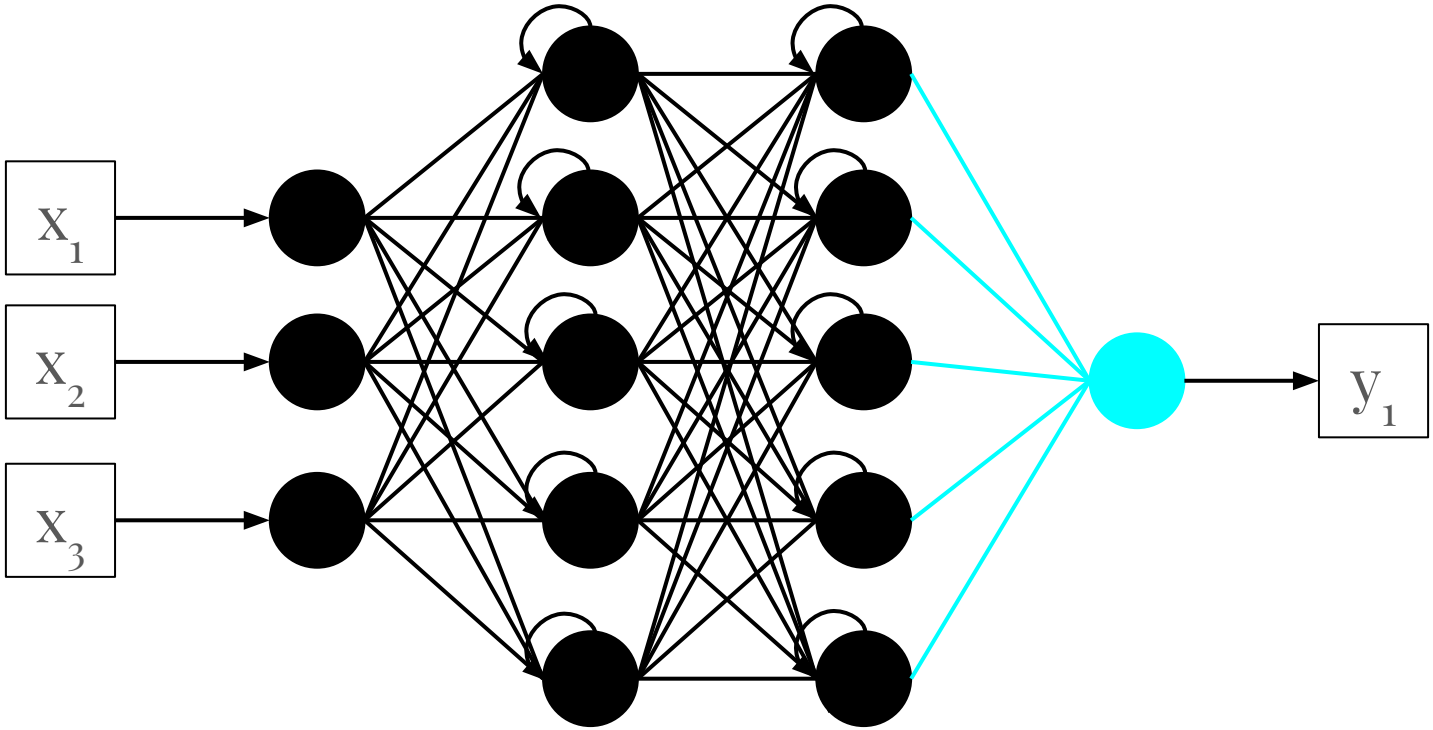
Simple Recurrent Neural Network



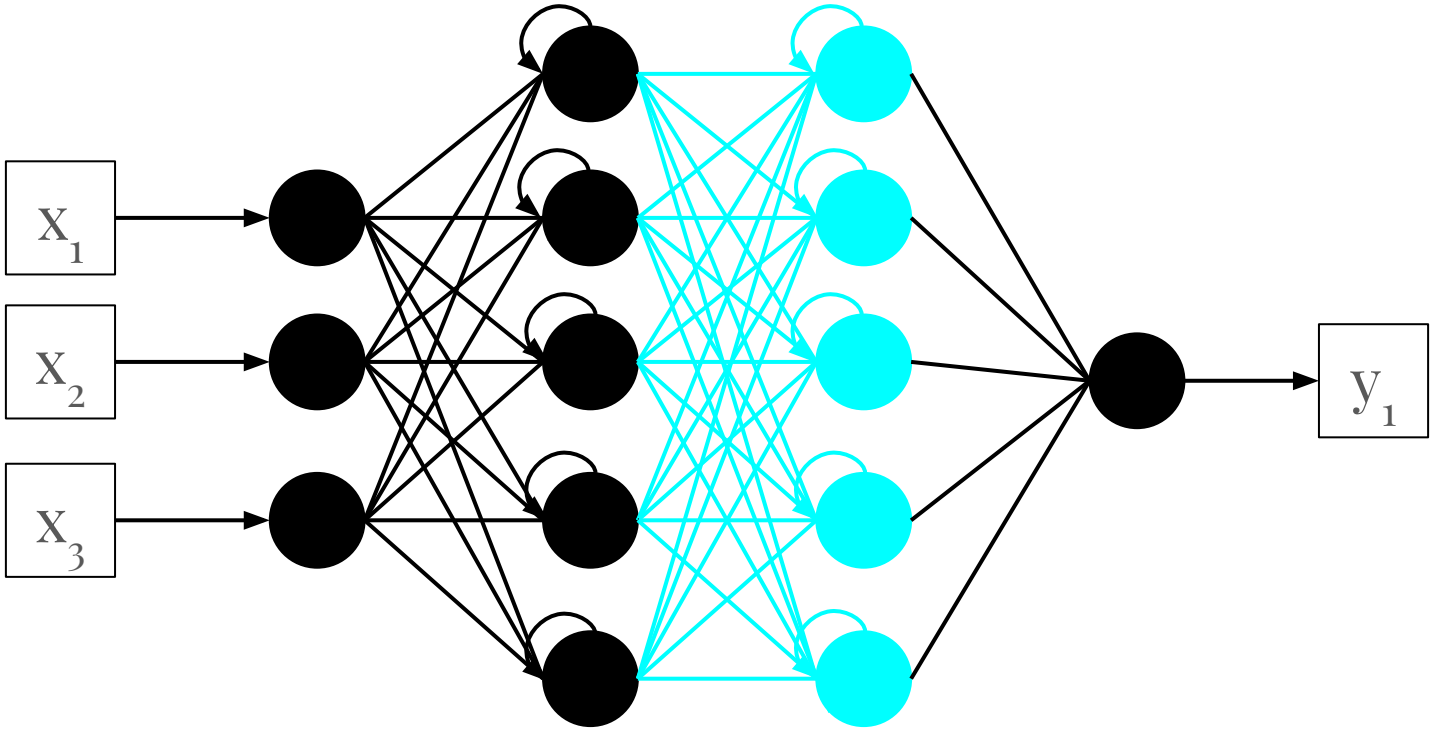
Simple Recurrent Neural Network



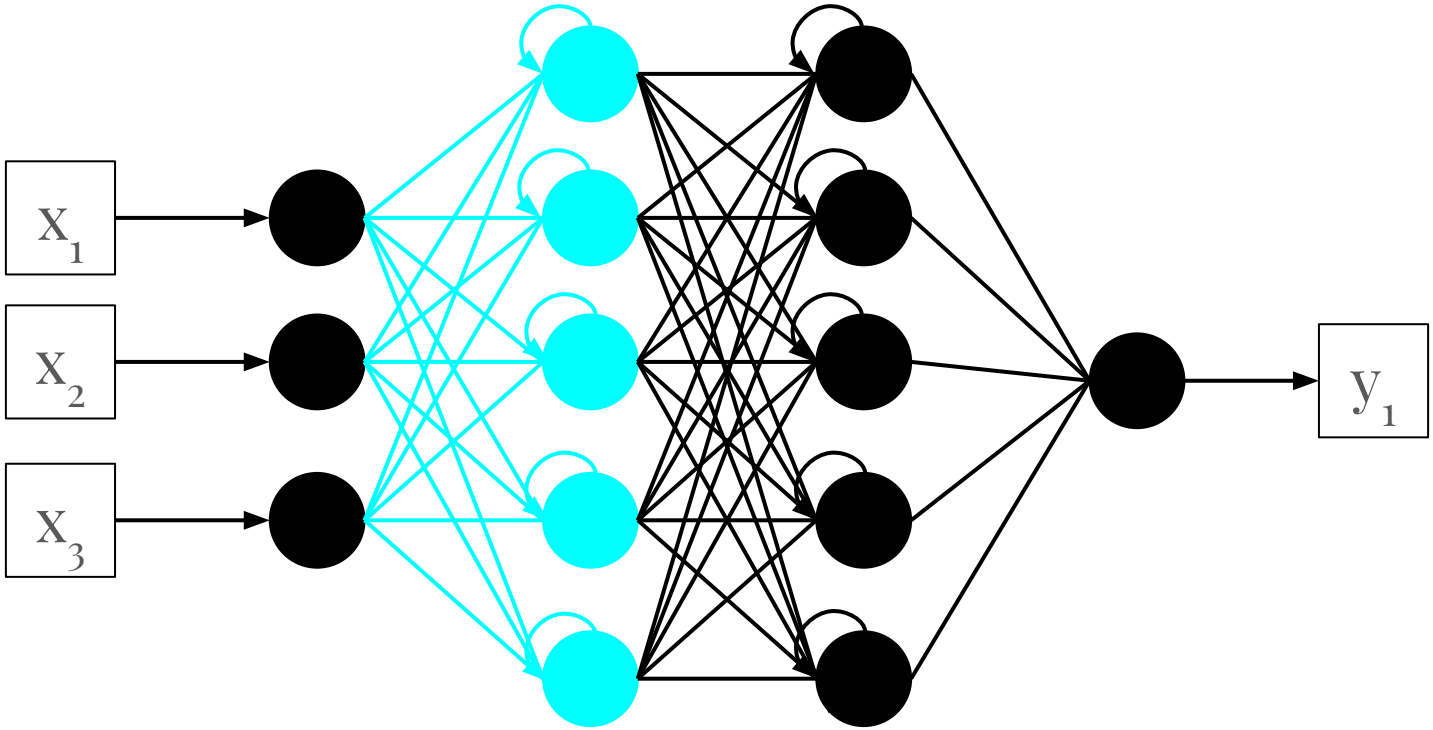
Simple Recurrent Neural Network



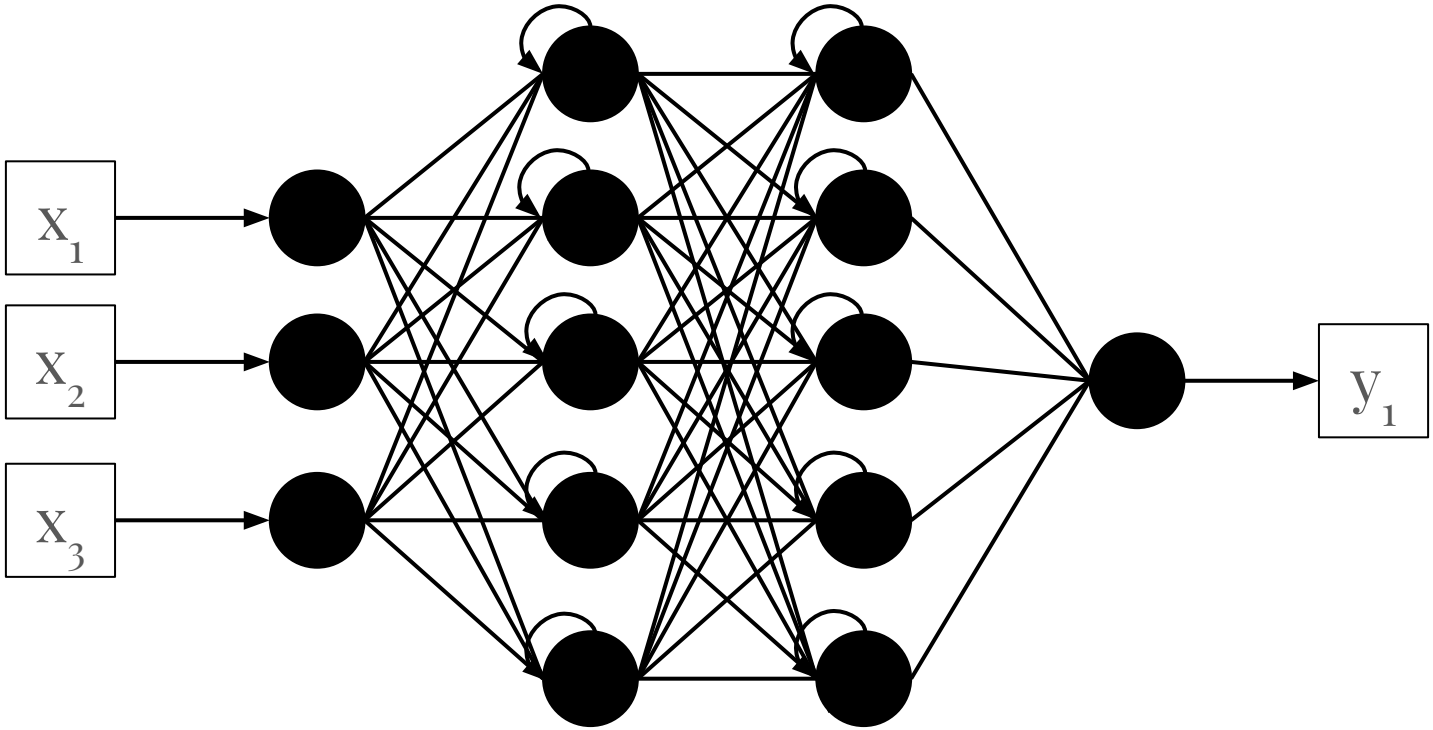
Simple Recurrent Neural Network



Simple Recurrent Neural Network

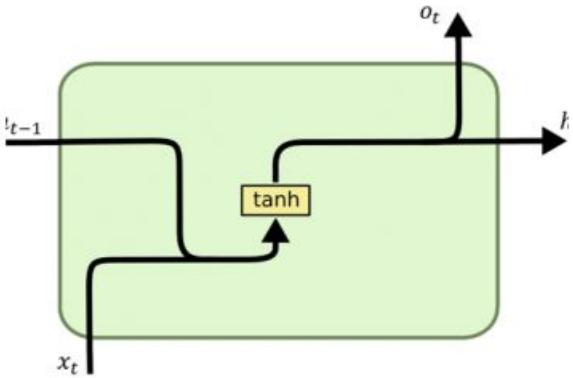


Simple Recurrent Neural Network

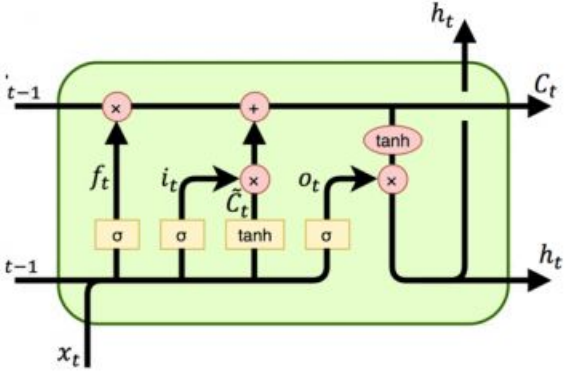


Simple Recurrent Neural Network

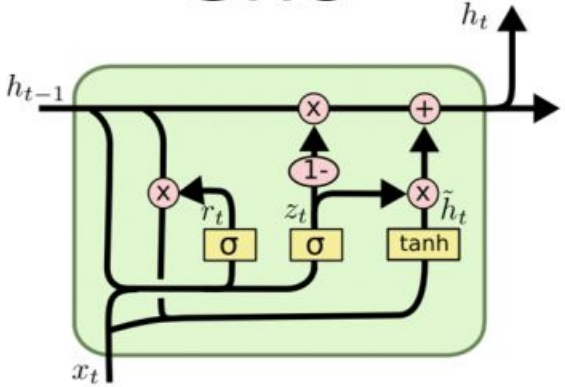
RNN



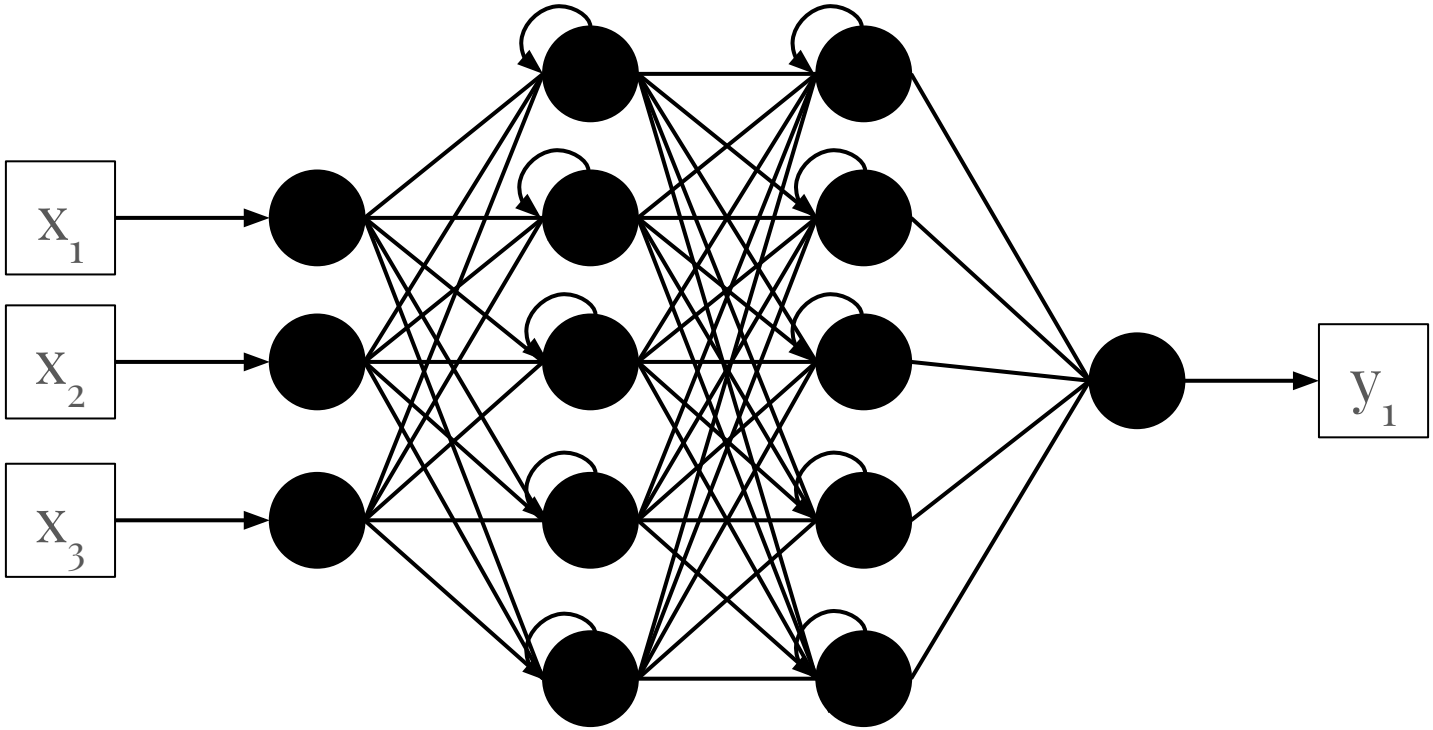
LSTM



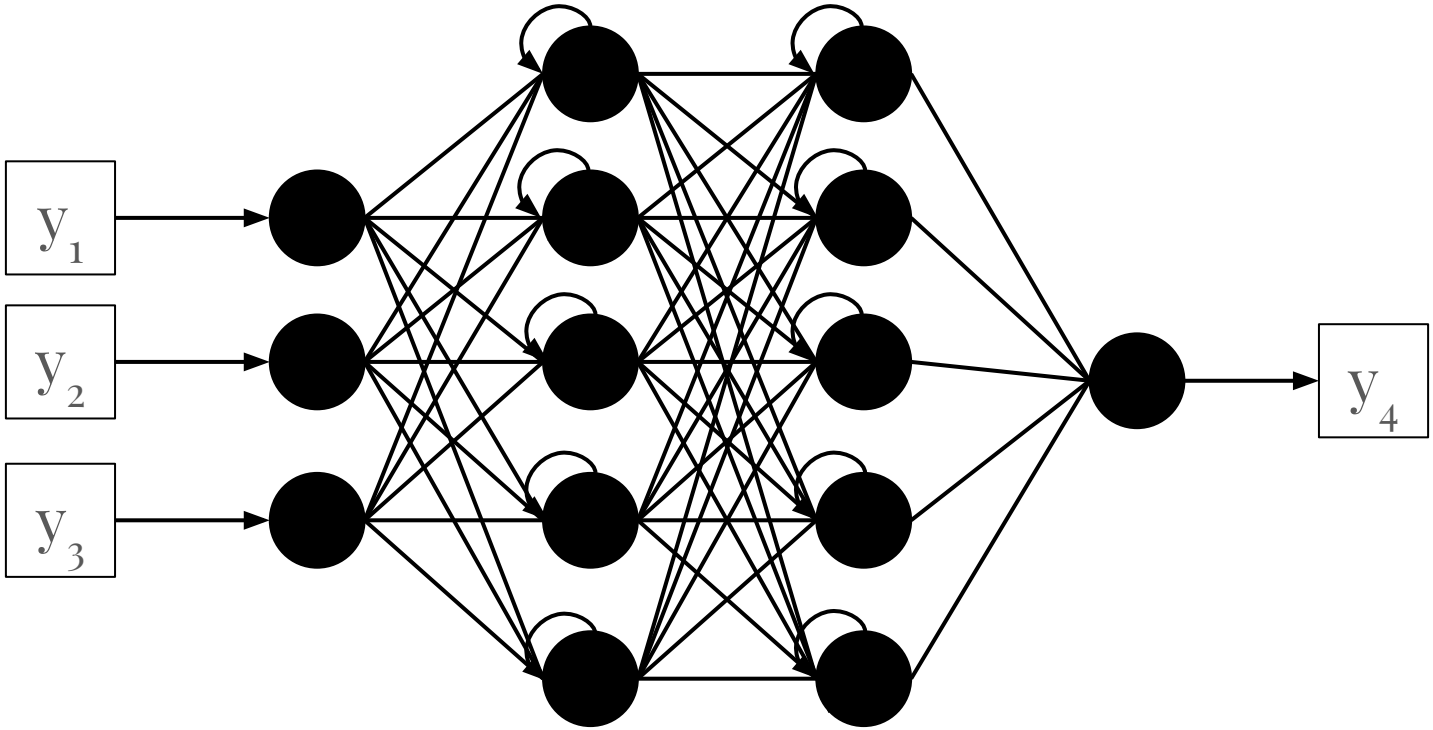
GRU



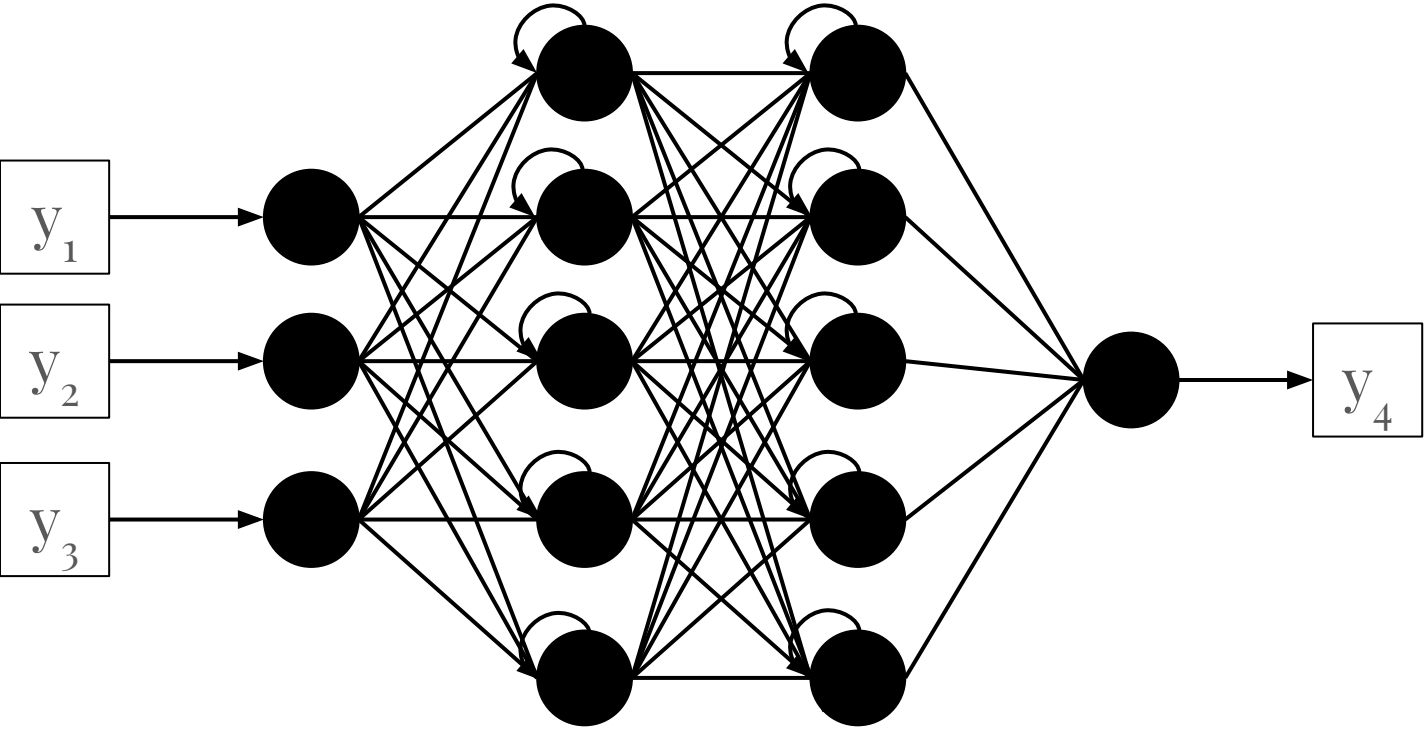
More Complicated Recurrent Neural Network Versions



Normal (Point-Wise Input)



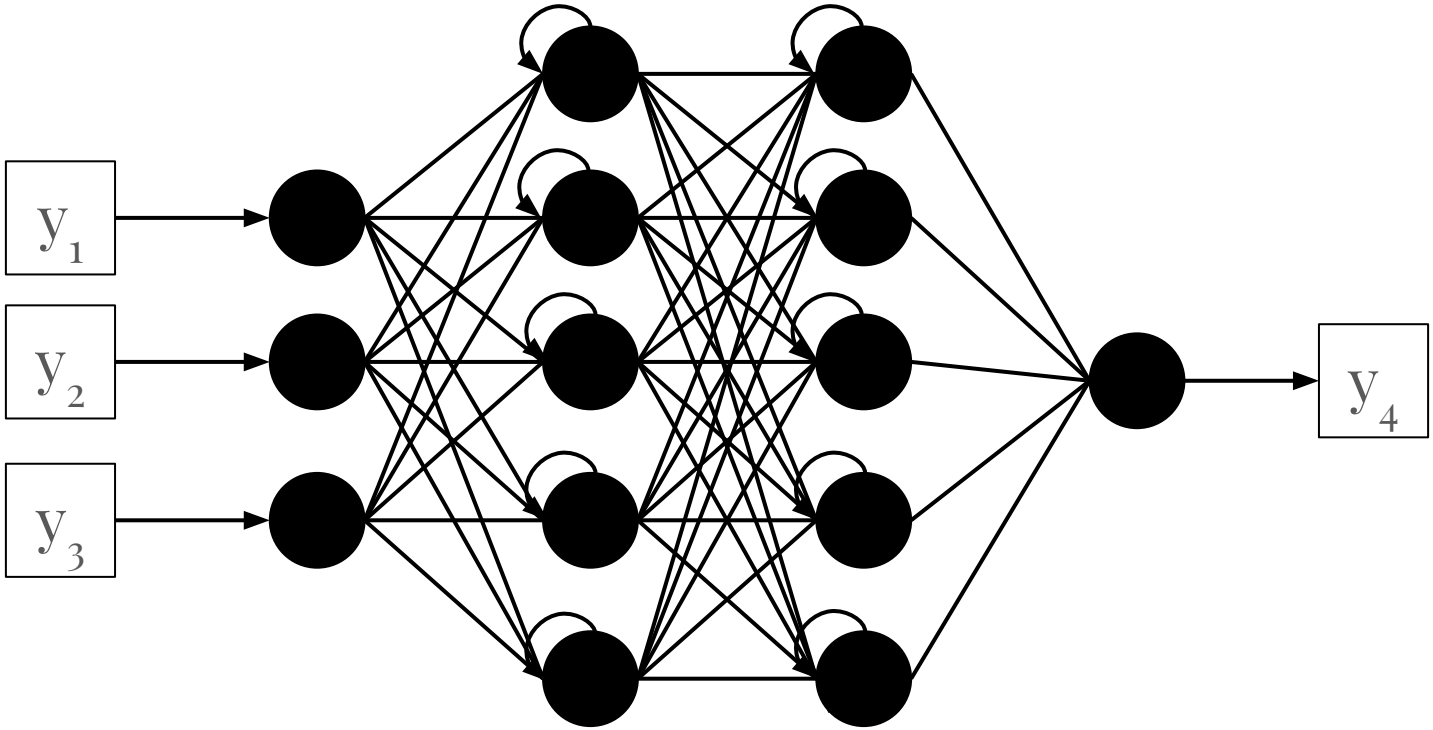
Time Series Formatting



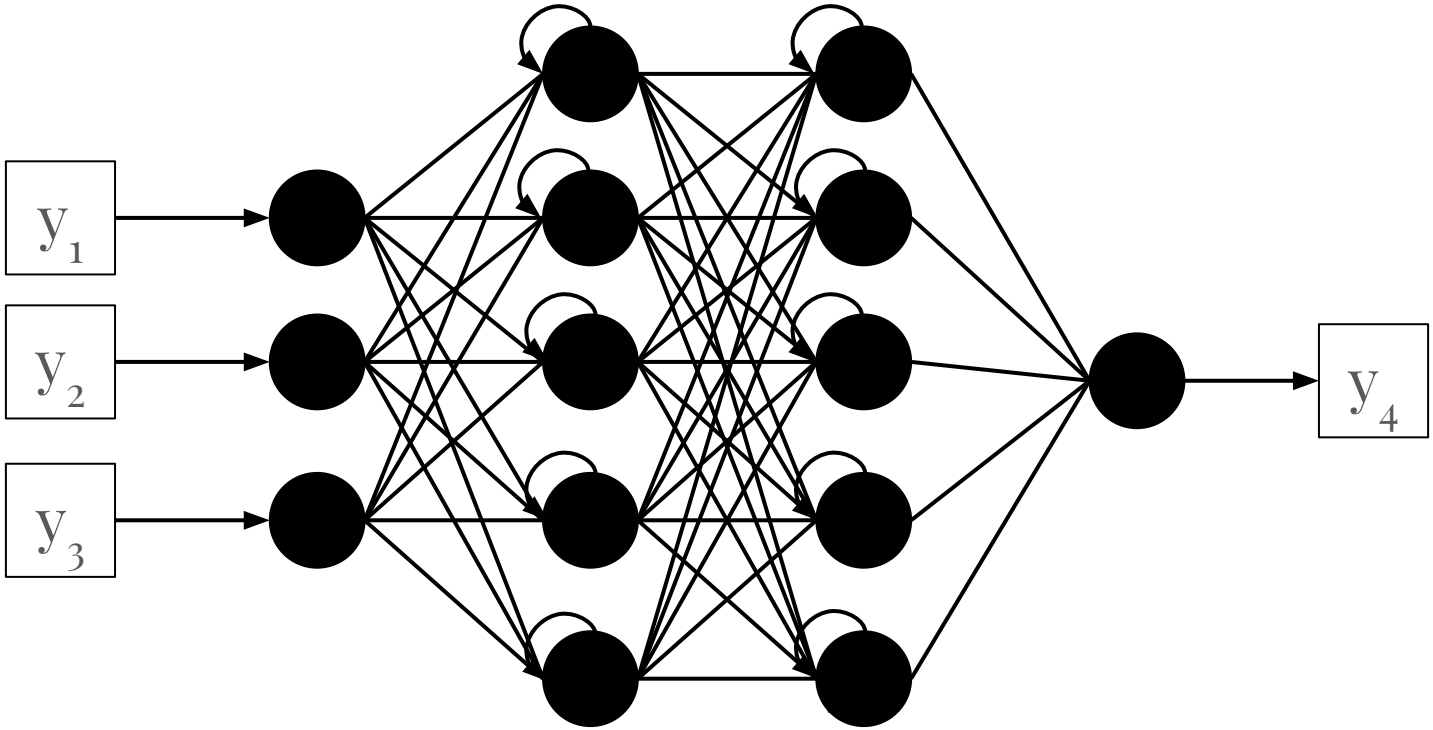
Text data can be considered time series (ordered) data

Time Series Formatting

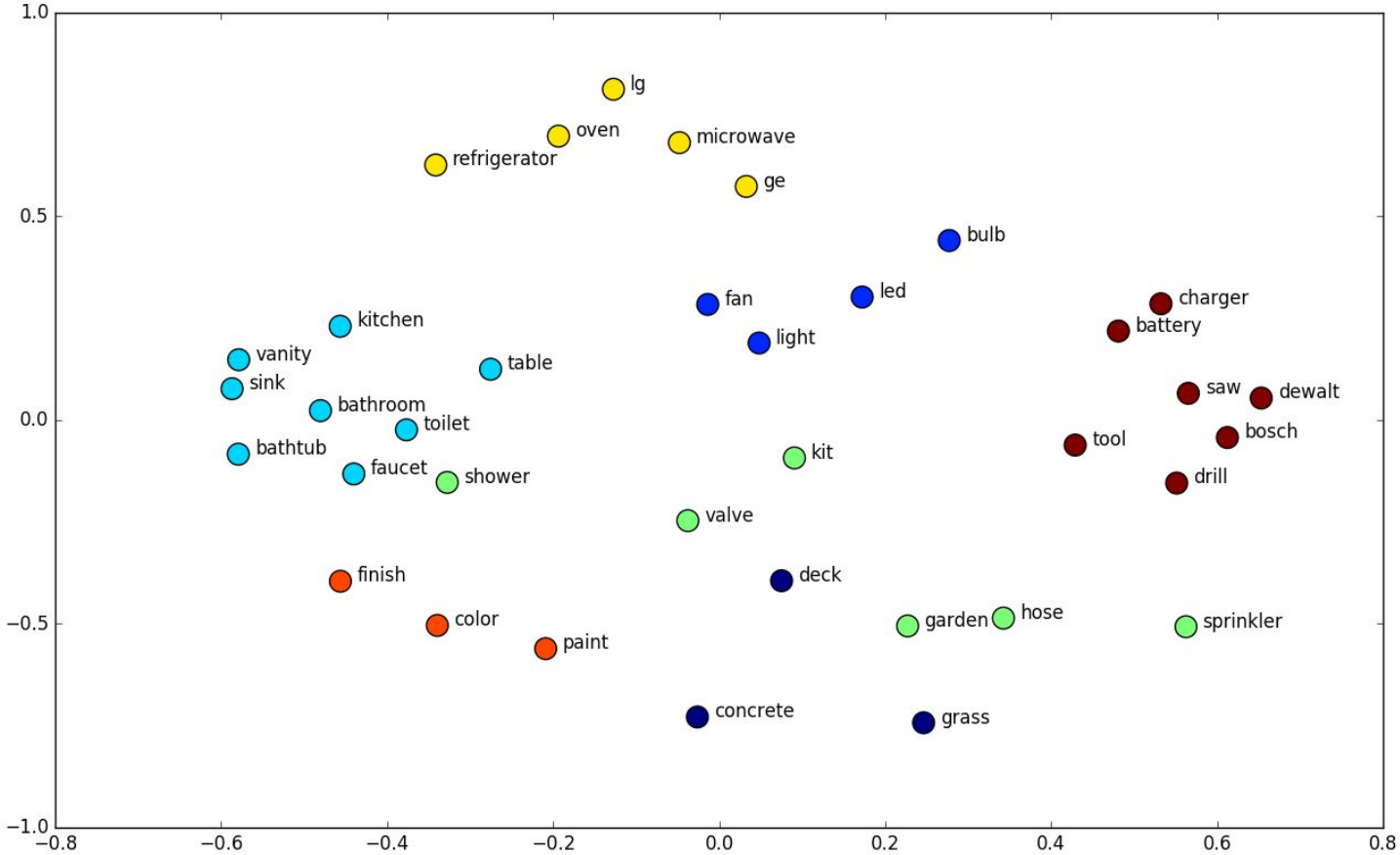
Working With Words



Numerical Data: (1,2,3) → 4



Word Data: (Mr., Worldwide, to) → Infinity



Word2Vec

Word Embedding

Predicting the Next Word In a Song

I don't want a lot for Christmas

There is just one thing I need

I don't care about the presents underneath the Christmas tree

I just want you for my own

More than you could ever know

Make my wish come true

All I want for Christmas is you

I don't want a lot for Christmas

There is just one thing I need

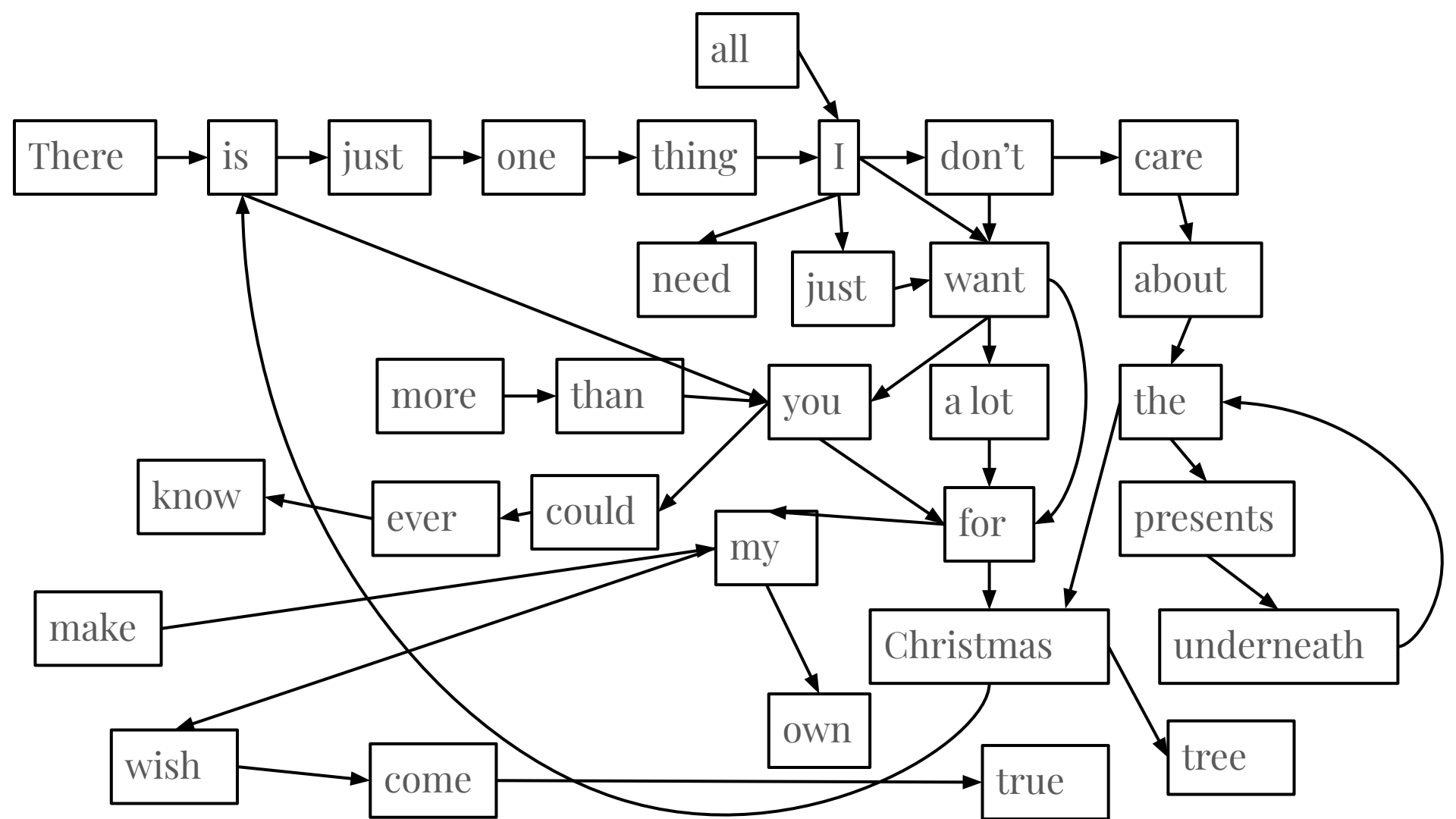
I don't care about the presents underneath the Christmas tree

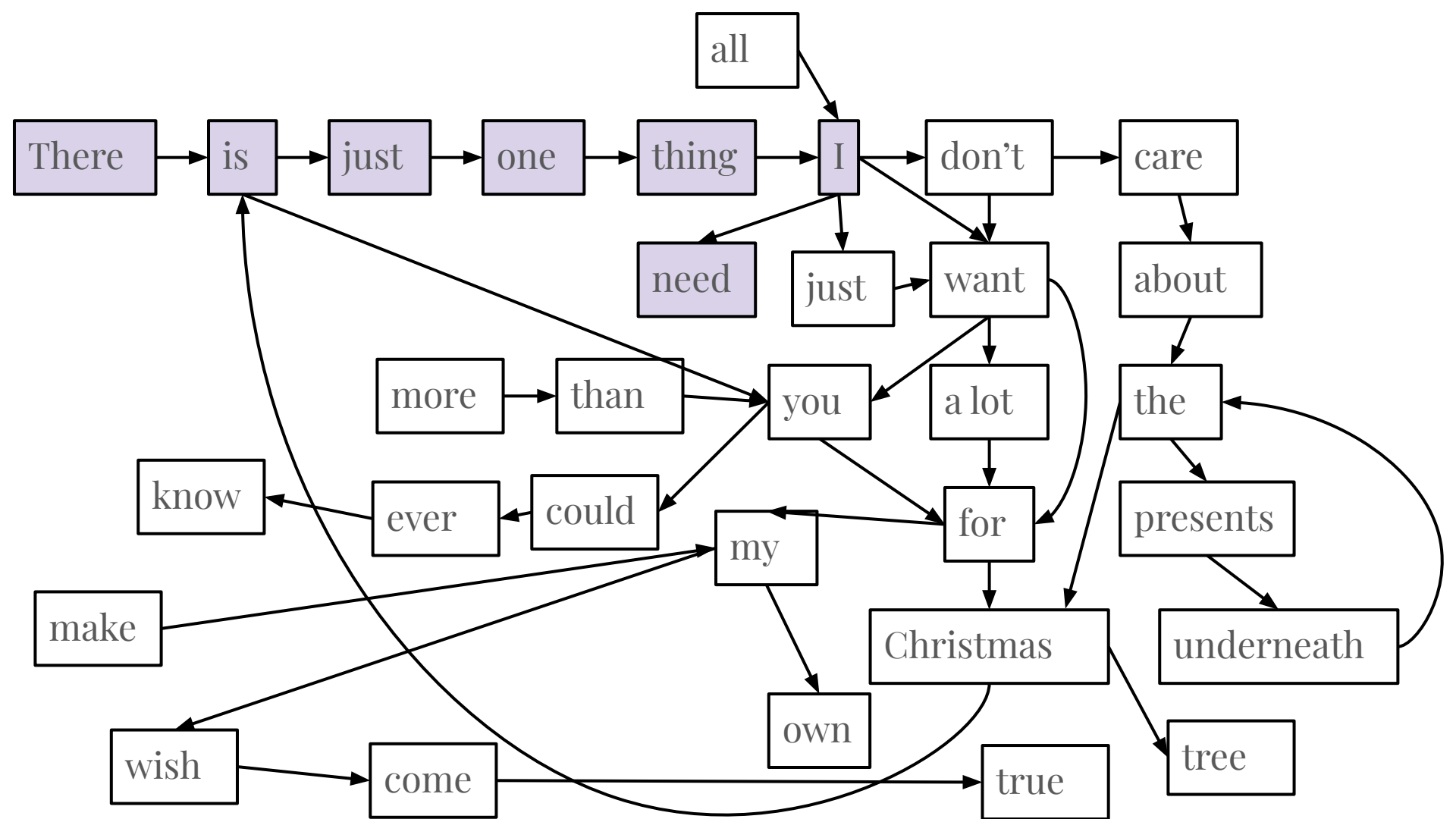
I just want you for my own

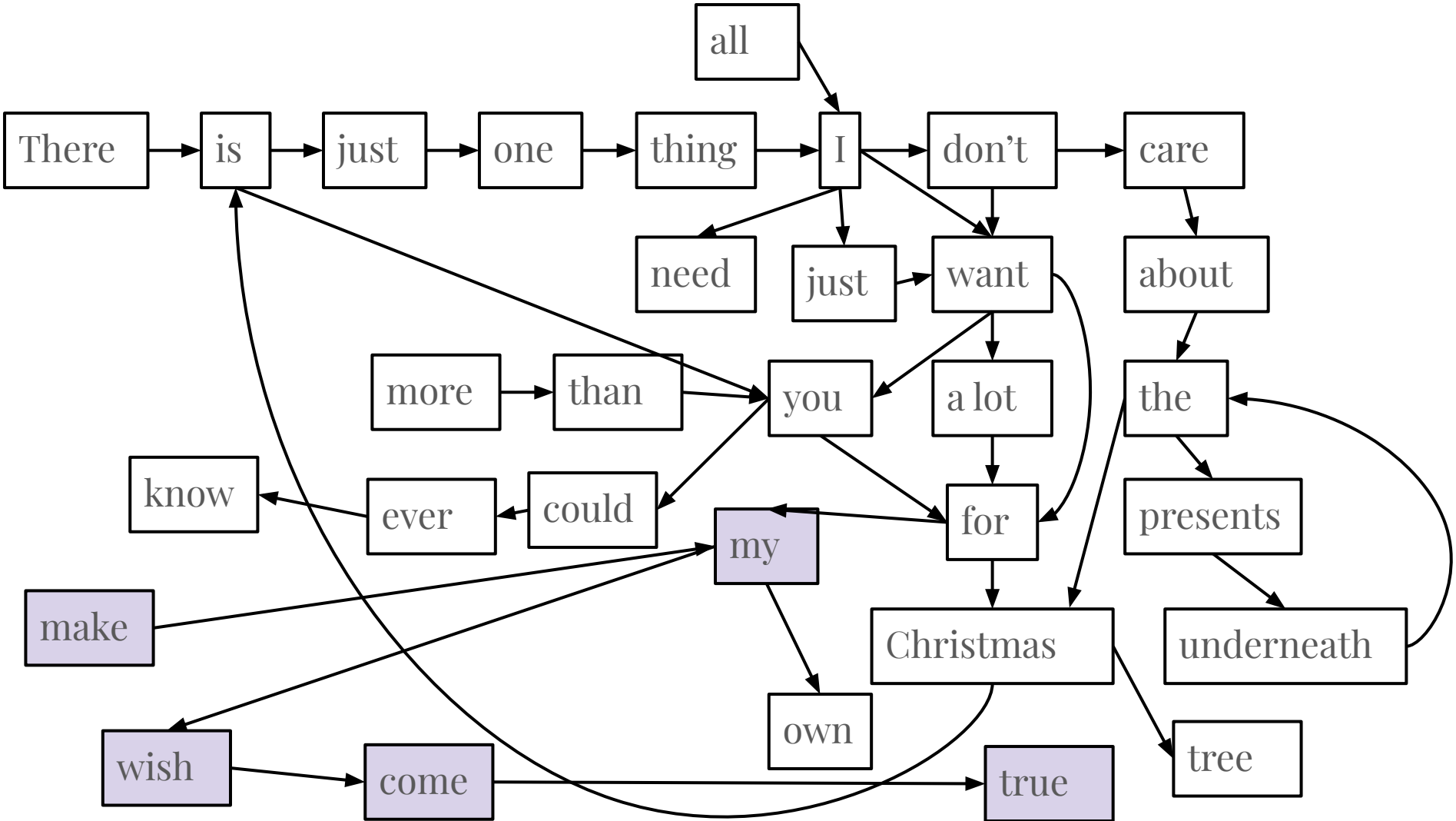
More than you could ever know

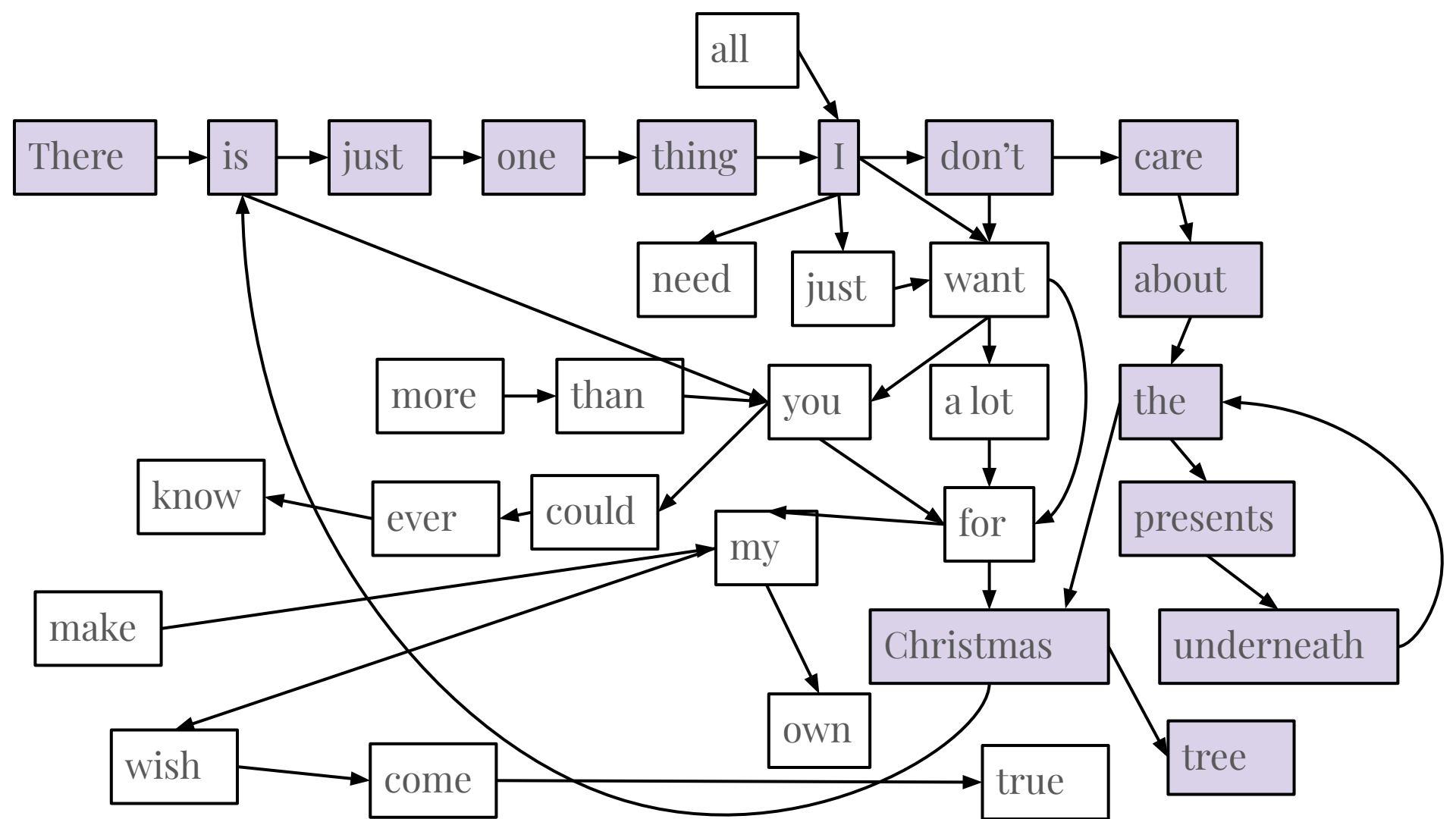
Make my wish come true

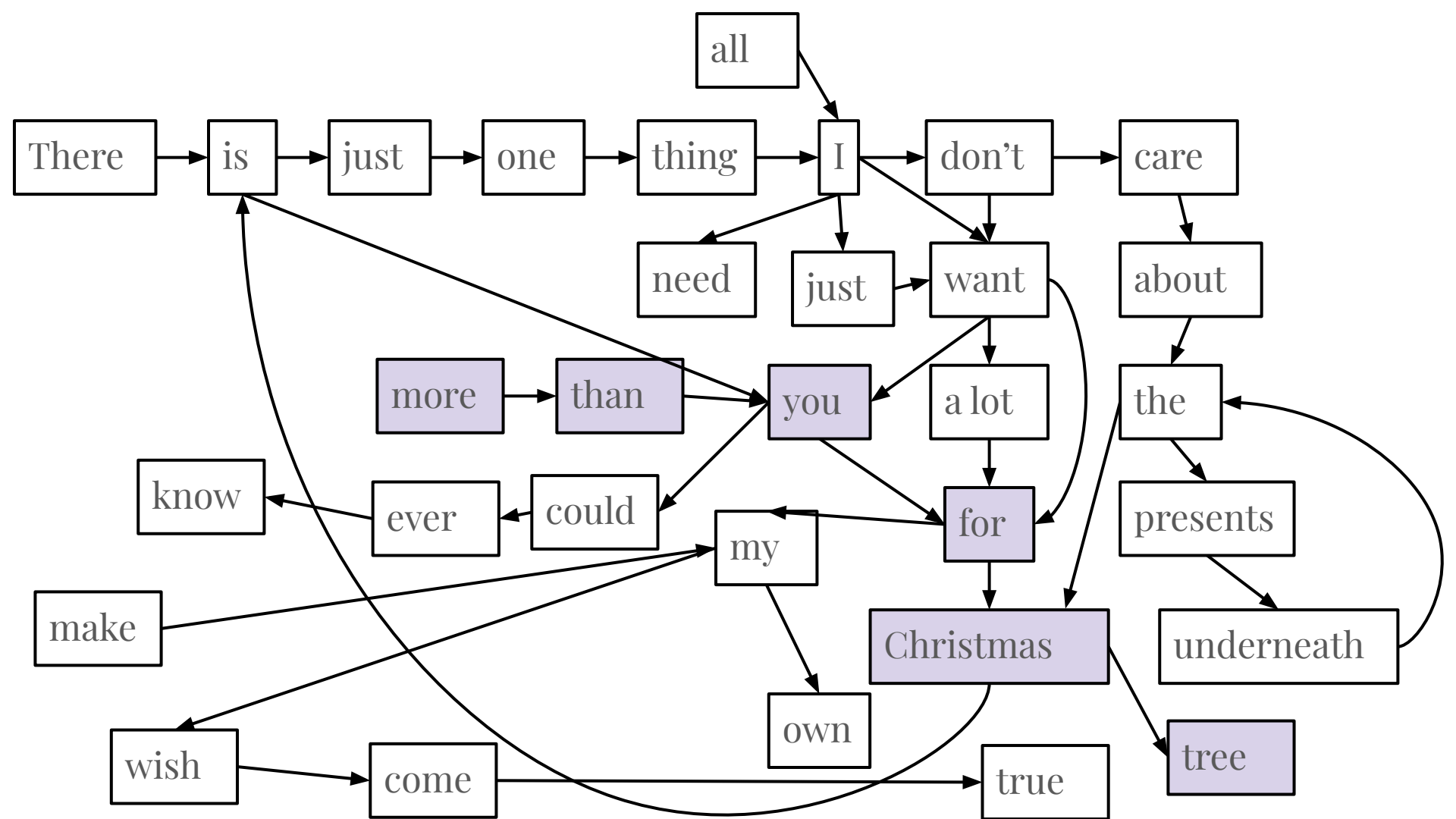
All I want for Christmas is you


all
There
is
just
one
thing
I
don’t
care
need
just
want
about
more
than
you
a lot
the
know
ever
could
for
presents
my
make
Christmas
underneath
own
wish
come
true
tree


all
There
is
just
one
thing
I
don’t
care
need
just
want
about
more
than
you
a lot
the
know
ever
could
for
presents
my
make
Christmas
underneath
own
wish
come
true
tree


all
There
is
just
one
thing
I
don’t
care
need
just
want
about
more
than
you
a lot
the
know
ever
could
for
presents
my
make
Christmas
underneath
own
wish
come
true
tree


all
There
is
just
one
thing
I
don’t
care
need
just
want
about
more
than
you
a lot
the
know
ever
could
for
presents
my
make
Christmas
underneath
own
wish
come
true
tree


all
There
is
just
one
thing
I
don’t
care
need
just
want
about
more
than
you
a lot
the
know
ever
could
for
presents
my
make
Christmas
underneath
own
wish
come
true
tree

I don't want a lot for Christmas

There is just one thing I need

I don't care about the presents underneath the Christmas tree

I just want you for my own

More than you could ever know

Make my wish come true

All I want for Christmas is you

I don't want a lot for Christmas

There is just one thing I need

I don't care about the presents underneath the Christmas tree

I just want you for my own

More than you could ever know

Make my wish come true

All I want for Christmas is you

I don't want a lot for Christmas

There is just one thing I need

I don't care about the presents underneath the Christmas tree

I just want you for my own

More than you could ever know

Make my wish come true

All I want for Christmas is you

I don't want a lot for Christmas

There is just one thing I need

I don't care about the presents underneath the Christmas tree

I just want you for my own

More than you could ever know

Make my wish come true

All I want for Christmas is you

Attention Networks

Attention Networks

- To properly interpret and produce intelligible sentences, you need to consider variable sequence lengths
  - Not something a normal neural network is capable of doing
- Solution: Attention networks!

Attention Is All You Need

Ashish Vaswani* Google Brain avaswani@google.com

Noam Shazeer* Google Brain noam@google.com

Niki Parmar* Google Research nikip@google.com

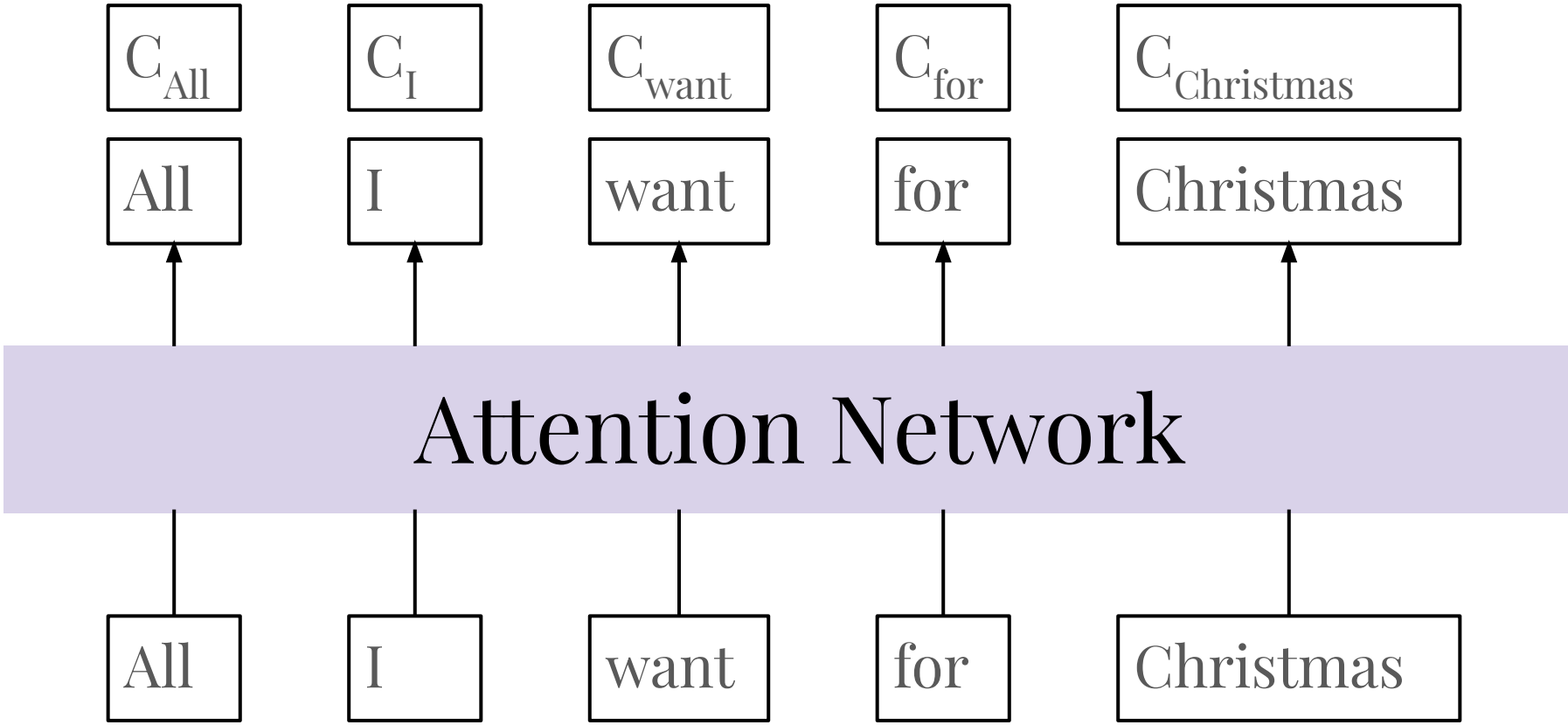
Jakob Uszkoreit* Google Research usz@google.com

Llion Jones* Google Research llion@google.com

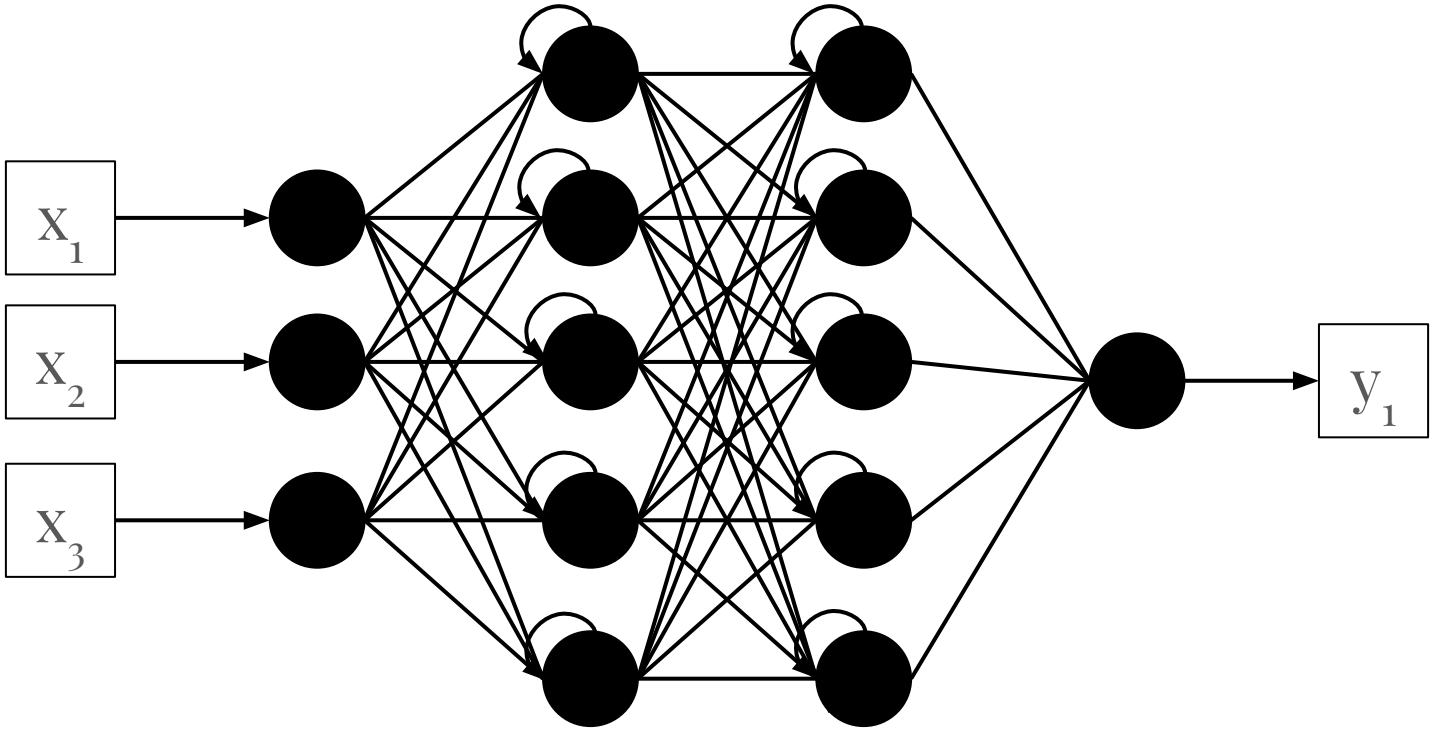
Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu

Łukasz Kaiser* Google Brain lukaszkaiser@google.com

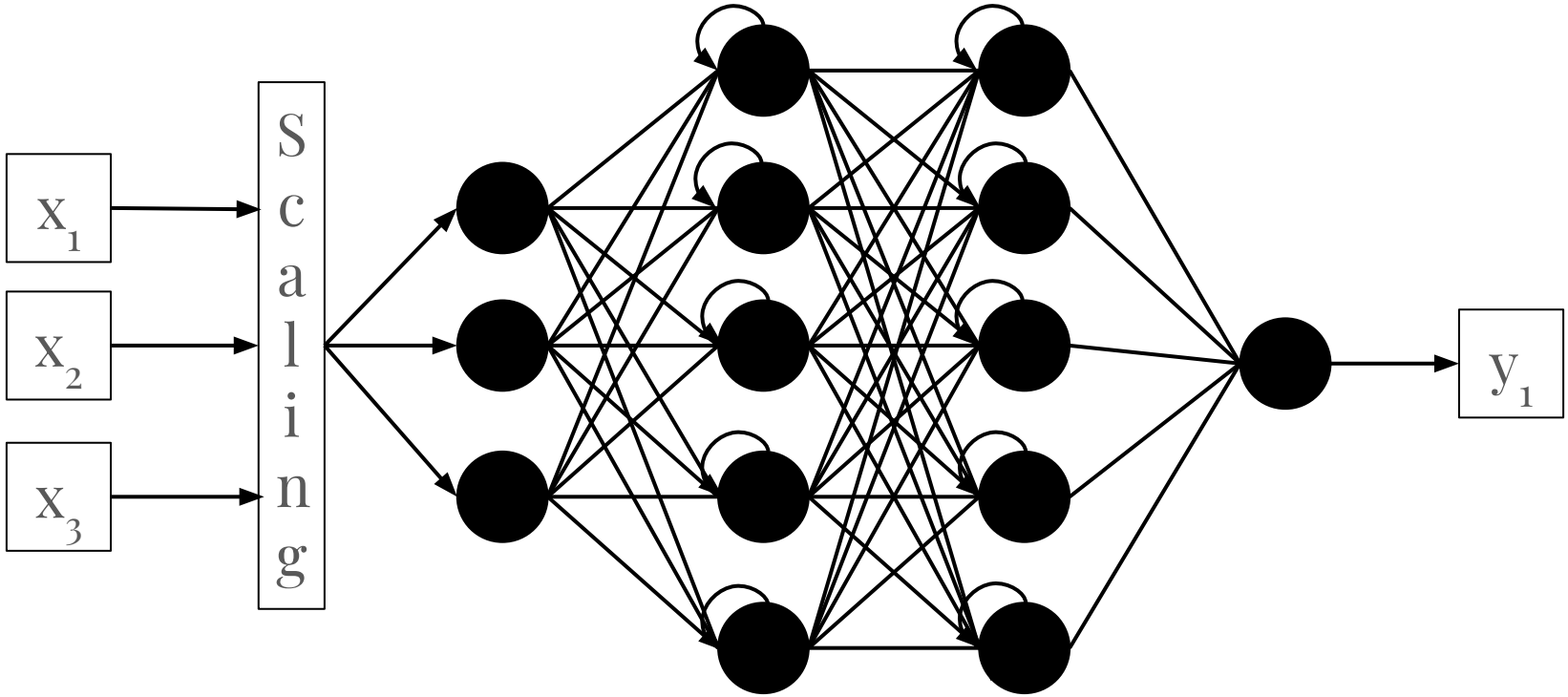
Illia Polosukhin* ‡ illia.polosukhin@gmail.com


C_{All}
C_{I}
C_{want}
C_{for}
$C_{Christmas}$
All
I
want
for
Christmas
Attention Network
All
I
want
for
Christmas

Transformer Networks

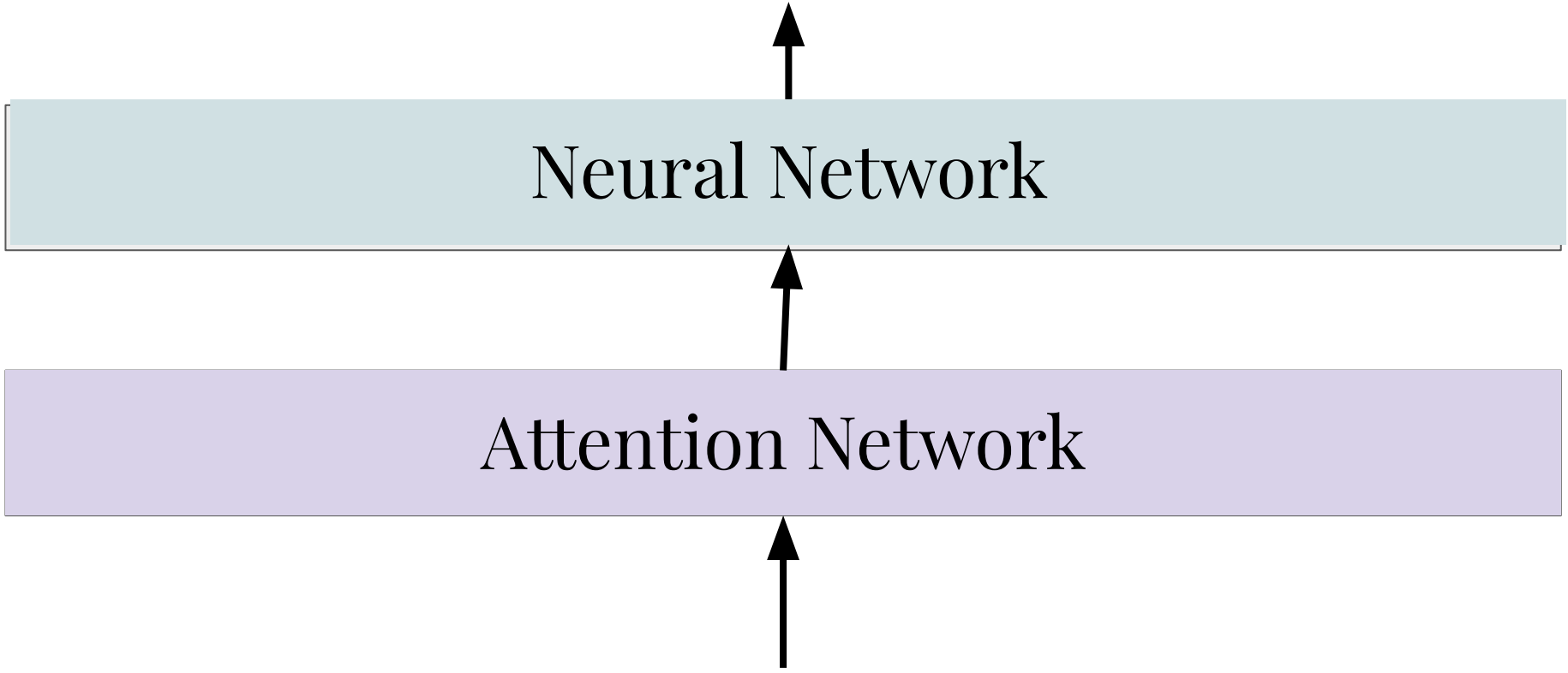


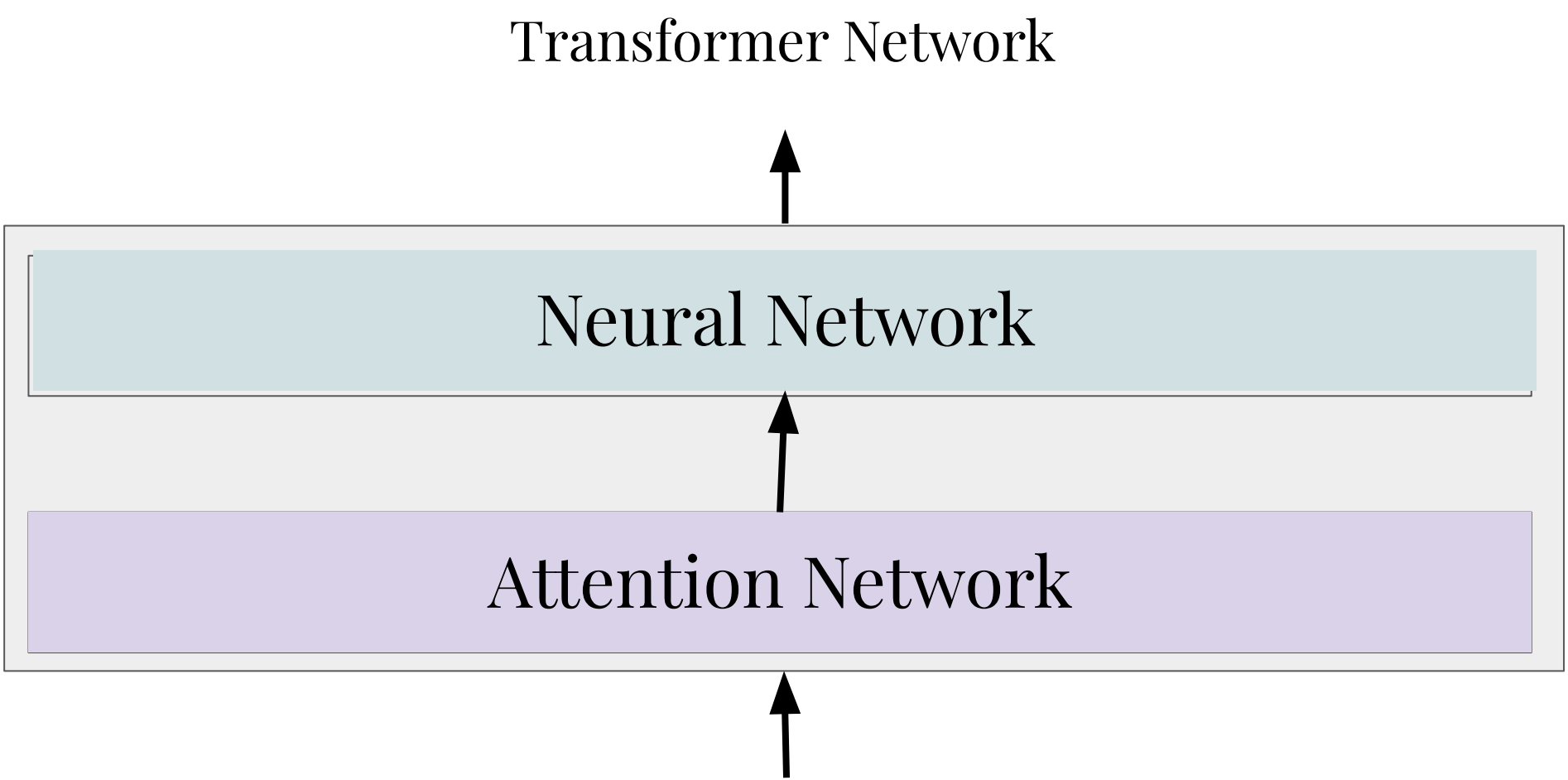
Simple Recurrent Neural Network



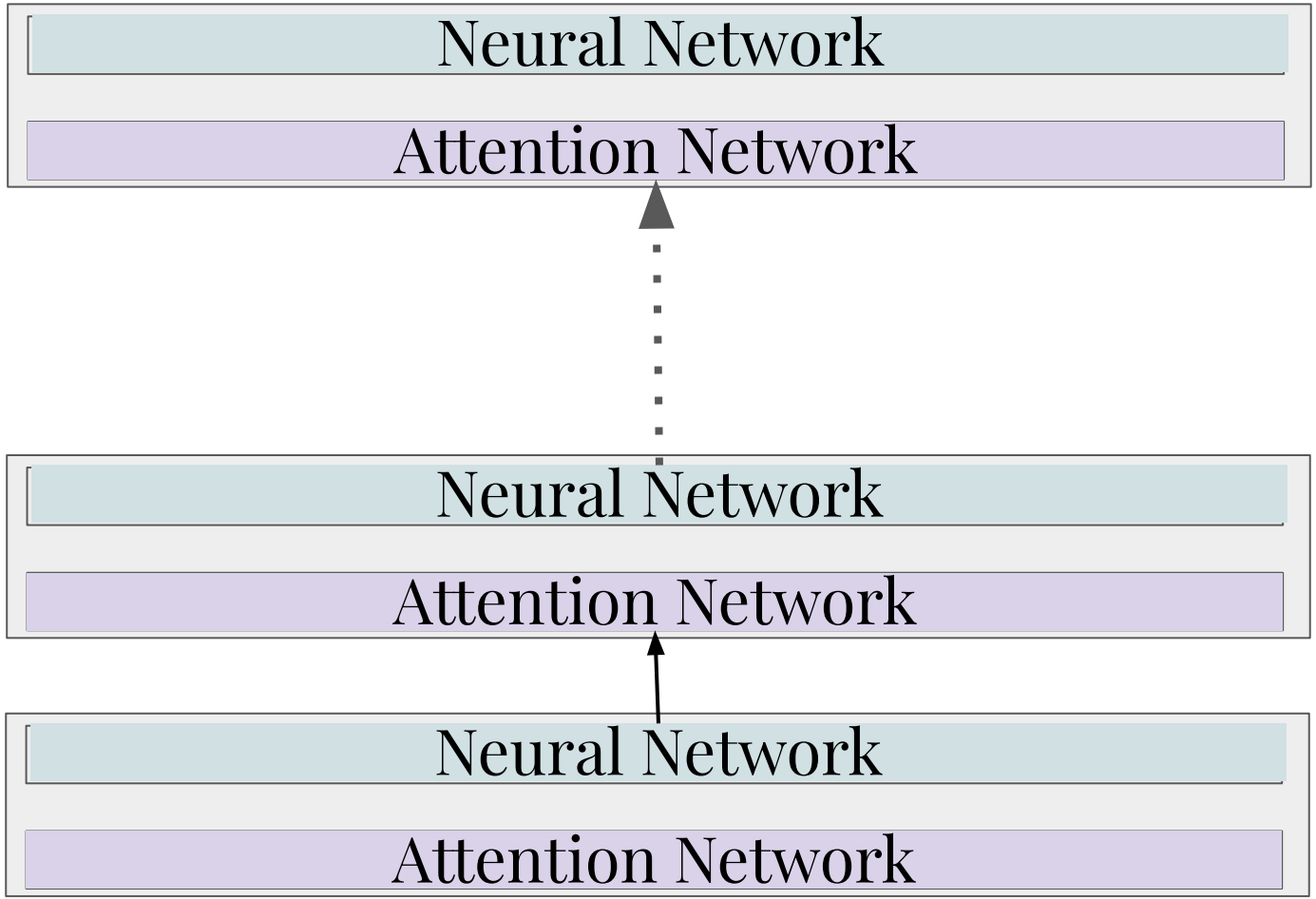
Preprocessing of data has been used throughout the course

Transformer Network: Attention network creates the connections, neural network finds the patterns

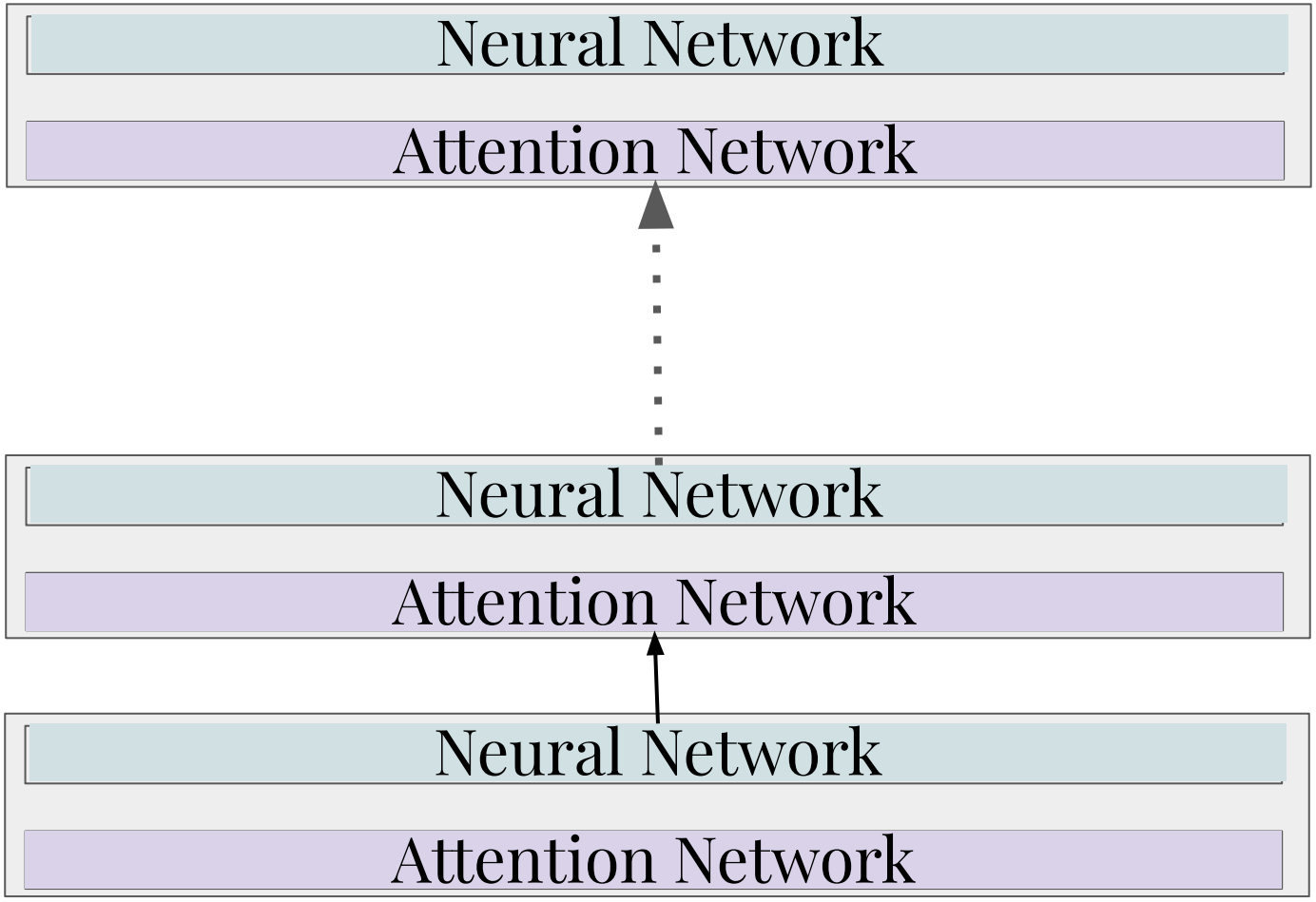



Transformer Network
Neural Network
Attention Network

Chat Bots


Neural Network
Attention Network
Neural Network
Attention Network
Neural Network
Attention Network

Semantics



Syntax

ChatGTP-3

- 96 Transformers
  - Hundreds of billions of parameters
- Trained with the entire internet and all openly available books
  - Trillions of training points
- Trains in 355 years on a single GPU
  - Trains in a month with thousands of GPUs