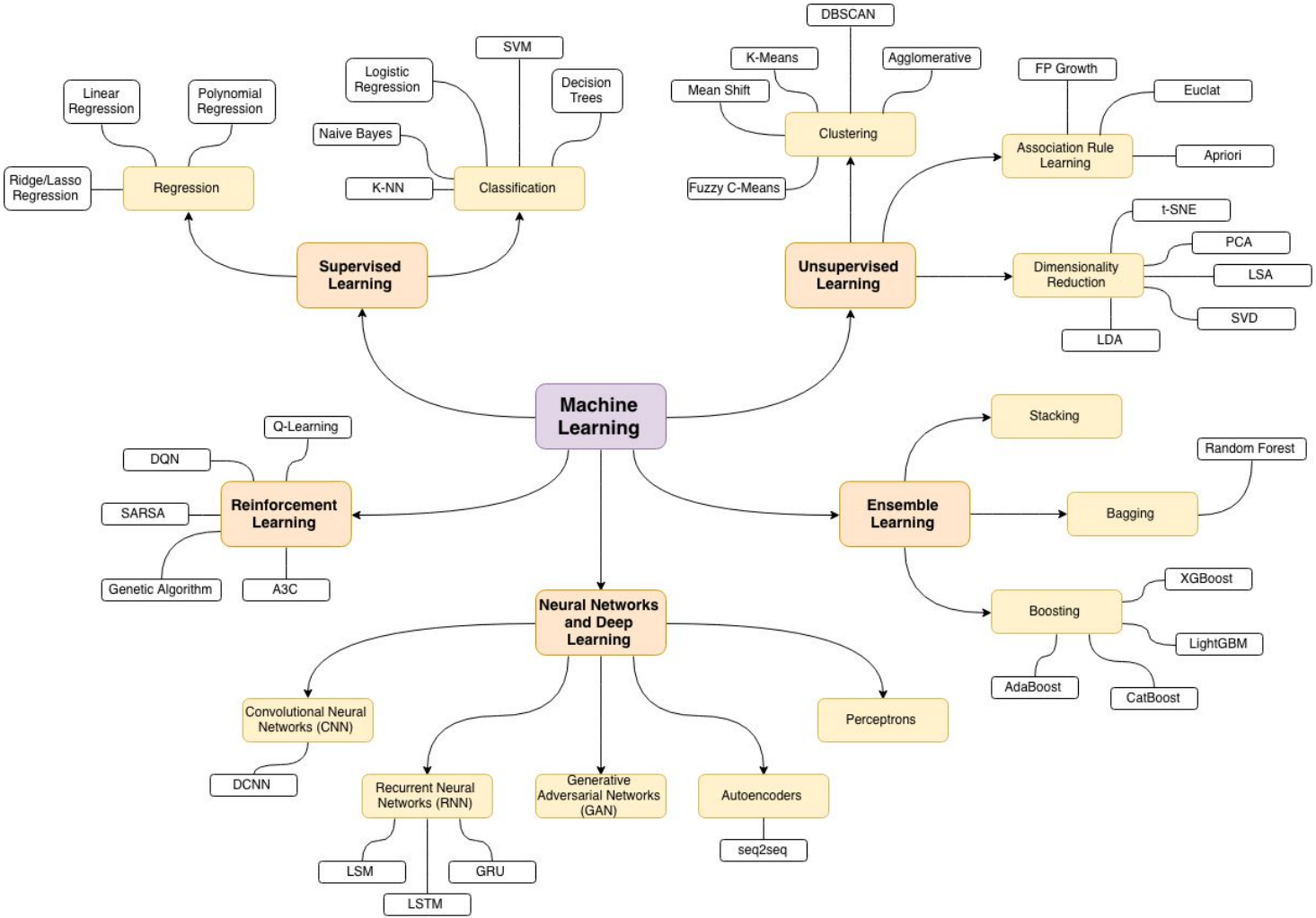
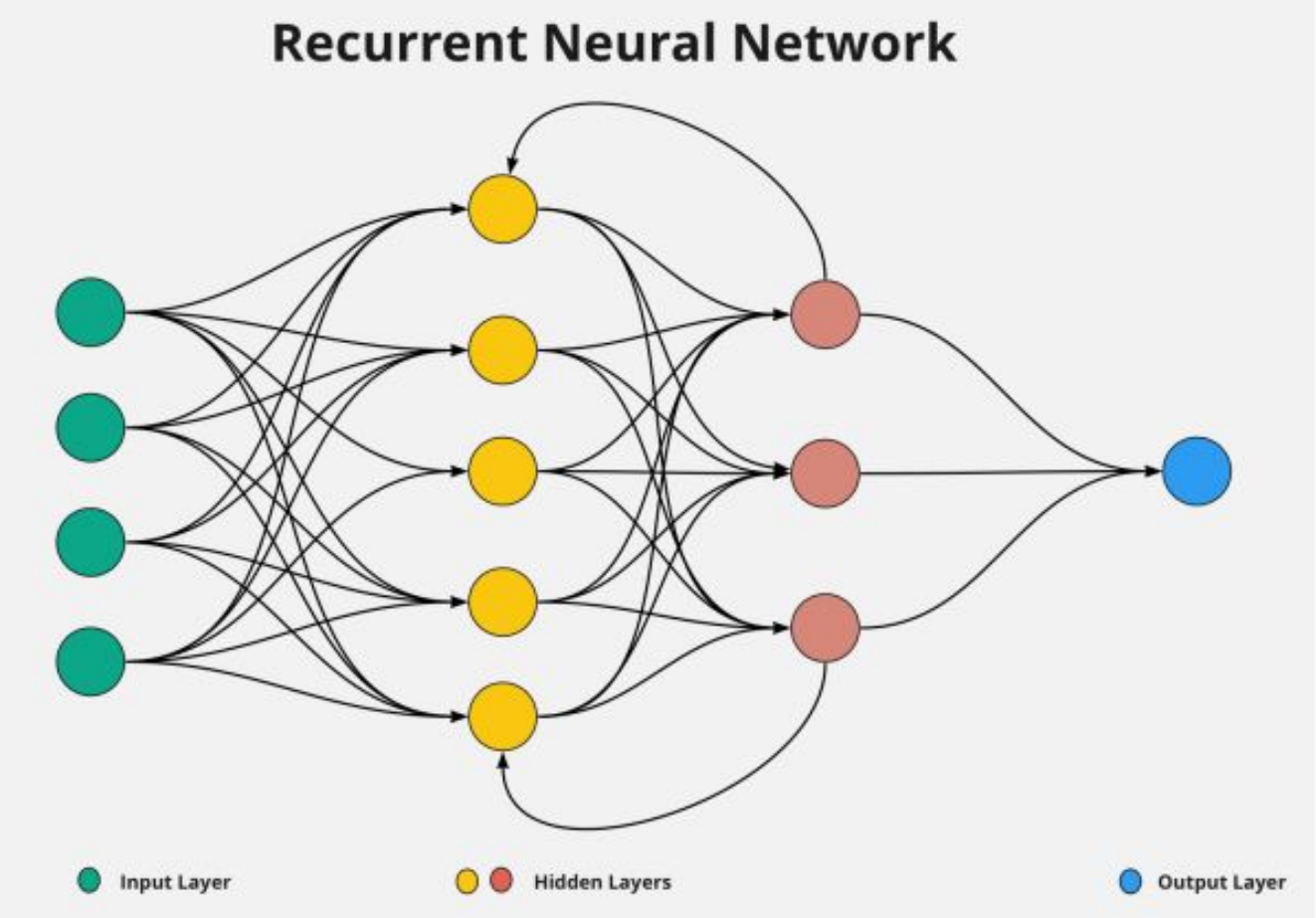
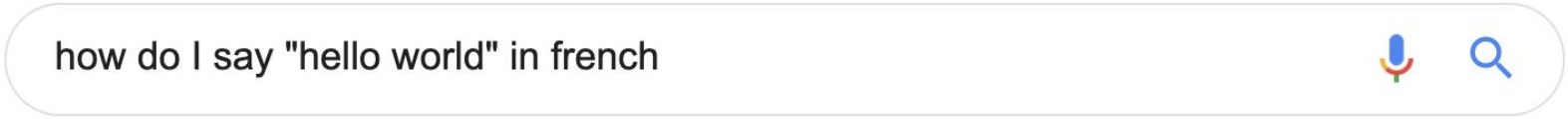
# MIDS W207
# Applied Machine Learning

Week 13
Live Session Slides


Machine Learning
Supervised Learning
Regression
Linear Regression
Polynomial Regression
Ridge/Lasso Regression
Classification
Logistic Regression
SVM
Decision Trees
Naive Bayes
K-NN
Unsupervised Learning
Clustering
DBSCAN
K-Means
Agglomerative
Mean Shift
Fuzzy C-Means
Association Rule Learning
FP Growth
Euclat
Apriori
Dimensionality Reduction
t-SNE
PCA
LSA
SVD
LDA
Reinforcement Learning
Q-Learning
DQN
SARSA
Genetic Algorithm
A3C
Neural Networks and Deep Learning
Convolutional Neural Networks (CNN)
DCNN
Recurrent Neural Networks (RNN)
LSM
GRU
LSTM
Generative Adversarial Networks (GAN)
Autoencoders
seq2seq
Perceptrons
Ensemble Learning
Stacking
Bagging
Random Forest
Boosting
XGBoost
LightGBM
AdaBoost
CatBoost

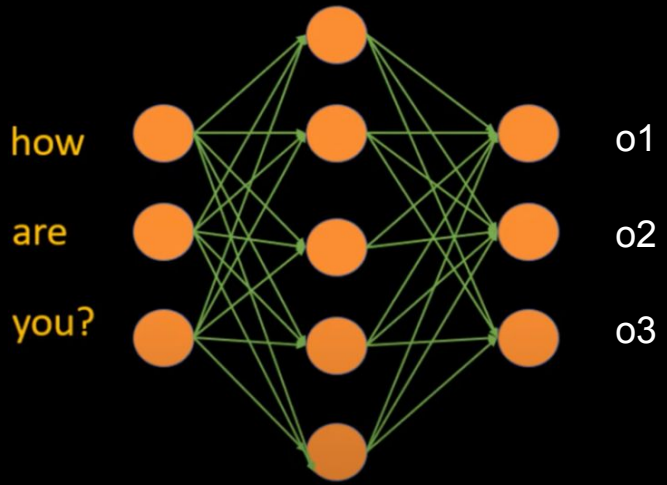

Recurrent Neural Network
Input Layer
Hidden Layers
Output Layer

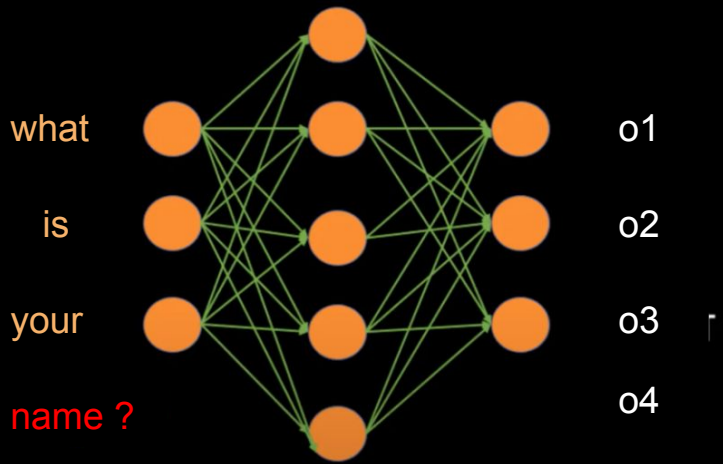
| x | | y |
| --- | --- | --- |
| not interested at | → | this time |
| will let you know if it changes | → | in the future |


how do I say "hello world" in french
All
Images
Shopping
Videos
News
More
Settings
Tools
About 2,660,000 results (0.71 seconds)
English - detected
French
hello world
Bonjour le monde
Open in Google Translate
Feedback


x
Rudolph Smith bought 1000 shares of tesla Inc. in March 2020
Person
Company
time
y
Rudolph Smith bought 1000 shares of tesla Inc. in March 2020

# Issue # 1: No fixed size of neurons in a layer



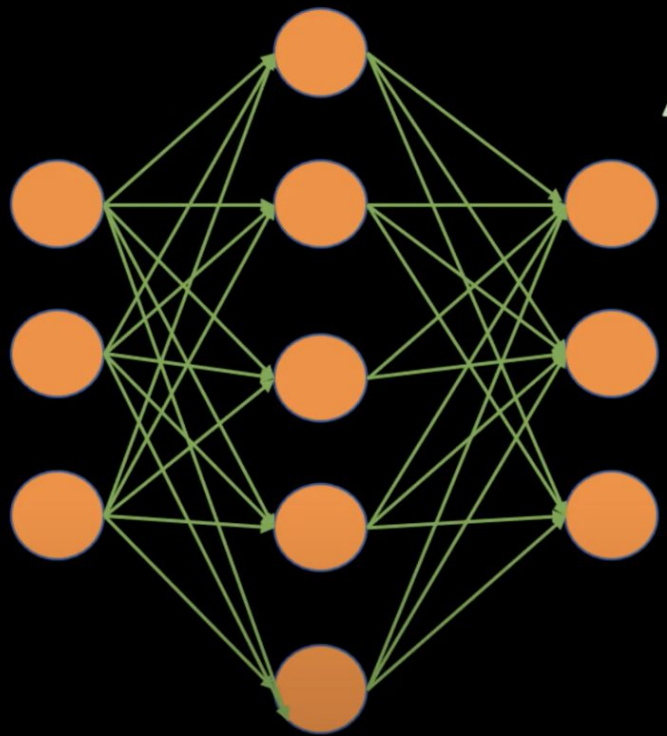
# Issue # 2: Too much computation

25000 words in vocabulary

how → [0,0,0,..,1,0,0,...,0 ]

are → [0, 1,0,0,0,..0,0,...,0 ]

you? → [0, 0,0,0,..0,0,1,0,0 ]

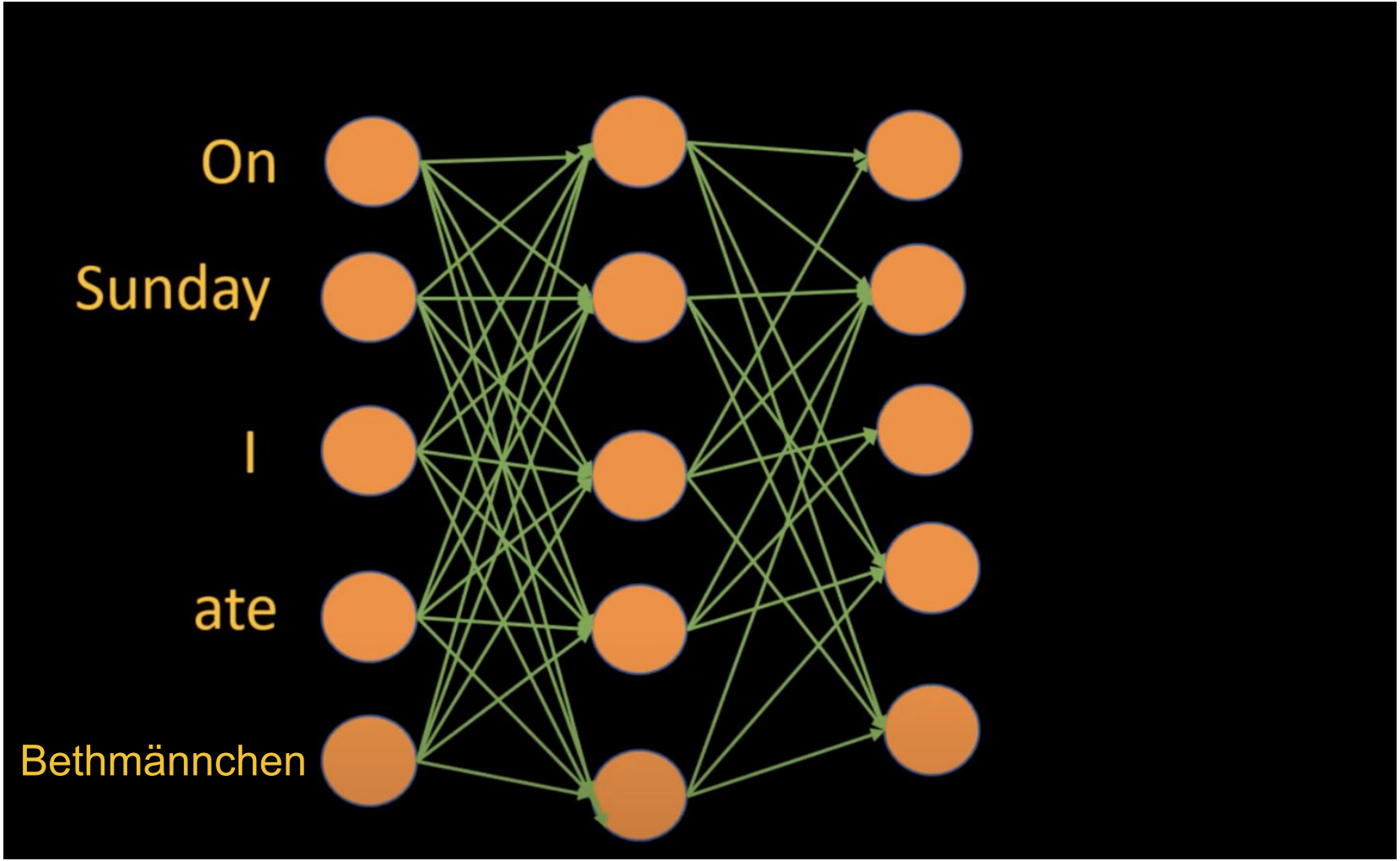


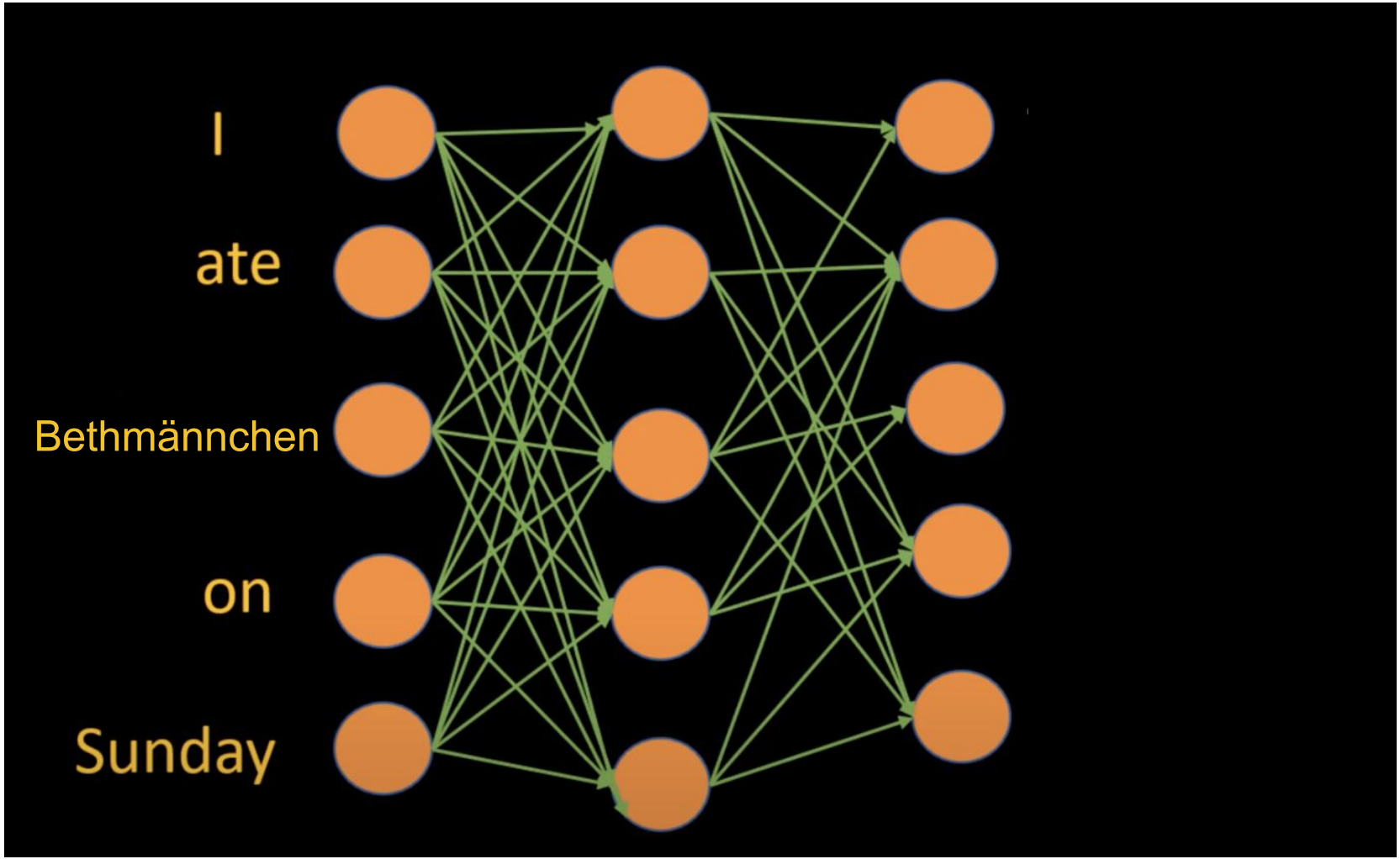
42000 words in vocabulary

[0,0,0,..,1,0,0,...,0 ]

[0,0,0,..,1,0,0,...,0 ]

[0,0,0,..,1,0,0,...,0 ]


On
Sunday
I
ate
Bethmännchen


I
ate
Bethmännchen
on
Sunday

Sarah loves Michael Jackson

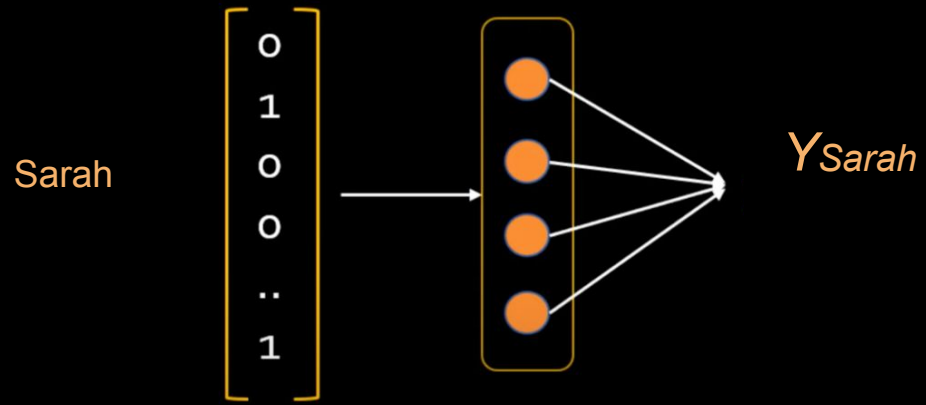
person person

Sarah loves Michael Jackson

1 0 1 1

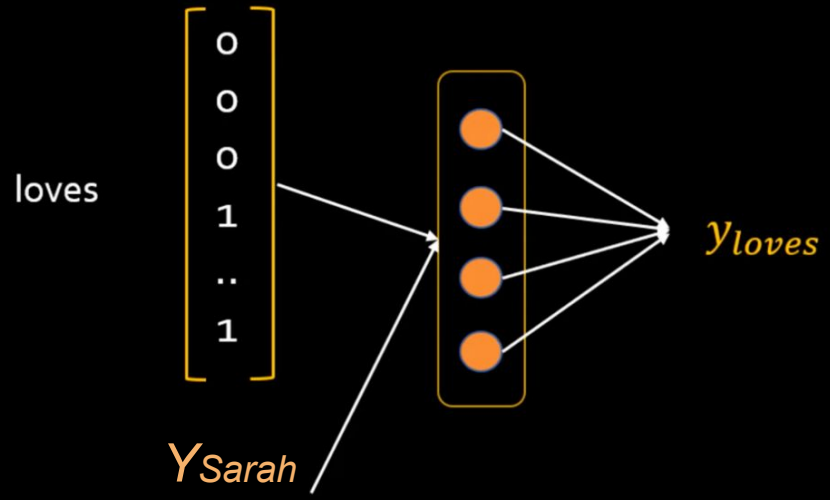
# Named Entity Recognition

Sarah loves Michael Jackson



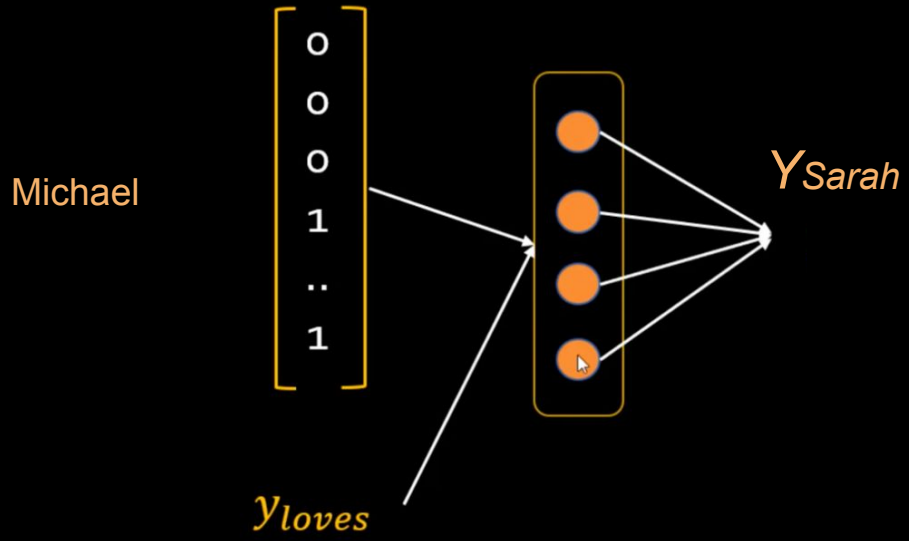
# Named Entity Recognition

Sarah loves Michael Jackson

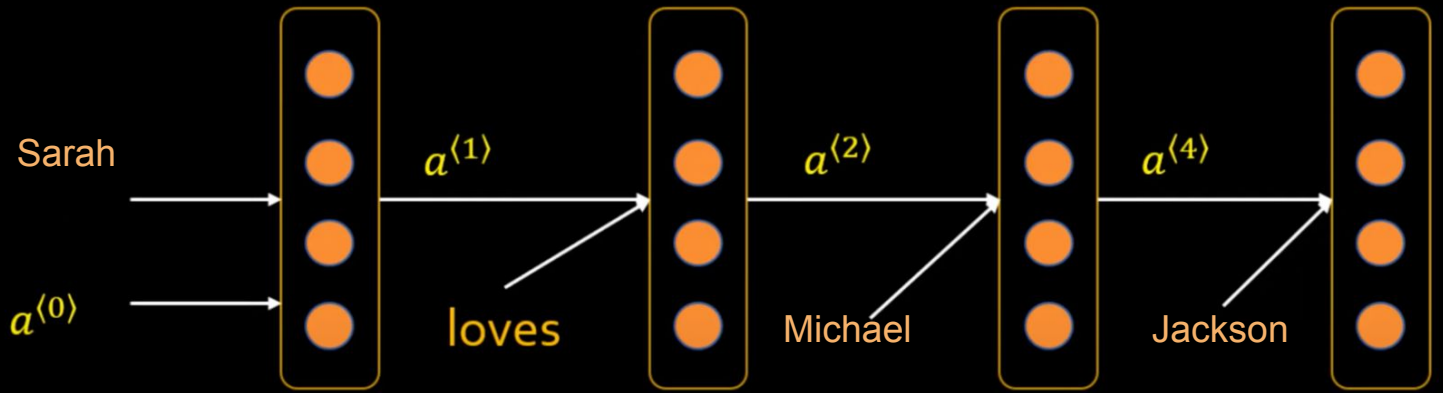


# Named Entity Recognition

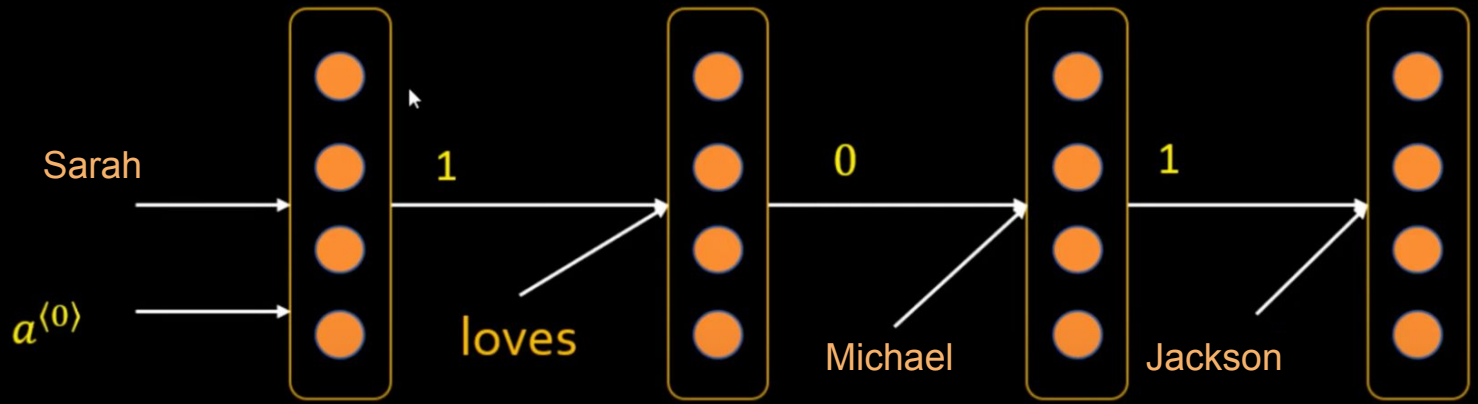
Sarah loves Michael Jackson

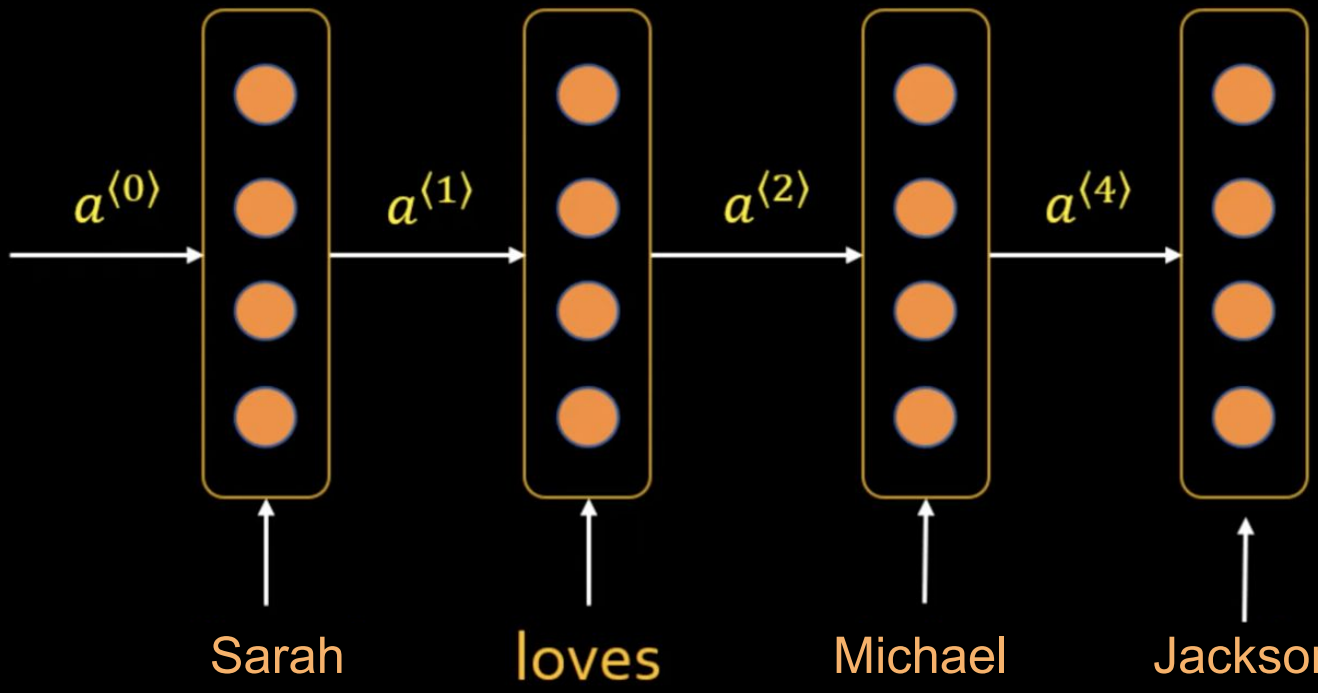


# Named Entity Recognition



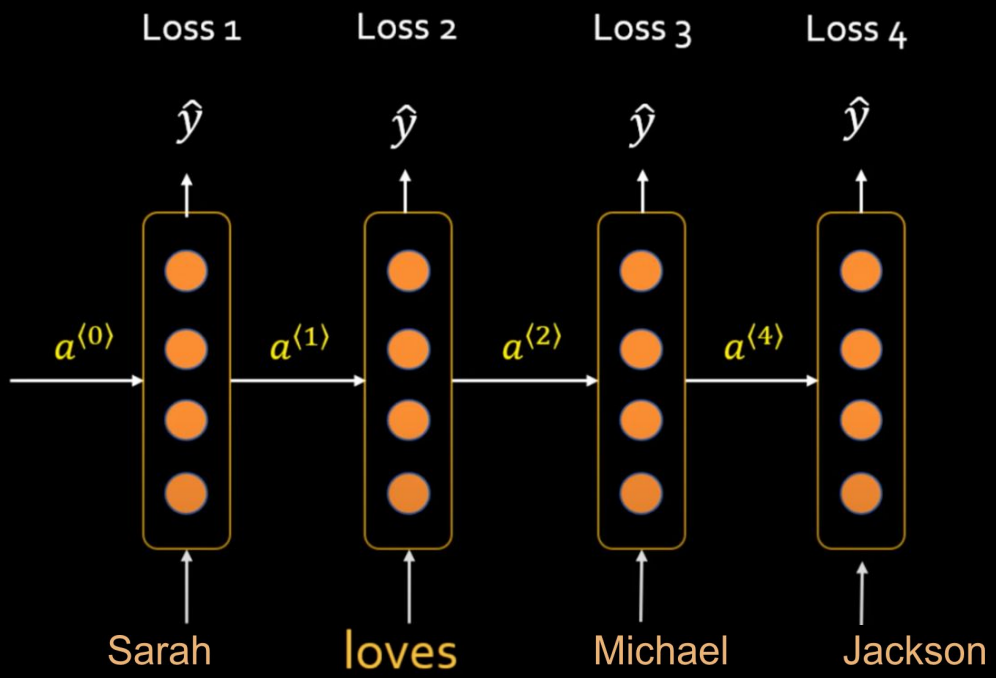
# Named Entity Recognition: once network is trained




$a^{\langle 0 \rangle}$
$a^{\langle 1 \rangle}$
$a^{\langle 2 \rangle}$
$a^{\langle 4 \rangle}$
Sarah
loves
Michael
Jackson

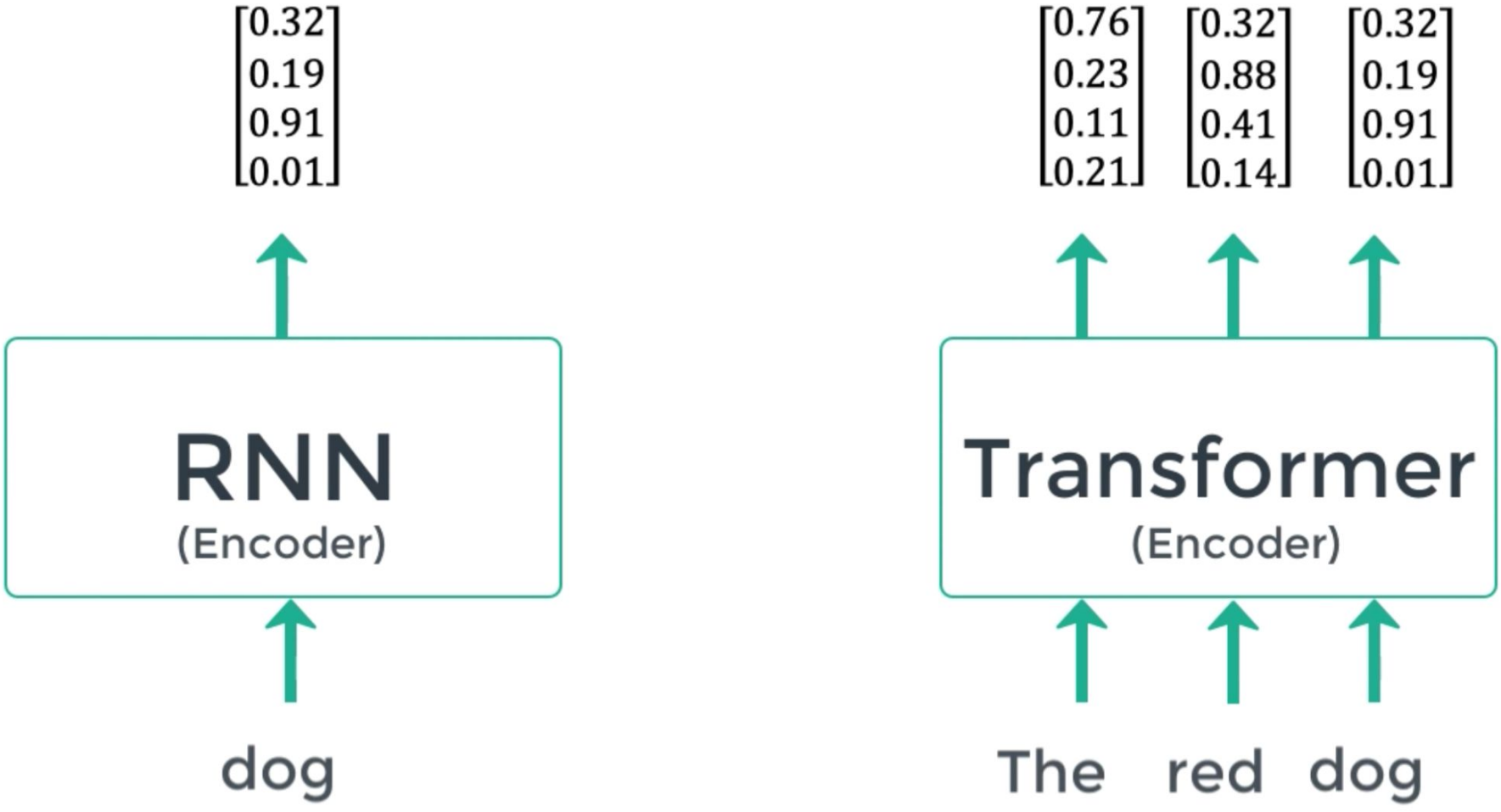
# Training

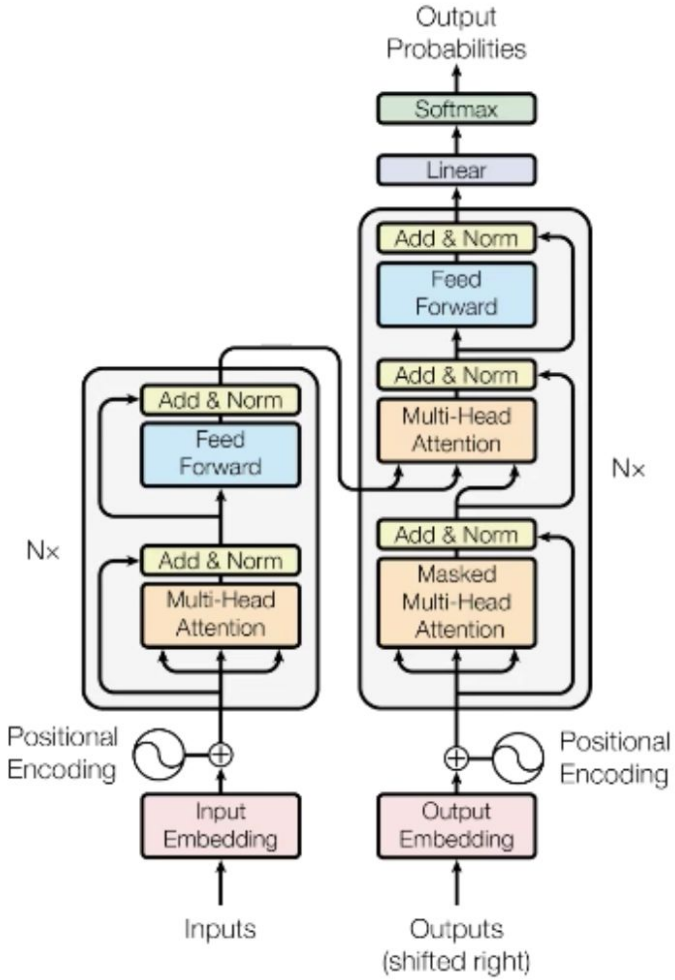
Sarah loves Michael Jackson → 1 0 1 1




Total Loss = Loss 1 + Loss 2 + Loss 3 + Loss4

# Transformers


0.32
0.19
0.91
0.01
RNN
(Encoder)
dog
0.76
0.23
0.11
0.21
0.32
0.88
0.41
0.14
0.32
0.19
0.91
0.01
Transformer
(Encoder)
The
red
dog


Output Probabilities
Softmax
Linear
Add & Norm
Feed Forward
Add & Norm
Multi-Head Attention
Add & Norm
Masked Multi-Head Attention
Nx
Add & Norm
Feed Forward
Add & Norm
Multi-Head Attention
Nx
Positional Encoding
Positional Encoding
Input Embedding
Output Embedding
Inputs
Outputs (shifted right)

# Transformer Components

## Input Embedding

dog →

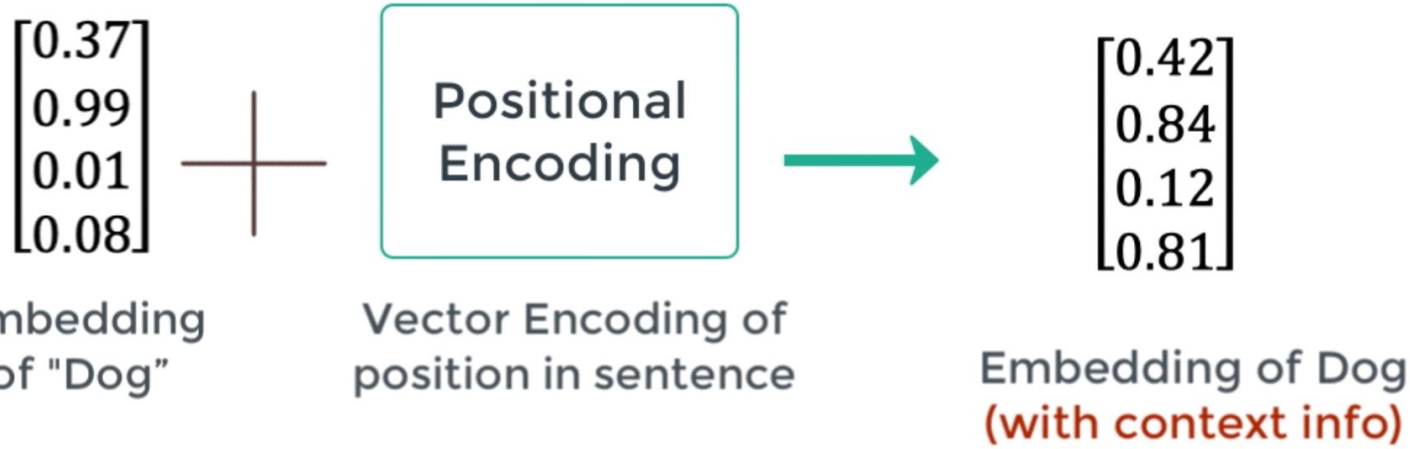

→ $\begin{bmatrix} 0.37 \\ 0.99 \\ 0.01 \\ 0.08 \end{bmatrix}$

AJ's dog is a cutie

AJ looks like a dog

# Transformer Components

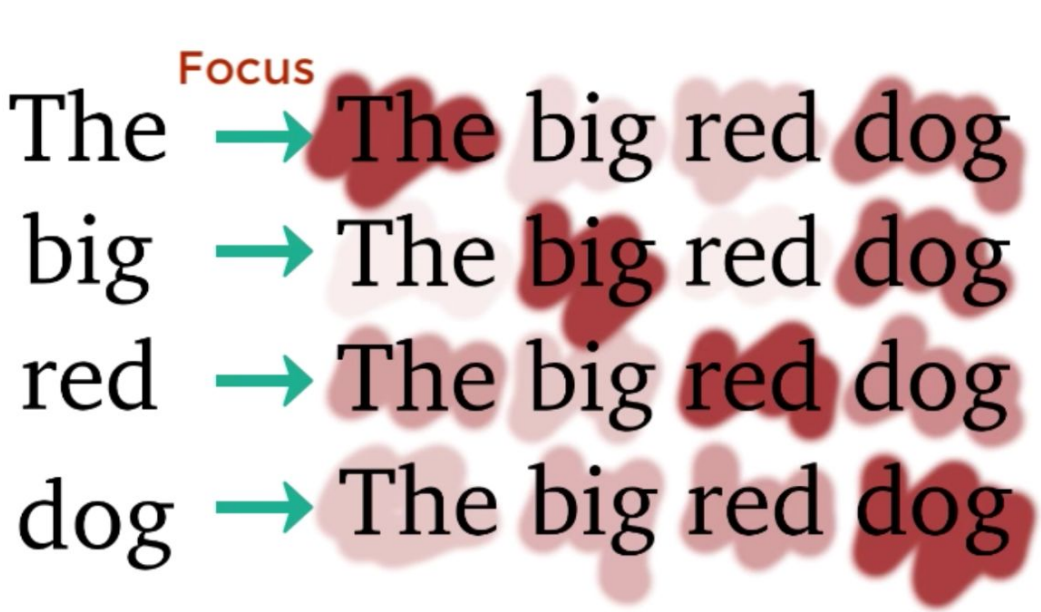
**Positional Encoder** :vector that gives context based on position of word in sentence




$$PE_{(pos,2i)} = sin(pos/10000^{2i/d_{\text{model}}})$$
$$PE_{(pos,2i+1)} = cos(pos/10000^{2i/d_{\text{model}}})$$

# Transformer Components

**Attention** : What part of the input should we focus?




Attention Vectors

$[0.71 \quad 0.04 \quad 0.07 \quad 0.18]^T$

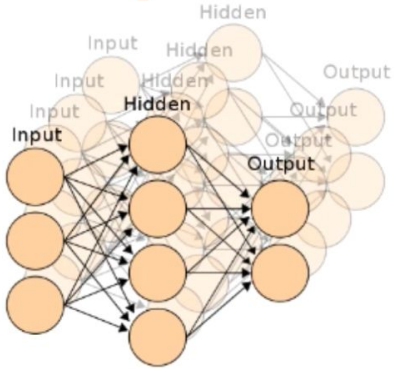
$[0.01 \quad 0.84 \quad 0.02 \quad 0.13]^T$

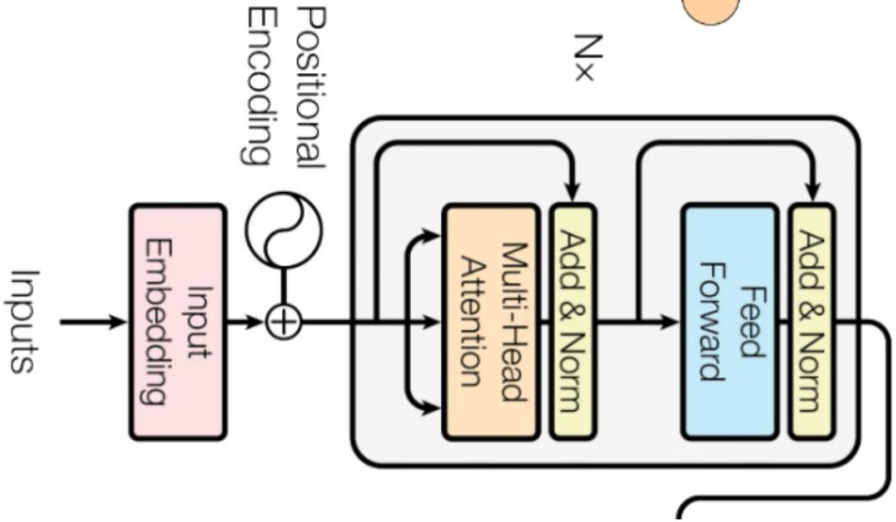
$[0.09 \quad 0.05 \quad 0.62 \quad 0.24]^T$

$[0.03 \quad 0.03 \quad 0.03 \quad 0.91]^T$

# Transformer Components

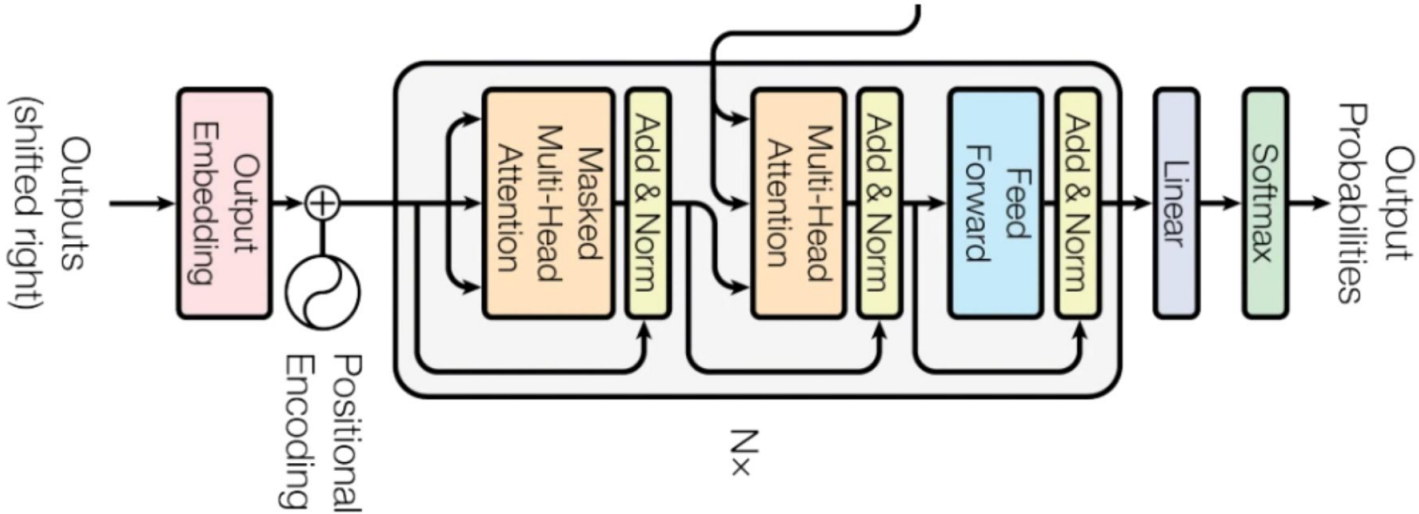




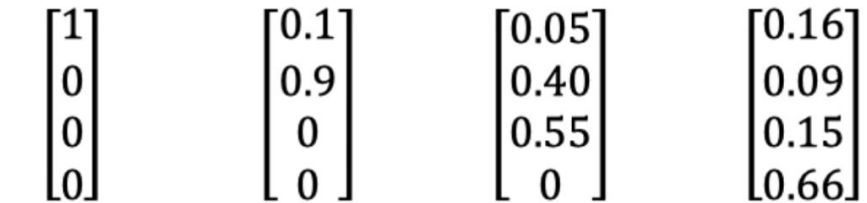
# Transformer Components

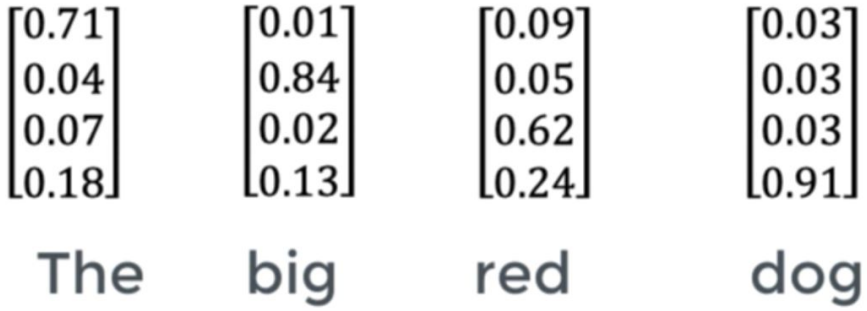
## Decoder



# Transformer Components

## Decoder



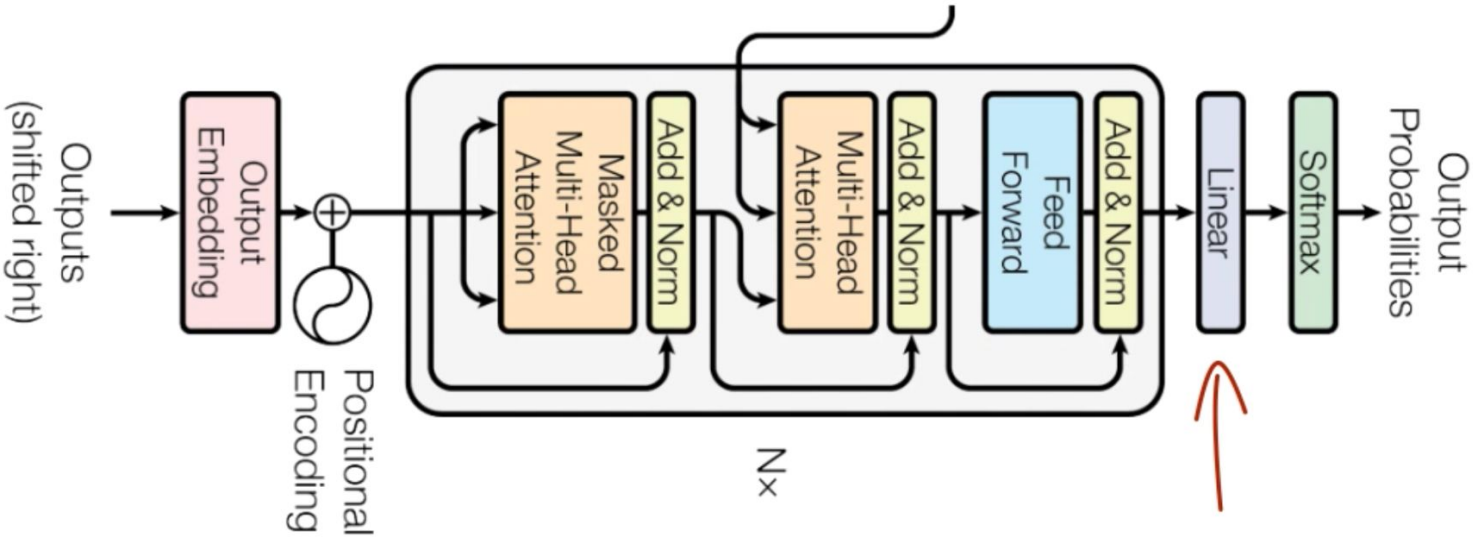




Encapsulates English-French Interactions

Encoder-Decoder Attention

# Transformer Components

Le




Feed Forward Layer

# Neurons =
# words in French