

Text-Mining code demos:

Basic NLP

Structure of the Day

- Background
- Data Analysis
- Simple Models

LUNCH

- Deep Learning
- Prompting LLMs

Fundamental Text Tasks

- Translation
- Summarization
- **Classification**
- **Information Extraction**

Classification



Can't fault it



Camera can easily be disabled by thieves



Excellent, effective, product.



Not fit for purpose as a security camera



Does what you ask



Not worth the time



Classification

**RescuePH** @RescuePH 15h

#RescuePH, **#ReliefPH**, and **#TracingPH** are /SERIOUSLY/ monitored hashtags. Please refrain from tweeting useless info. Pls RT

 Retweeted 431 times

[Collapse](#) [↩ Reply](#) [↻ Retweet](#) [★ Favorite](#) [⋮ More](#)

431 RETWEETS	38 FAVORITES	
------------------------	------------------------	--

12:32 AM - 14 Nov 13 · Details

Classification



Classification

I have been working with Archie, a 10-year-old diabetic cat. He currently receives 3 units of ProZinc insulin

Cat vs Dog
Neutered / Spayed?

Information Extraction

I have been working with Archie, a 10-year-old **diabetic** cat.
He currently receives **3 units** of **ProZinc insulin**

DOSE **DRUG** **DISEASE**

Information Extraction

I have been working with Archie, a 10-year-old **diabetic** cat.
He currently receives **3 units** of **ProZinc insulin**

DOSE **DRUG**



Information Extraction

●

Diabetes mellitus (disorder)

☆

📄

SCTID: 73211009

73211009 | Diabetes mellitus (disorder) |

en

 Diabetes mellitus (disorder)

en

 Diabetes mellitus

en

 DM - Diabetes mellitus

DISEASE

I have been working with Archie, a 10-year-old **diabetic** cat.
He currently receives **3 units** of **ProZinc insulin**

DOSE

DRUG

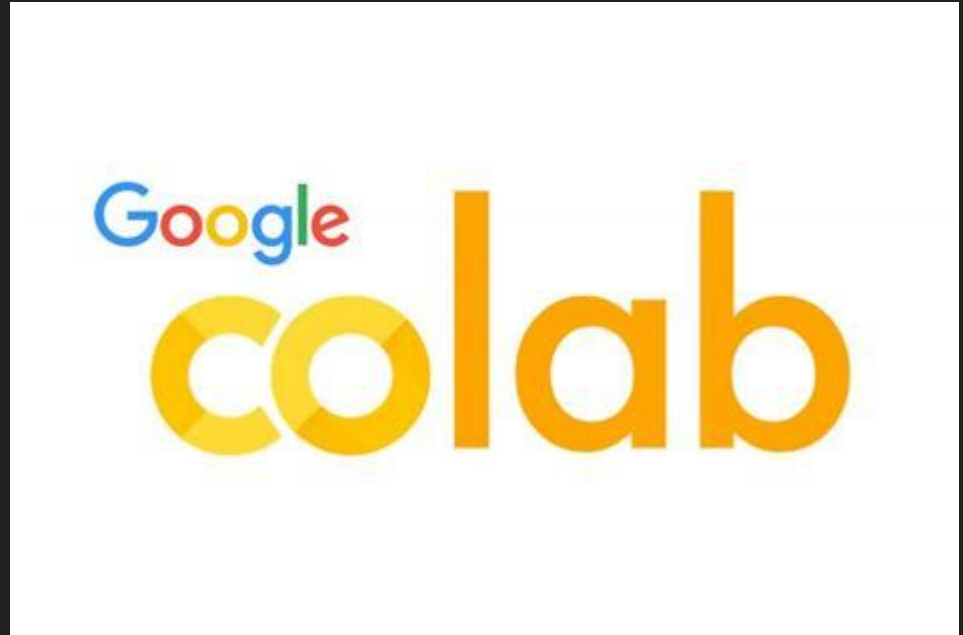


ProZinc 40 IU/ml Suspension for Injection for Cats and Dogs

MA holder	Boehringer Ingelheim Vetmedica GmbH
VM number	EU/2/13/152/001-002
Territory	Northern Ireland
Aligned product	
Associated documents	View EPAR
Date of issue	12/07/2013
Authorisation route	Centralised
Active substances	Insulin

Python Environment

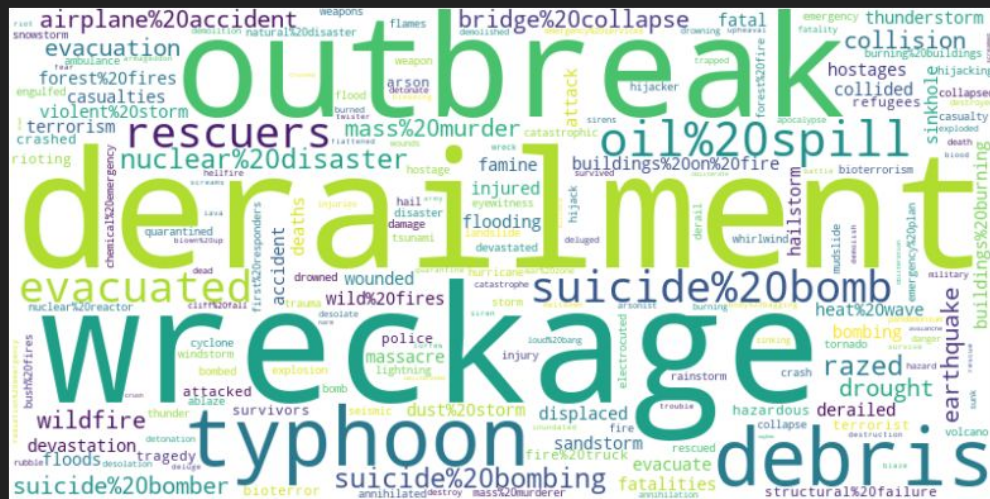
- Colab
- Local Machine
- Bede



<https://t.ly/FmAjD>

<https://github.com/ghomasHudson/text-mining-demos-workshop>

Notebook 1: Analysing the datasets

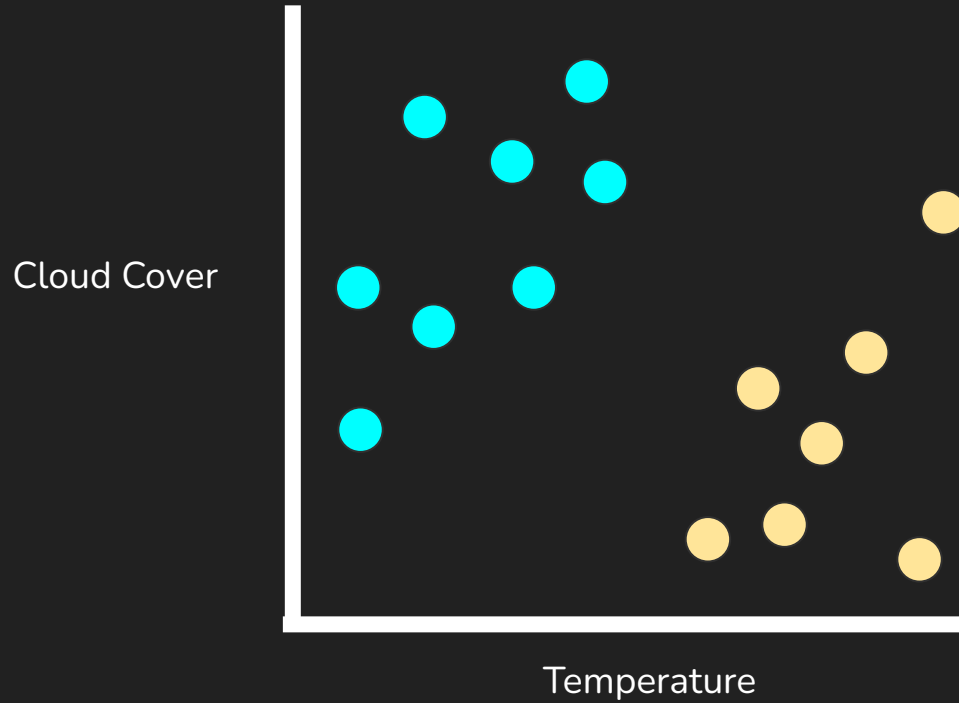


Simple Models

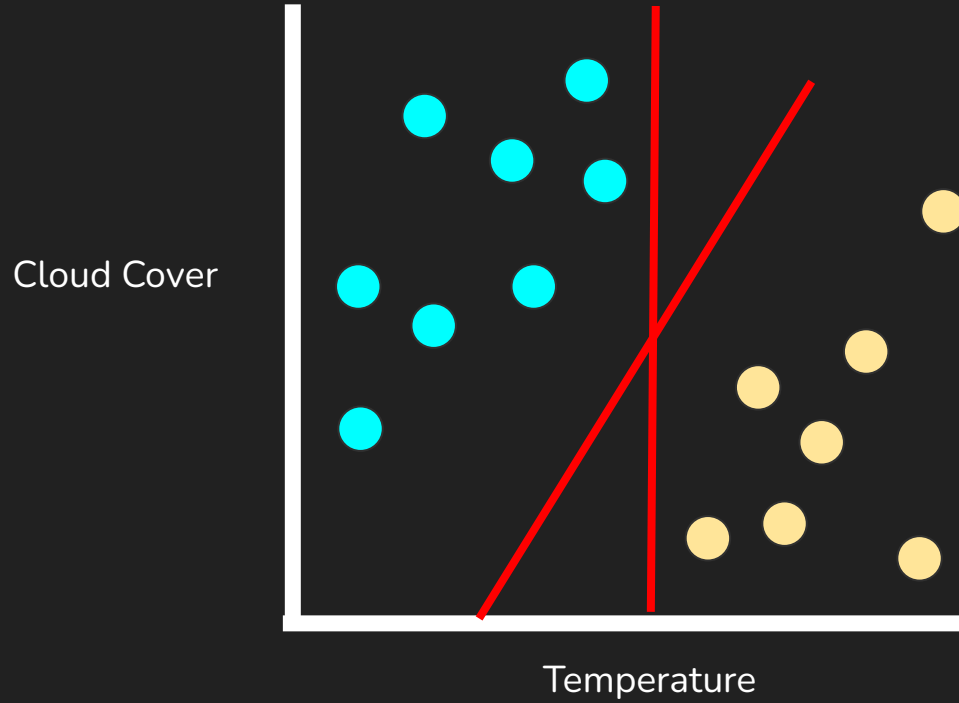
A Simple Model

Temperature	Cloud Cover	Rain
3.3	10	N
0.4	80	Y
...

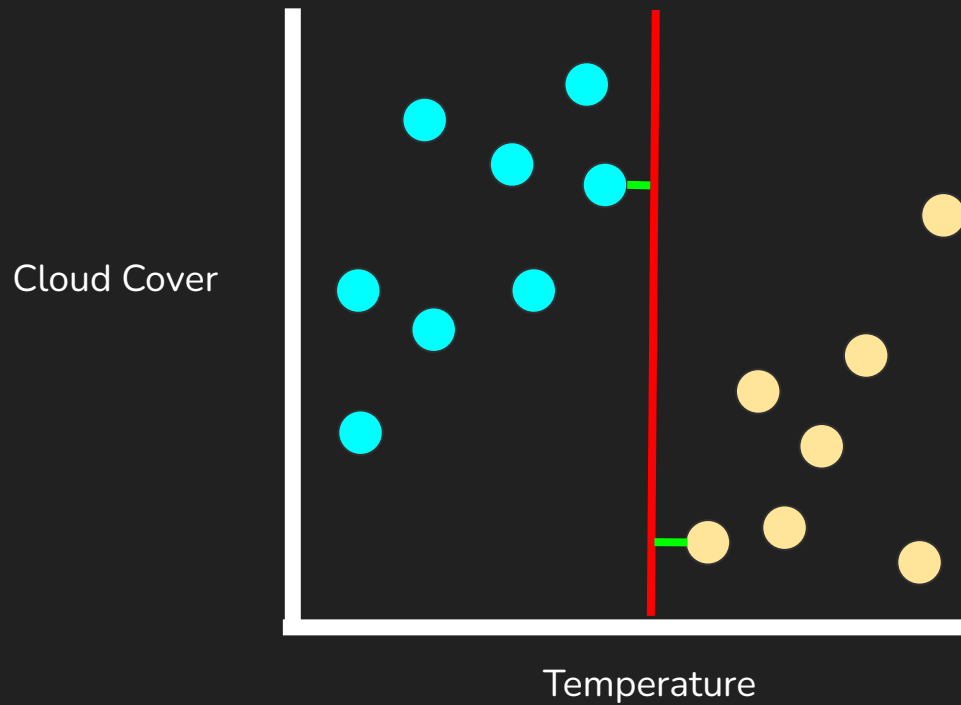
A Simple Model



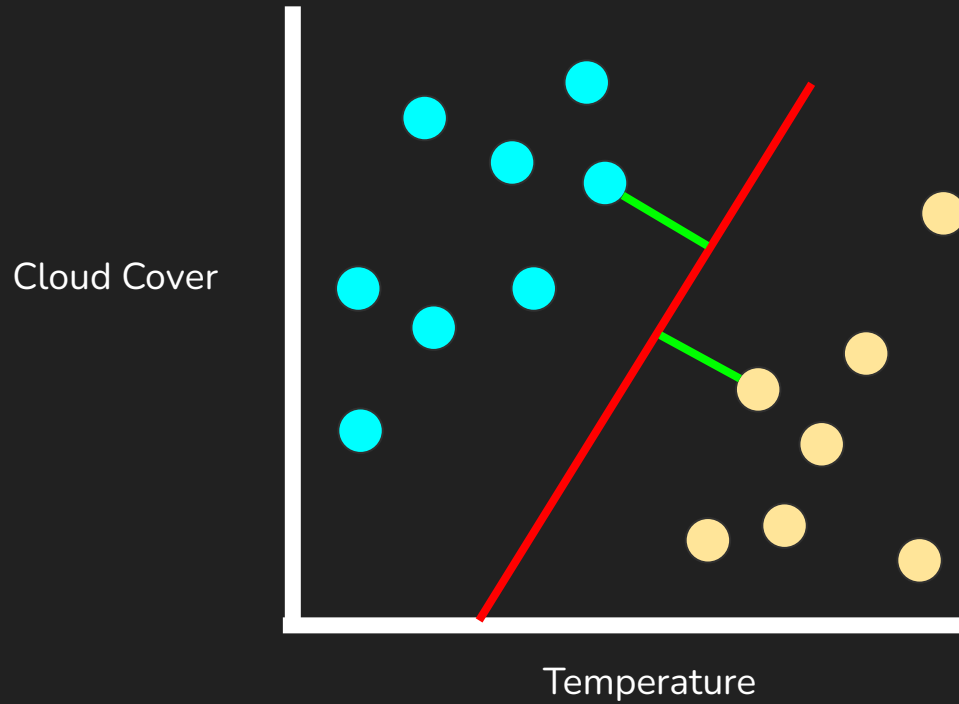
A Simple Model



A Simple Model



A Simple Model



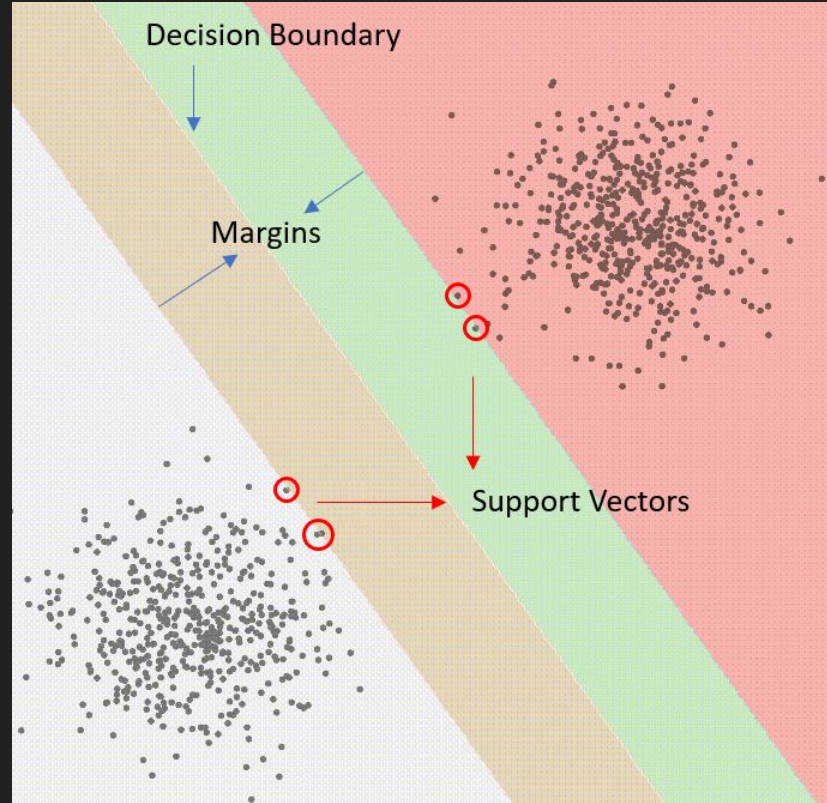
SVMs for Text

Forest fire near La Ronge Sask. Canada

Help my house is on fire. The fire is huge.

forest	fire	near	...	help	...
1	1	1		0	
0	2	0		1	

Notebook 2: SVMs for disaster tweet classification



The Problem with Bag of Words

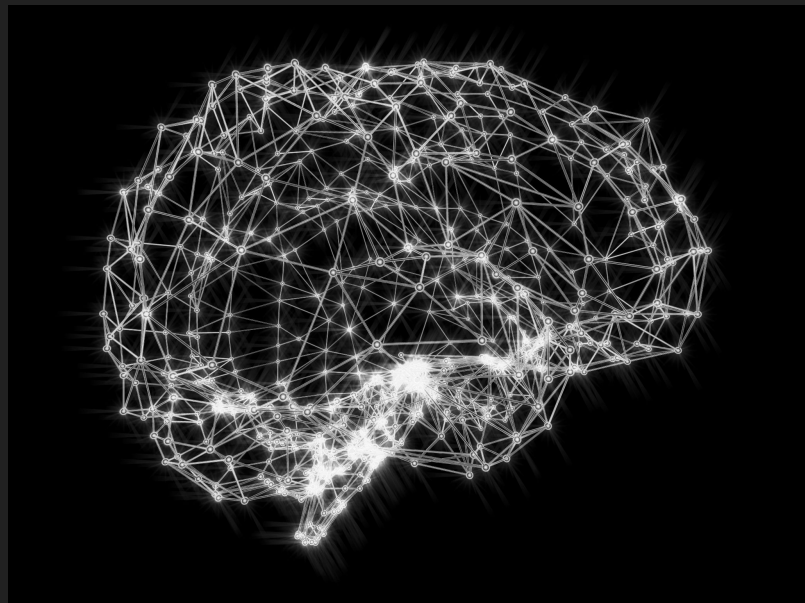
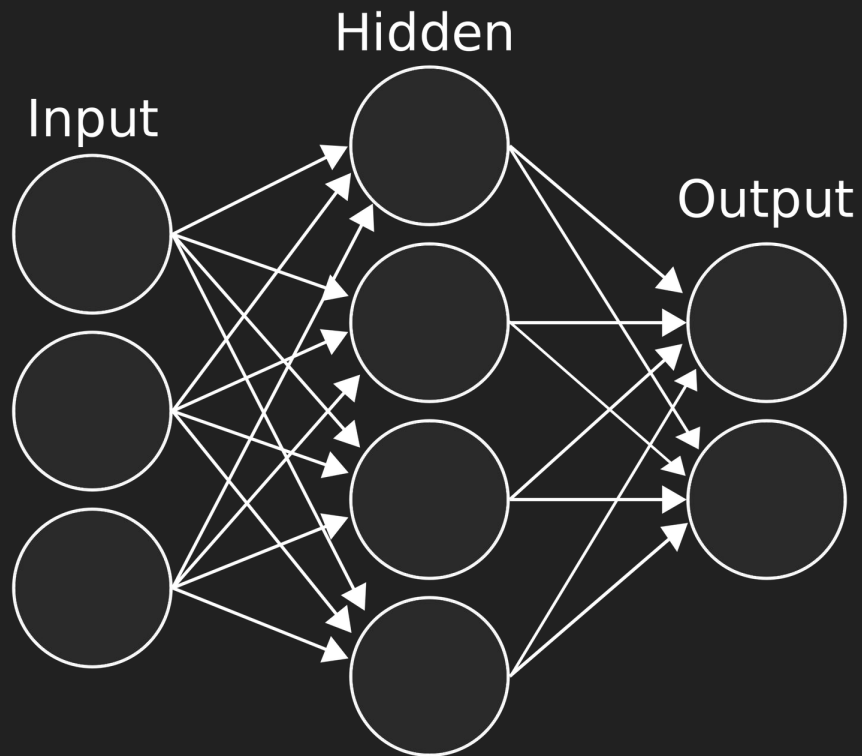
I have been working with a 10-year-old diabetic cat. **He** is treated with 3 units of ProZinc insulin.

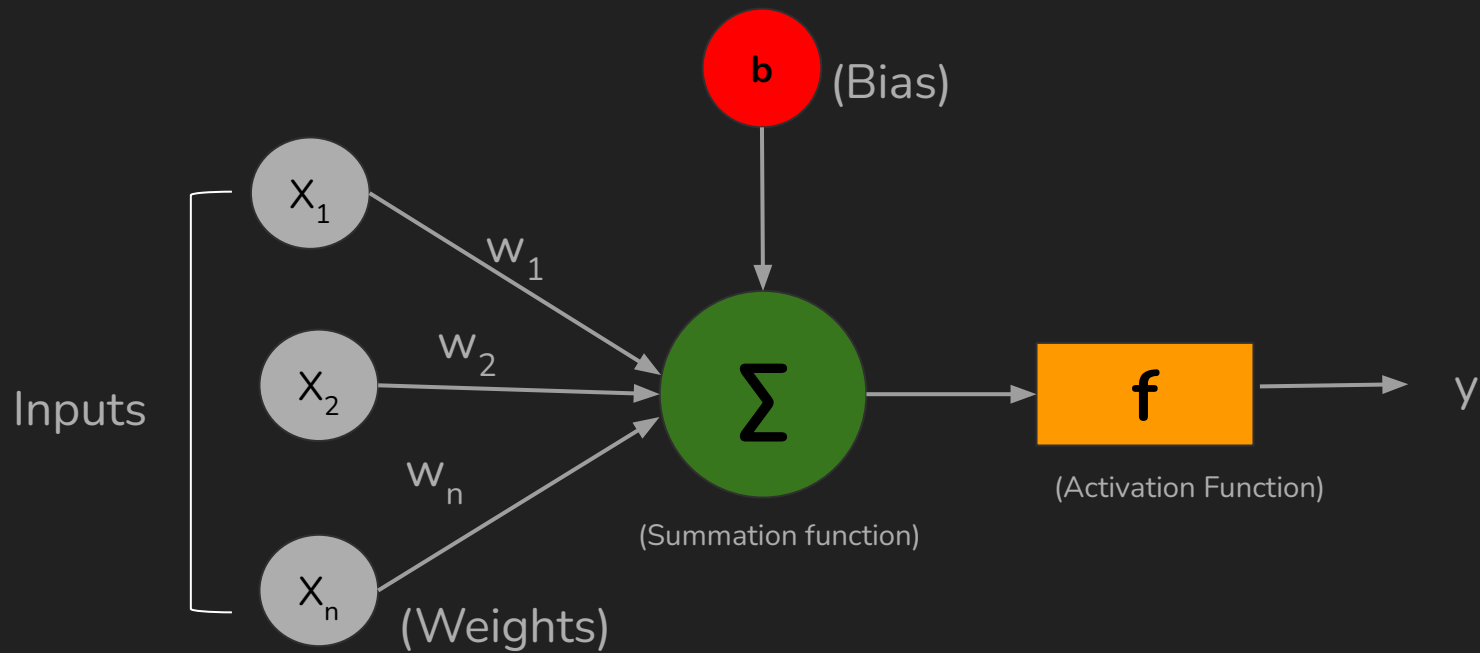
The Problem with Bag of Words

The Owner brought Jess into the surgery yesterday. He mentioned a history with diabetes.

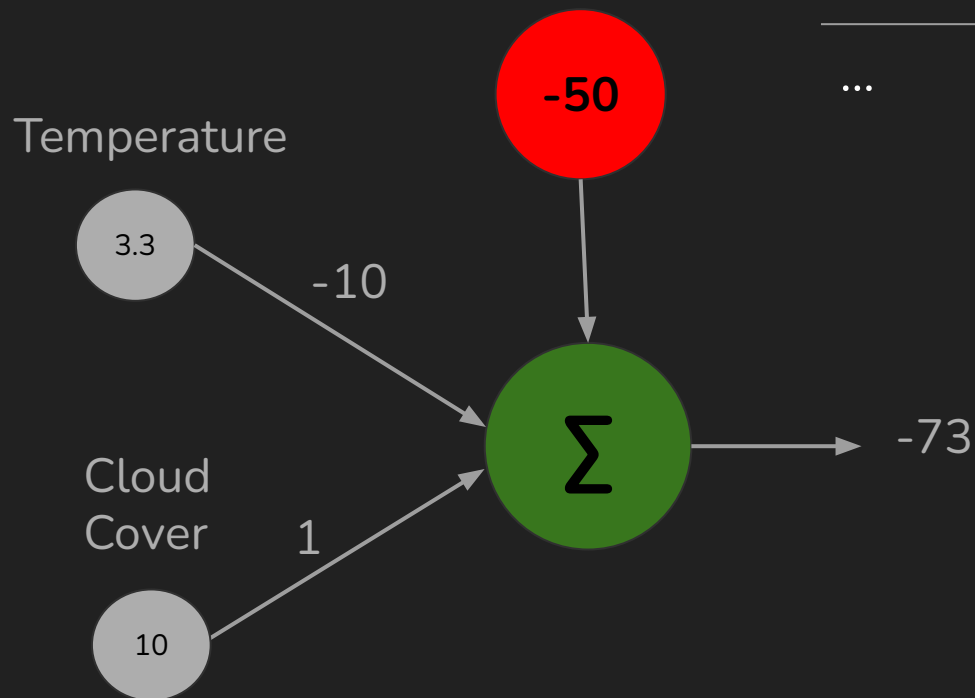
Neural Networks

Neural Networks

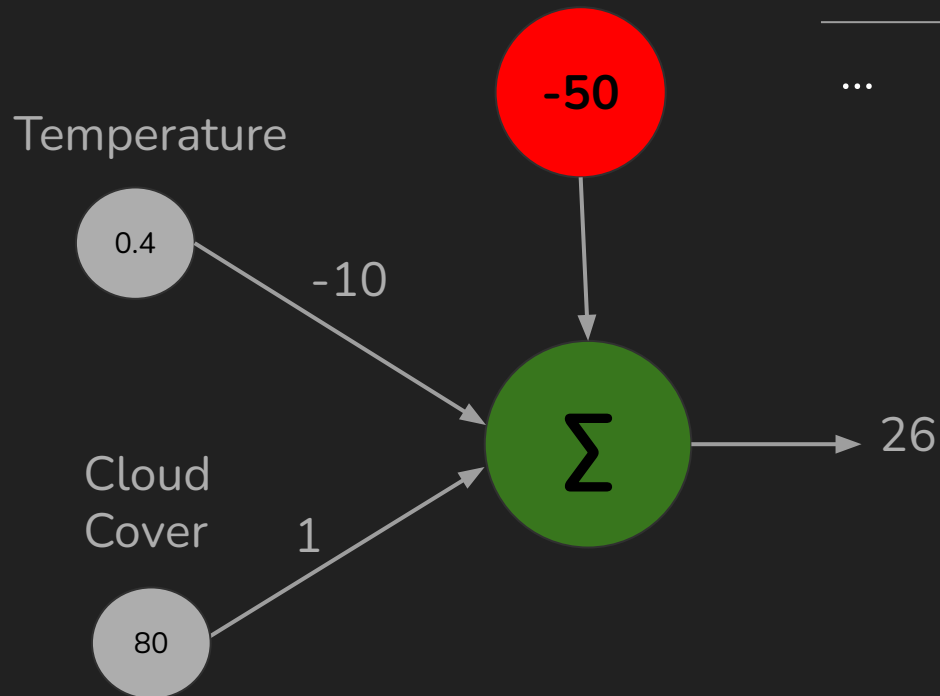




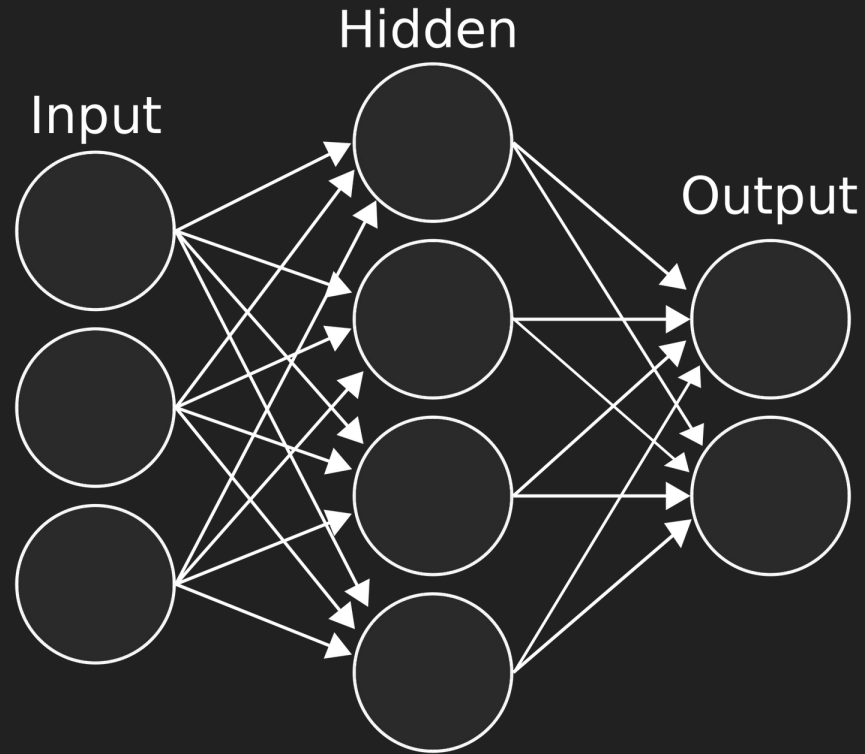
Temperature	Cloud Cover	Rain
3.3	10	N
0.4	80	Y
...



Temperature	Cloud Cover	Rain
3.3	10	N
0.4	80	Y
...



Neural Networks



Transformers

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

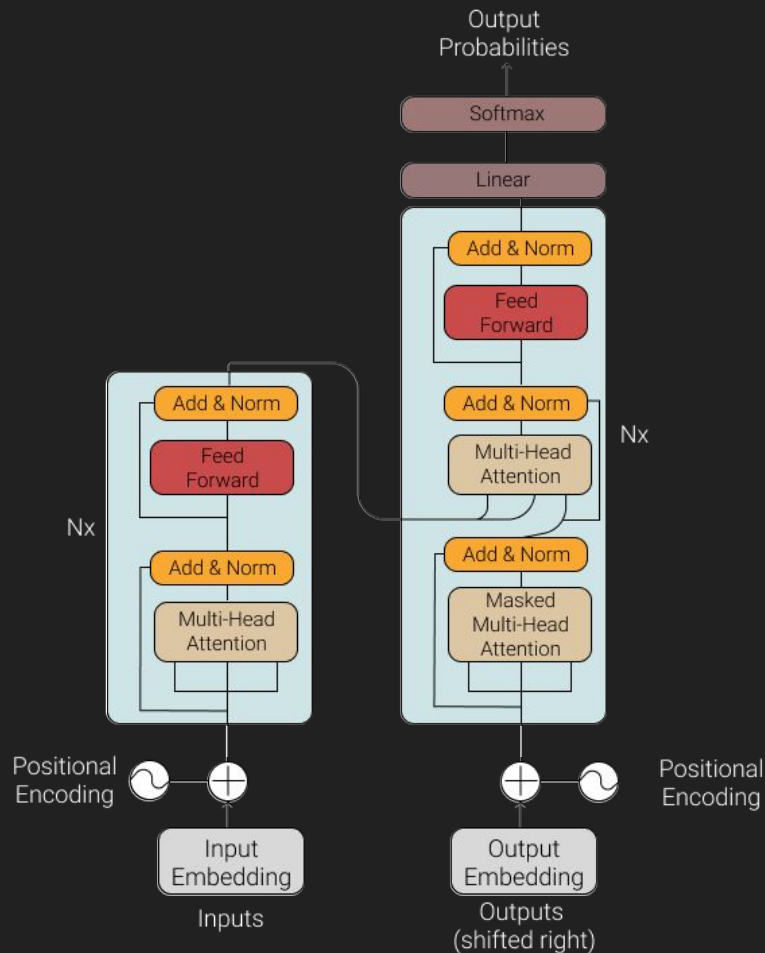
Aidan N. Gomez* †
University of Toronto
aidan@cs.toronto.edu

Lukasz Kaiser*
Google Brain
lukaszkaiser@google.com

Illia Polosukhin* ‡
illia.polosukhin@gmail.com

Abstract

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.8 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature. We show that the Transformer generalizes well to other tasks by applying it successfully to English constituency parsing both with large and limited training data.



The

animal

didn't

cross

the

street

because

it

was

too

tired

.

The

animal

didn't

cross

the

street

because

it

was

too

tired

.

The

animal

didn't

cross

the

street

because

it

was

too

wide

.

The

animal

didn't

cross

the

street

because

it

was

too

wide

.

Language Models



Notebook 3: Applying Transformers

Huggingface Hub

Tasks

Sizes

Sub-tasks

Languages

Licenses

Other

Filter Tasks by name

Multimodal

Feature Extraction

Text-to-Image

Image-to-Text

Text-to-Video

Visual Question Answering

Graph Machine Learning

Computer Vision

Depth Estimation

Image Classification

Object Detection

Image Segmentation

Image-to-Image

Unconditional Image Generation

Video Classification

Zero-Shot Image Classification

Natural Language Processing

Text Classification

Token Classification

Table Question Answering

Question Answering

Zero-Shot Classification

Translation

Summarization

Conversational

Datasets 74,115

Filter by name

new Full-text search

Sort: Trending

THUDM/AgentInstruct

Viewer • Updated 7 days ago • 870 • 92

fka/awesome-chatgpt-prompts

Viewer • Updated Mar 7 • 1.67k • 3.65k

Skywork/SkyPile-150B

Preview • Updated about 12 hours ago • 649 • 36

EleutherAI/proof-pile-2

Viewer • Updated 5 days ago • 1.78k • 71

openbmb/UltraFeedback

Viewer • Updated 30 days ago • 1.67k • 136

laion/dalle-3-dataset

Viewer • Updated about 2 hours ago • 1.45k • 165

HuggingFaceH4/ultrachat_200k






Viewer • Updated 3 days ago • 309 • 22

stingning/ultrachat

Viewer • Updated 18 days ago • 21.1k • 240

No labels?

Prompting

-  Reset Thread
-  Dark Mode
-  OpenAI Discord
-  FAQ
-  Log out

ChatGPT



Examples

"Explain quantum computing in simple terms"

"Got any creative ideas for a 10 year old's birthday?"

"How do I make an HTTP request in Javascript?"



Capabilities

Remembers what user said earlier in the conversation

Allows user to provide follow-up corrections

Trained to decline inappropriate requests



Limitations

May occasionally generate incorrect information

May occasionally produce harmful instructions or biased content

Limited knowledge of world and events after 2021



Free Research Preview: ChatGPT is optimized for dialogue. Our goal is to make AI systems more natural to interact with, and your feedback will help us improve our systems and make them safer.

Notebook 4: Prompting for Text-mining

Can we improve prompting further?

Yes

Lots of ways

www.promptingguide.ai

Chain of thought

Break down complex problems

Standard Prompting

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The answer is 27. ❌

Chain-of-Thought Prompting

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. $5 + 6 = 11$. The answer is 11.

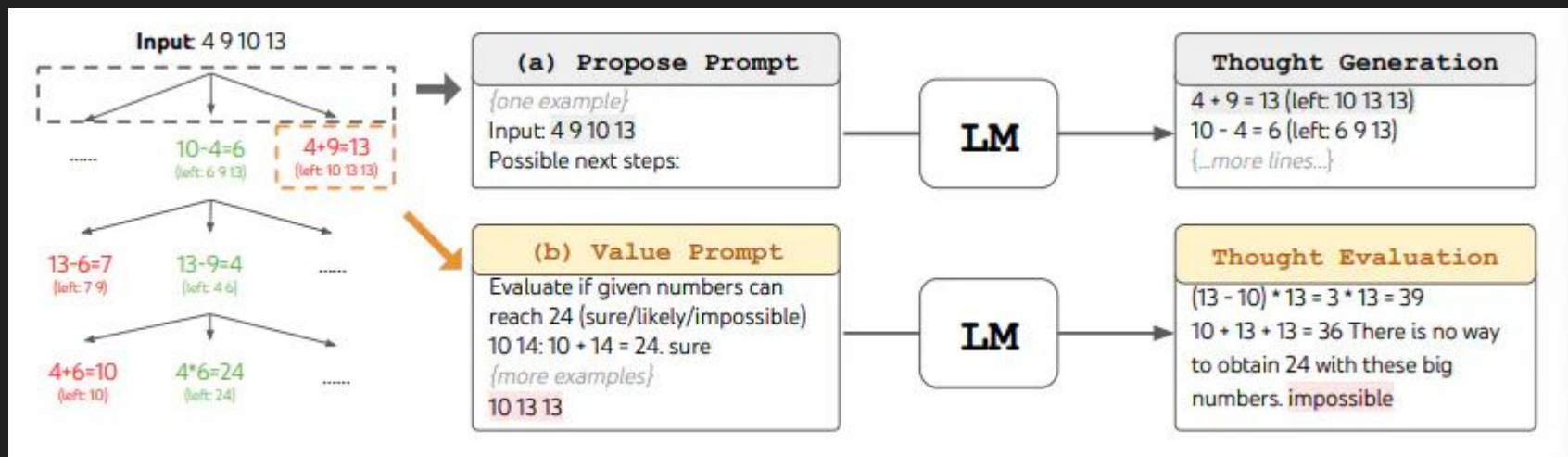
Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had $23 - 20 = 3$. They bought 6 more apples, so they have $3 + 6 = 9$. The answer is 9. ✅

Tree of thought

Pick the best way to break down complex problems



Notebook 5: Chain of Thought