



# Data-Science Karaoke

---

It is as awful as it sounds



# RULES

- 1) Team up in 4 teams (2-3 people is ideal)
- 2) Each team must present their part for 5 minutes minimum (more is fine) without seeing the presentation beforehand and acting as if they created it
- 3) Expect around 5-10 slides per presentation
- 4) There is a 2 minute Q&A after each presentation
- 5) Have fun!

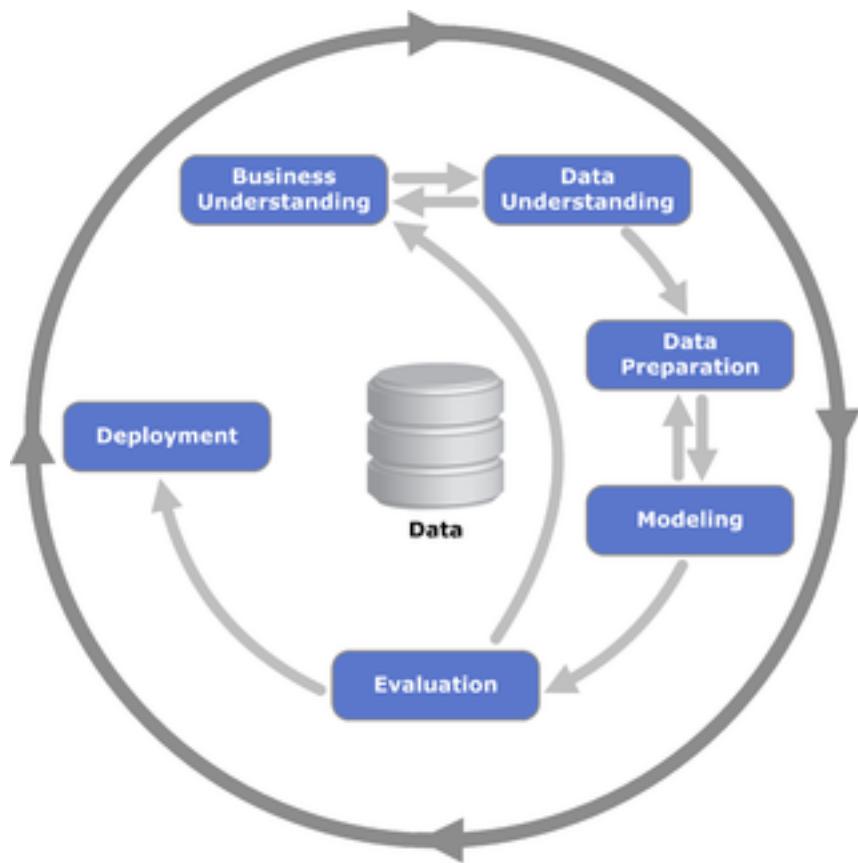
# STOP HERE

Do not look at the slides in advance to preserve the fun

# Data Science

Part I – Frameworks

# Cross-industry standard process for data mining (CRISP-DM)



# Cross-industry standard process for data mining (CRISP-DM)



# Cross-industry standard process for data mining (CRISP-DM)



- Old and outdated

# Cross-industry standard process for data mining (CRISP-DM)



- Old and outdated
- Needs a new hip approach

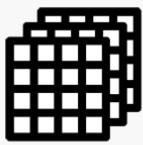
# Our new approach: MB-ML

Our new approach: MB-ML

*Meme Based Machine Learning*

DATASET

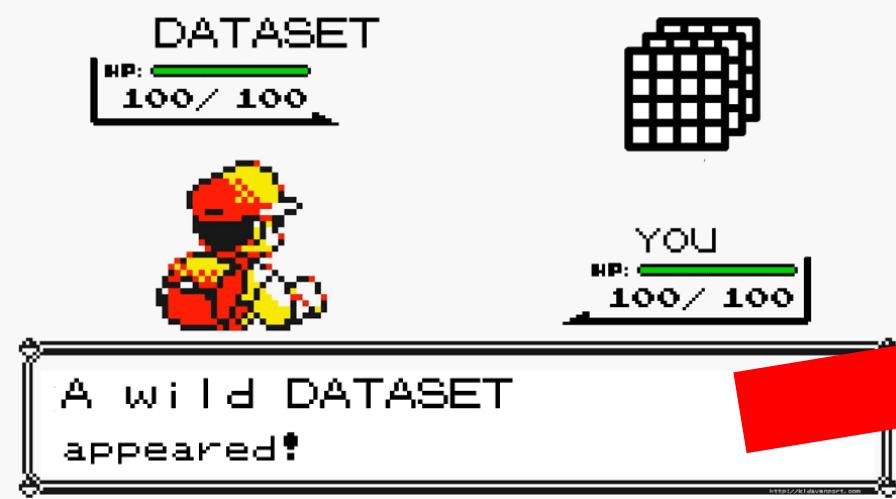
HP:  100 / 100



YOU

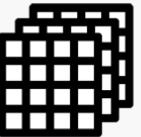
HP:  100 / 100

A wild DATASET  
appeared!



DATASET

HP: 100 / 100



A wild DATASET  
appeared!

YOU

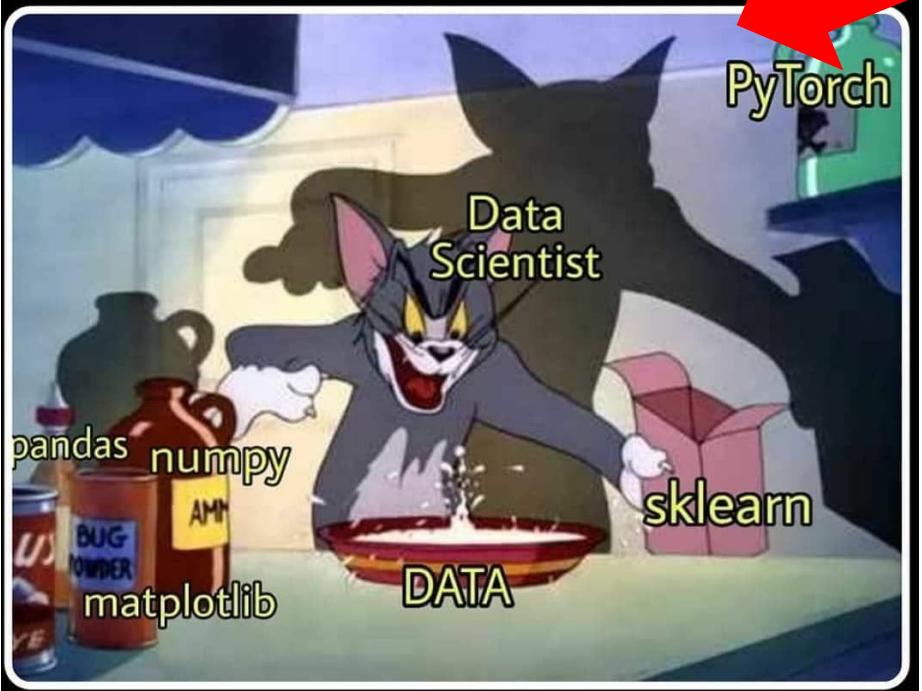
HP: 100 / 100



DIRTY DATA YOU  
HAVE

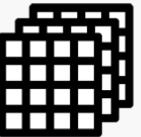
CLEAN IT UP YOU MUST

memegenerator.net



DATASET

HP: 100 / 100



A wild DATASET  
appeared!

YOU

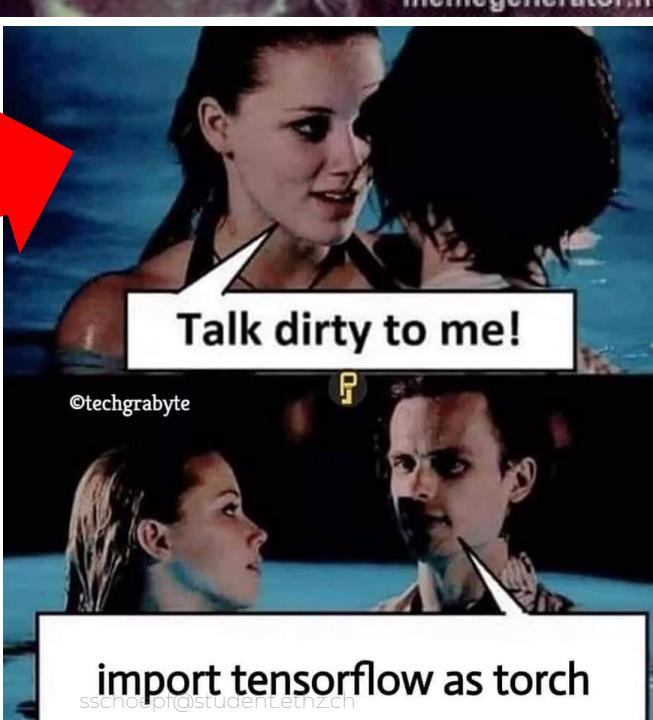
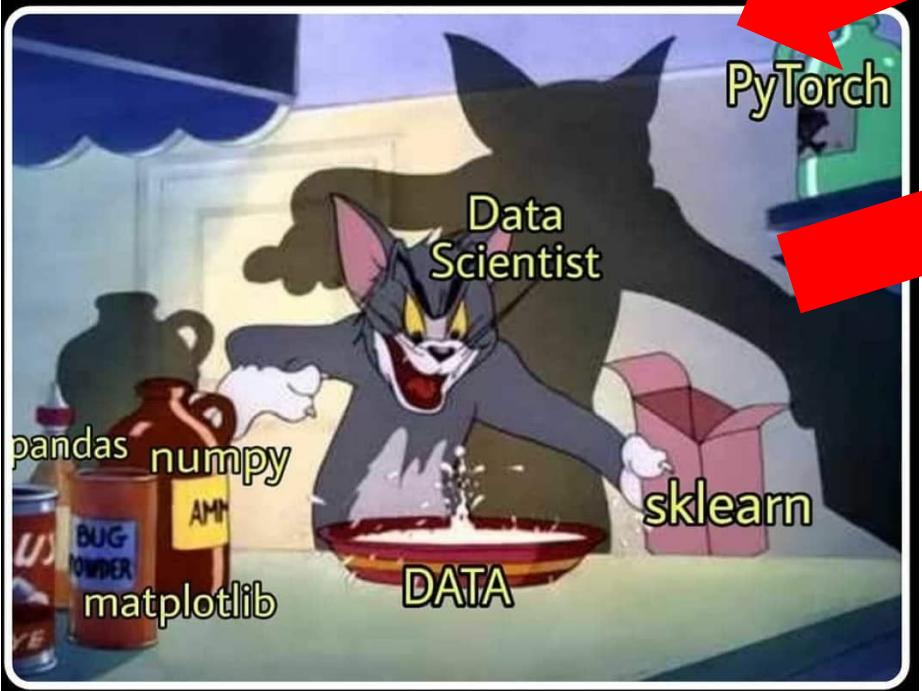
HP: 100 / 100

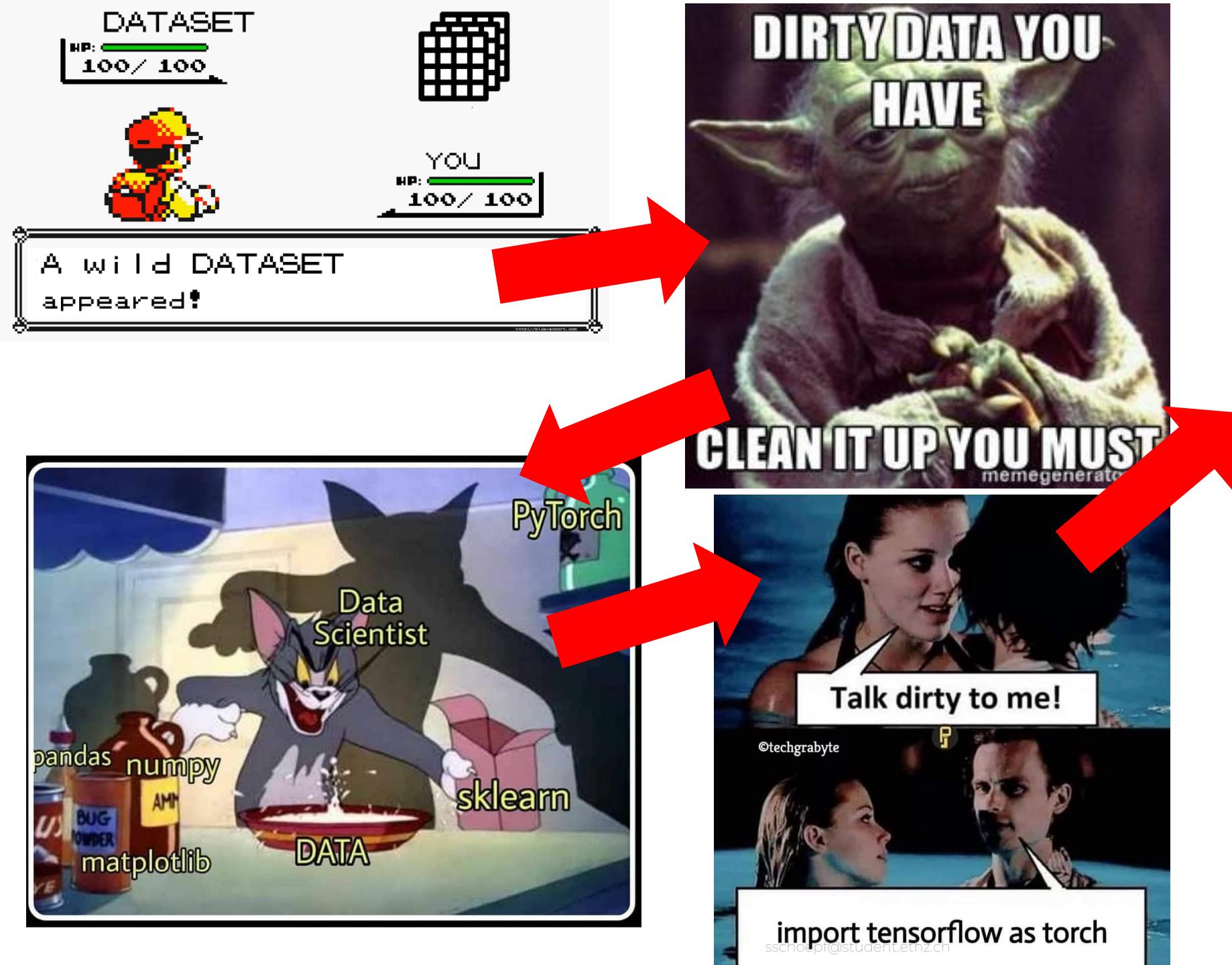


DIRTY DATA YOU  
HAVE

CLEAN IT UP YOU MUST

memegenerator.net





DATASET

HP: 100 / 100



A wild DATASET  
appeared!

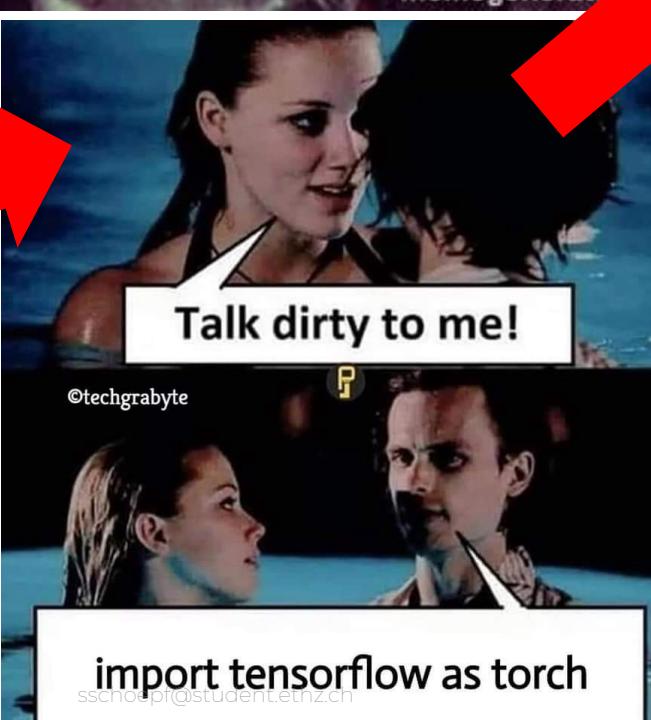
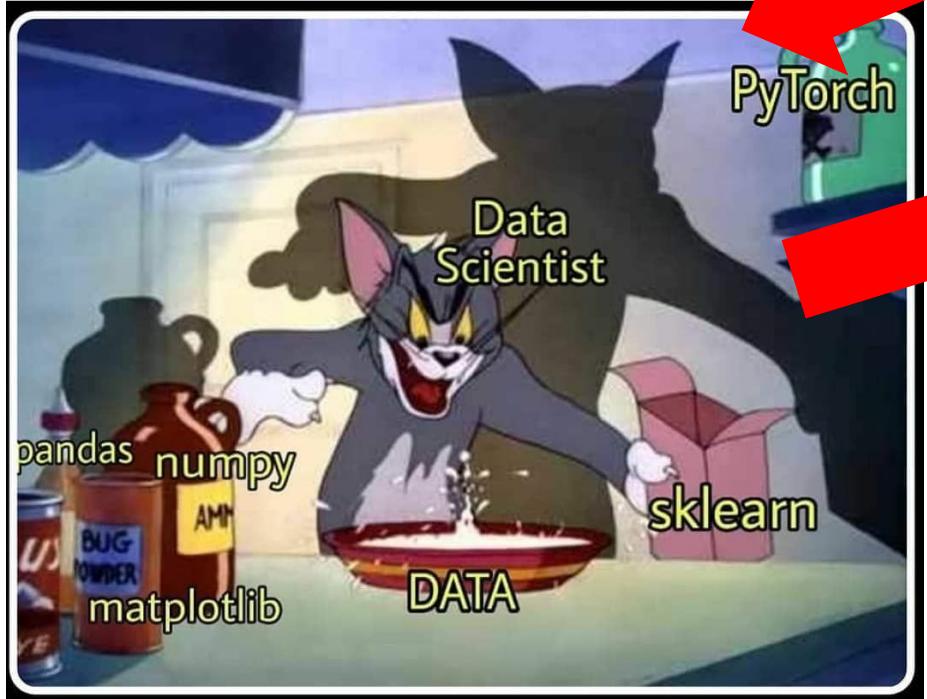
YOU

HP: 100 / 100

DIRTY DATA YOU  
HAVE

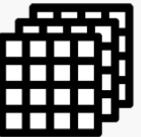
CLEAN IT UP YOU MUST

memegenerator.net



DATASET

HP: 100 / 100



A wild DATASET  
appeared!

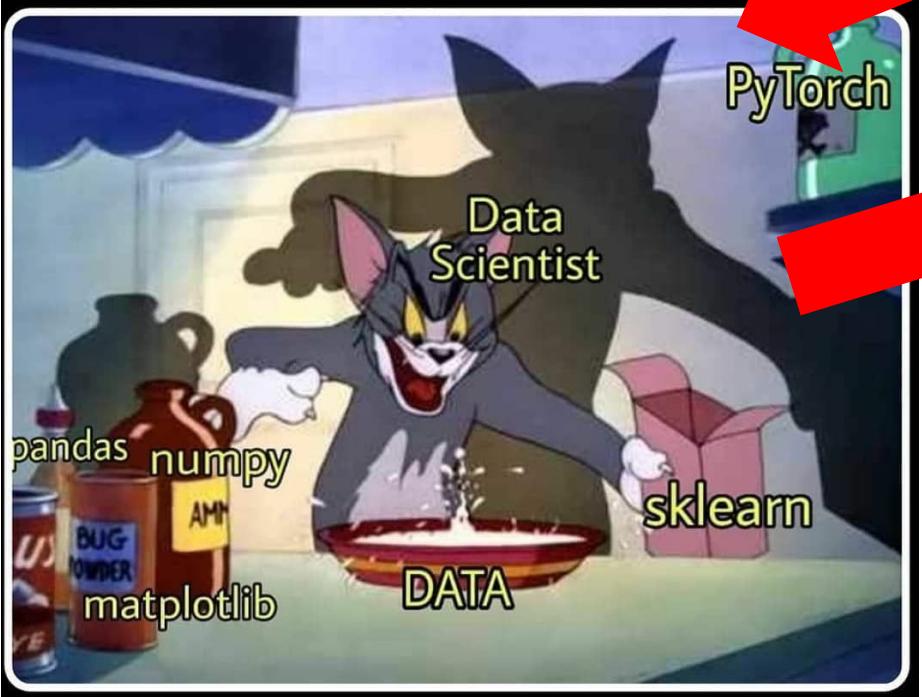
YOU

HP: 100 / 100

DIRTY DATA YOU  
HAVE

CLEAN IT UP YOU MUST

memegenerator.net



DATASET

HP: 100 / 100



A wild DATASET  
appeared!

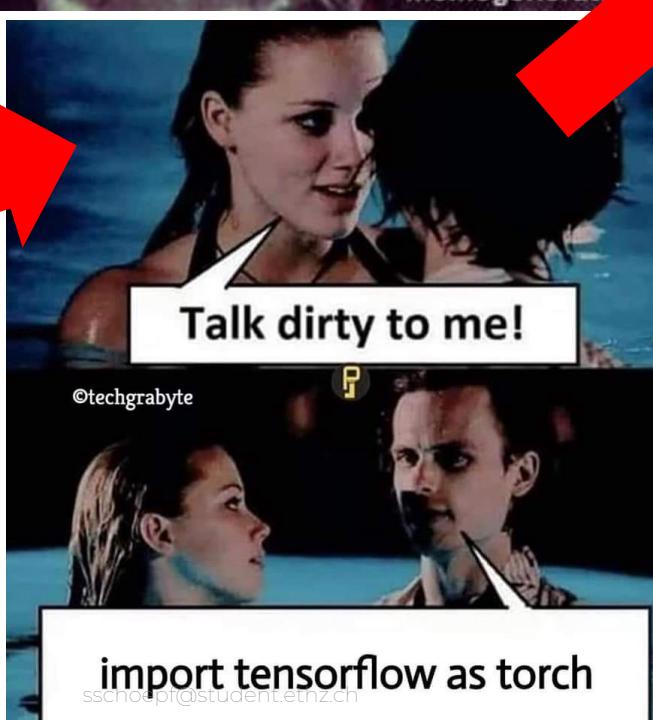
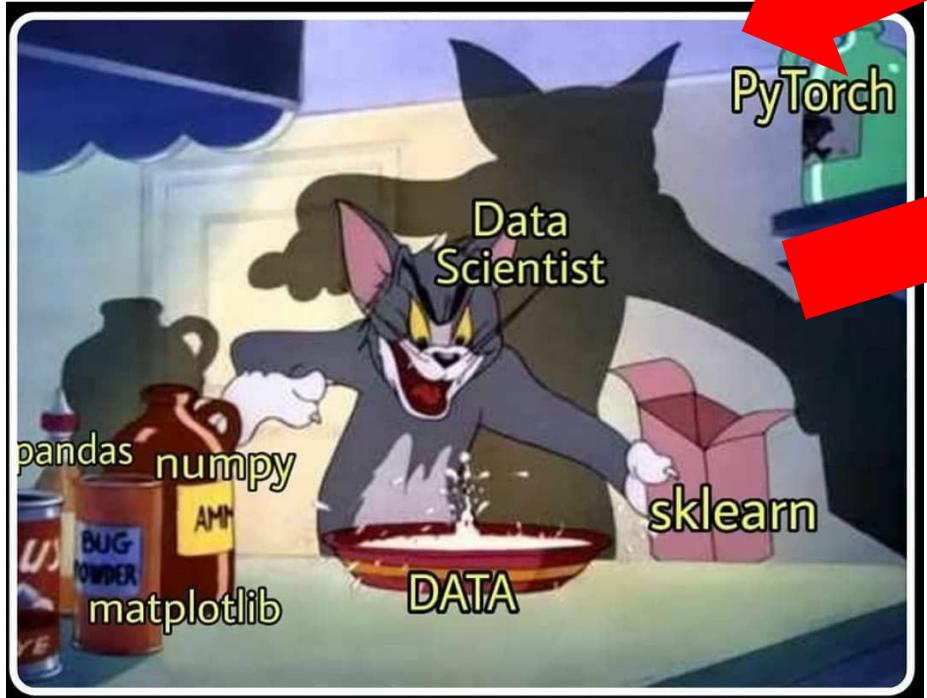
YOU

HP: 100 / 100

DIRTY DATA YOU  
HAVE

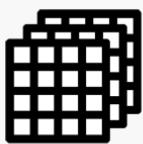
CLEAN IT UP YOU MUST

memegenerator.net



DATASET

HP: 100 / 100



A wild DATASET  
appeared!

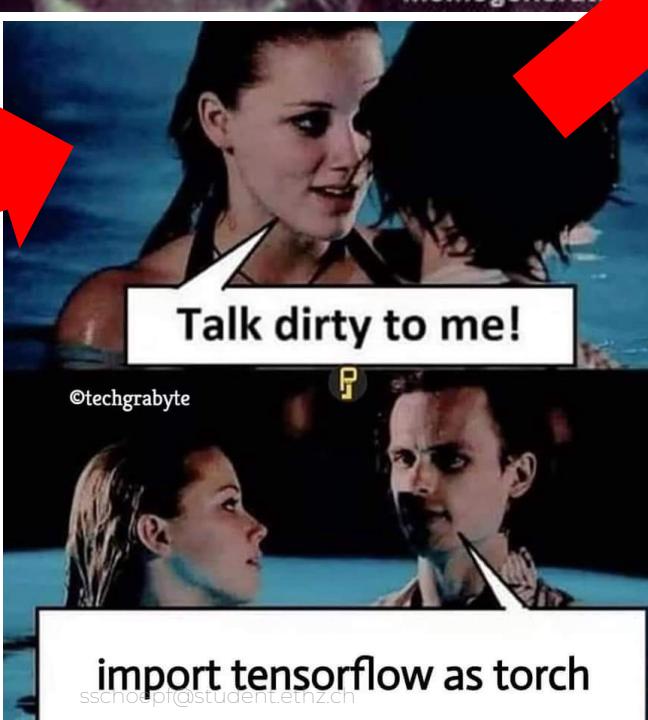
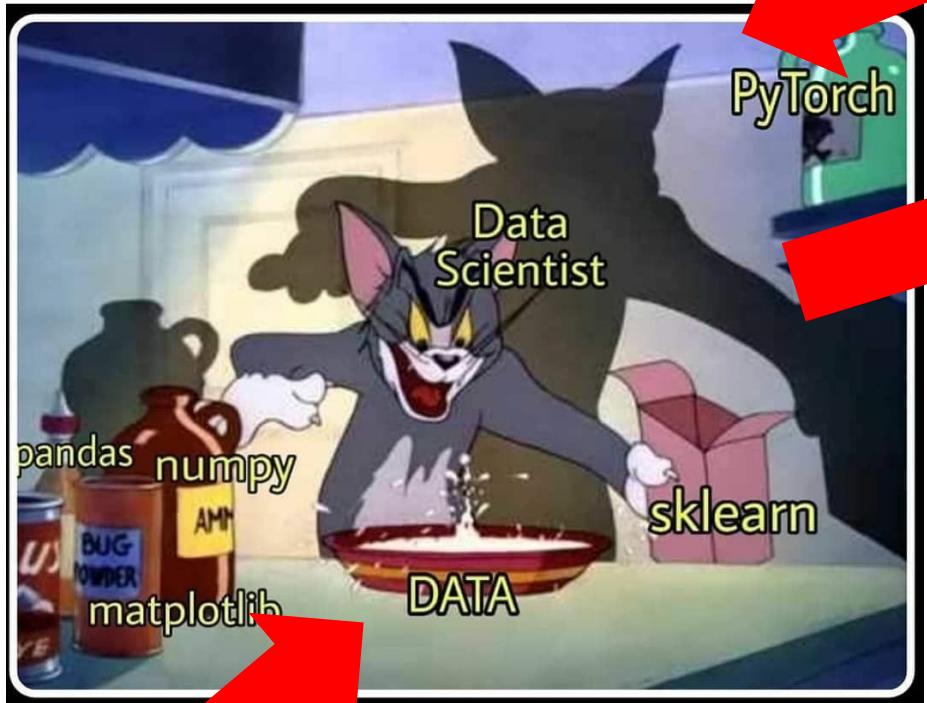
YOU

HP: 100 / 100

DIRTY DATA YOU  
HAVE

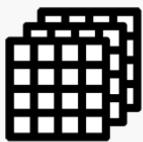
CLEAN IT UP YOU MUST

memegenerator.net



DATASET

HP: 100 / 100



YOU

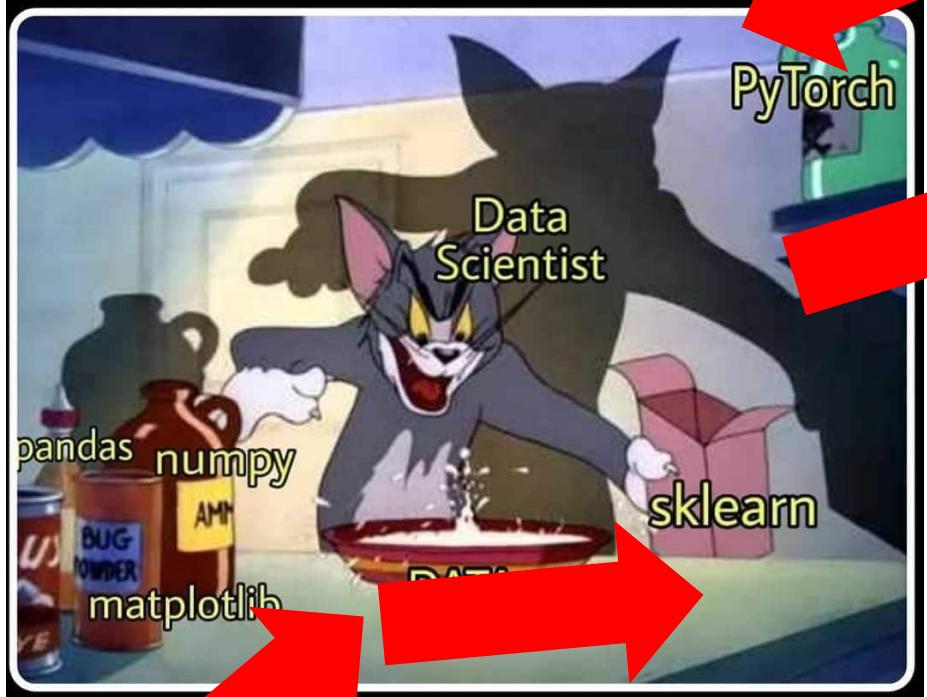
HP: 100 / 100

A wild DATASET  
appeared!

DIRTY DATA YOU  
HAVE

CLEAN IT UP YOU MUST

memegenerator.net



DATASET

HP: 100 / 100



YOU

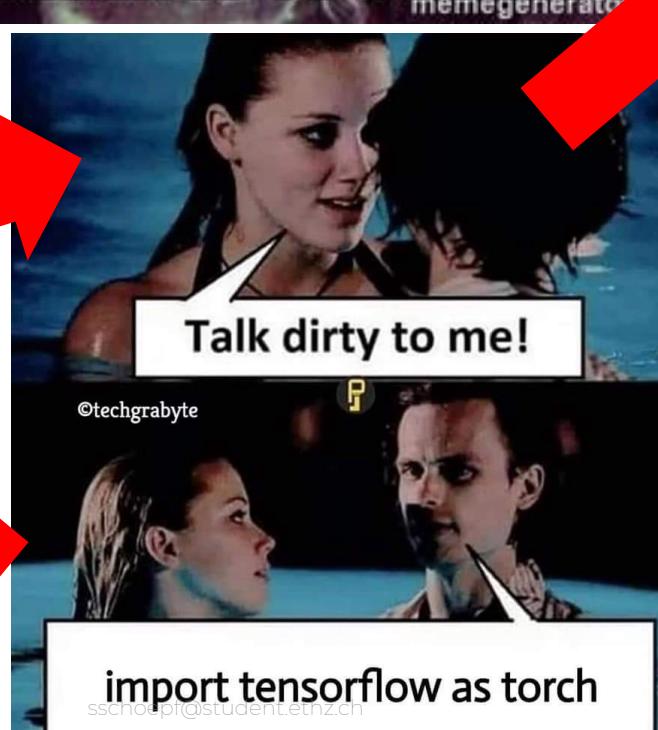
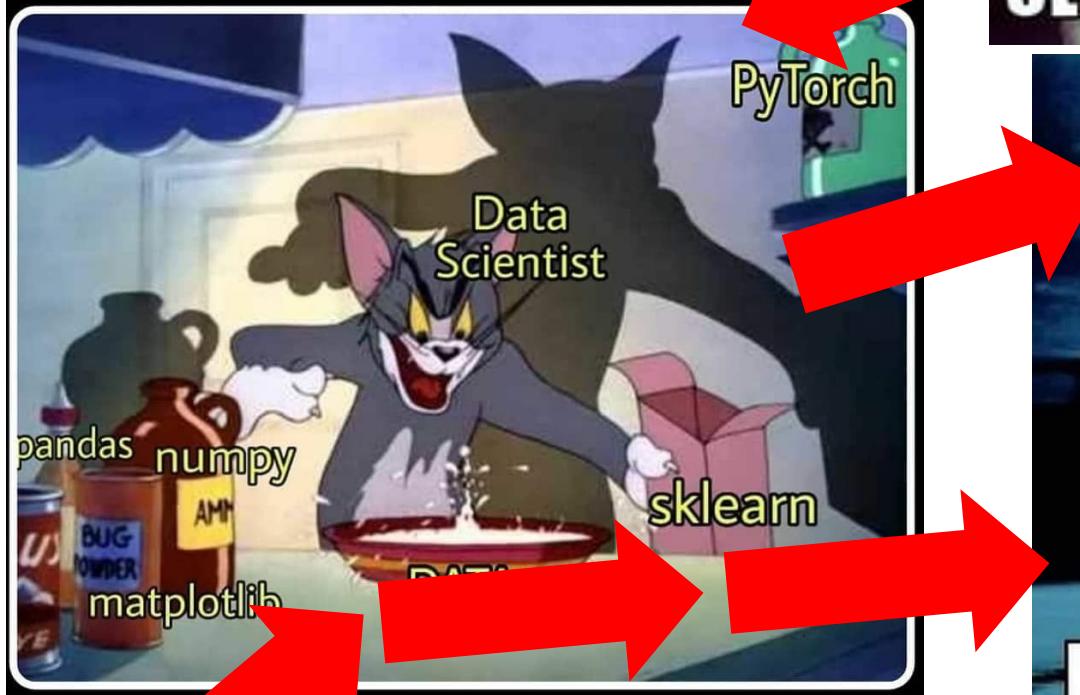
HP: 100 / 100

A wild DATASET  
appeared!

DIRTY DATA YOU  
HAVE

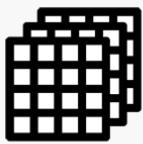
CLEAN IT UP YOU MUST

memegenerator.net



DATASET

HP: 100 / 100



YOU

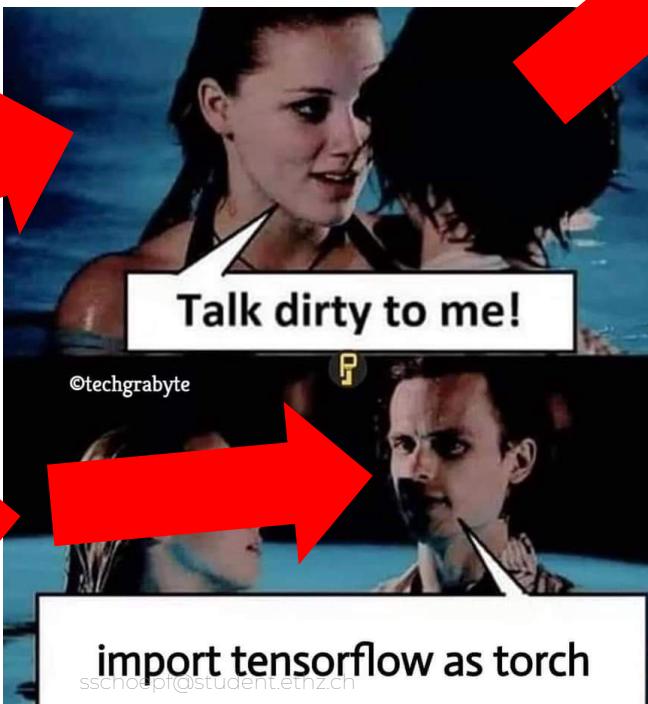
HP: 100 / 100

A wild DATASET  
appeared!

DIRTY DATA YOU  
HAVE

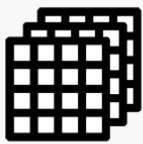
CLEAN IT UP YOU MUST

memegenerator



DATASET

HP: 100 / 100



YOU

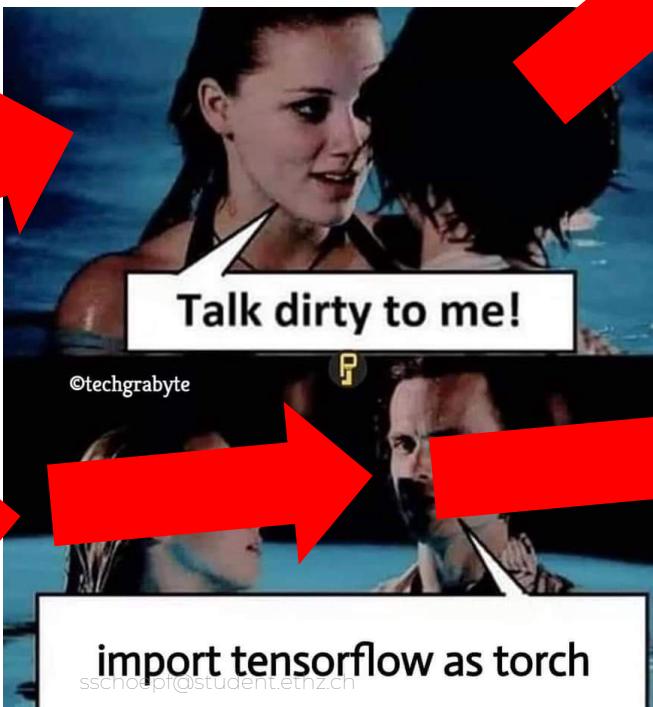
HP: 100 / 100

A wild DATASET  
appeared!

DIRTY DATA YOU  
HAVE

CLEAN IT UP YOU MUST

memegenerator.net



DATASET

HP: 100 / 100



YOU

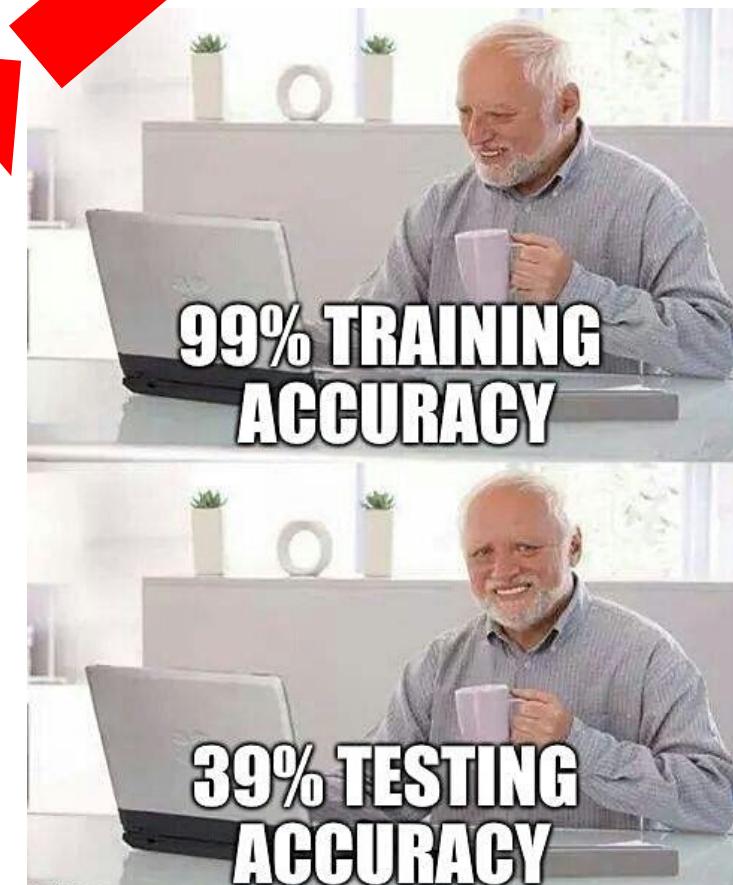
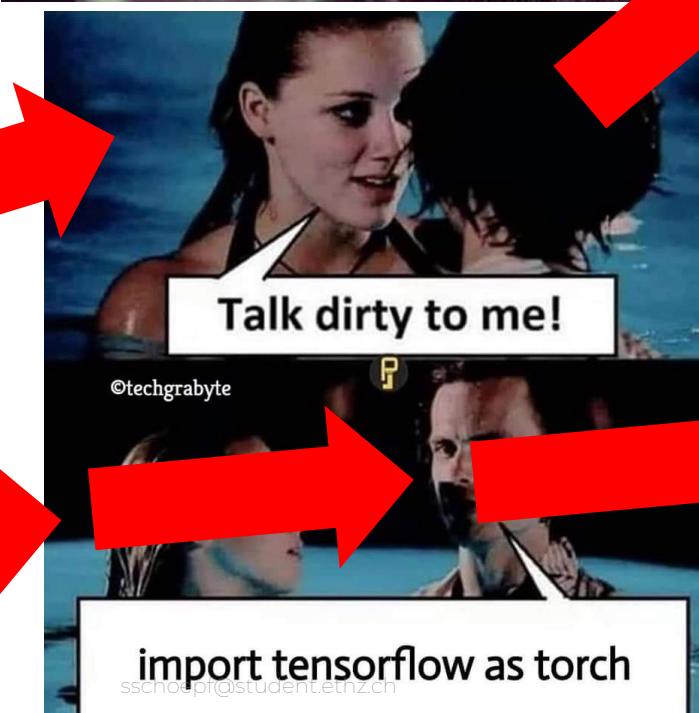
HP: 100 / 100

A wild DATASET  
appeared!

DIRTY DATA YOU  
HAVE

CLEAN IT UP YOU MUST

memegenerator.net



# Summary

# Summary

Me: \*uses machine learning\*

Machine: \*learns\*

Me:



# Q&A

## Part I – Frameworks

# Data Science

## Part II – Linear Regression

# Introduction

# Introduction

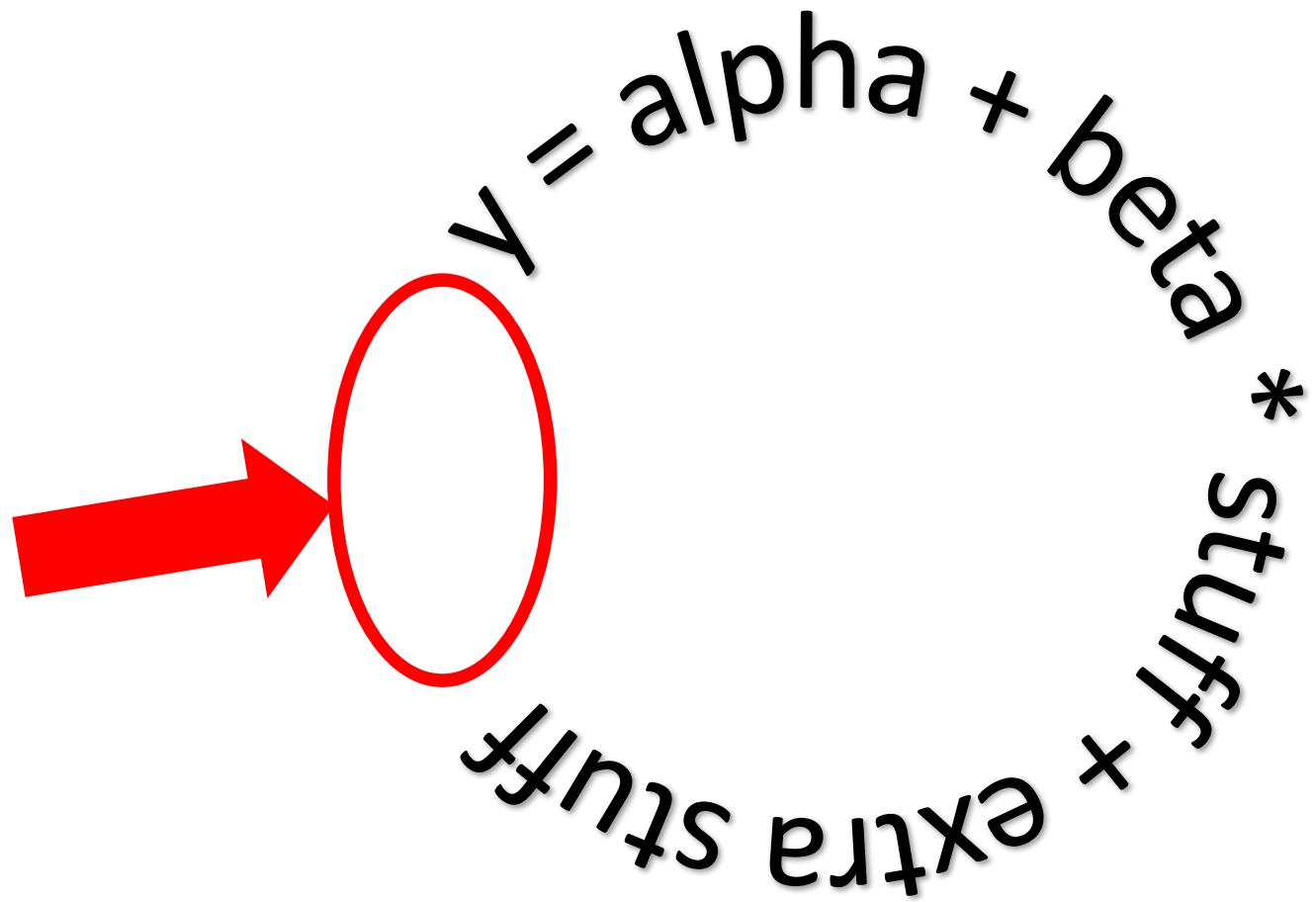
**y = alpha + beta \* stuff + extra stuff**

# Transformation Gap

# Transformation Gap

$y = \alpha + \beta * stuff + extra stuff$

# Transformation Gap



# Transformation Fill

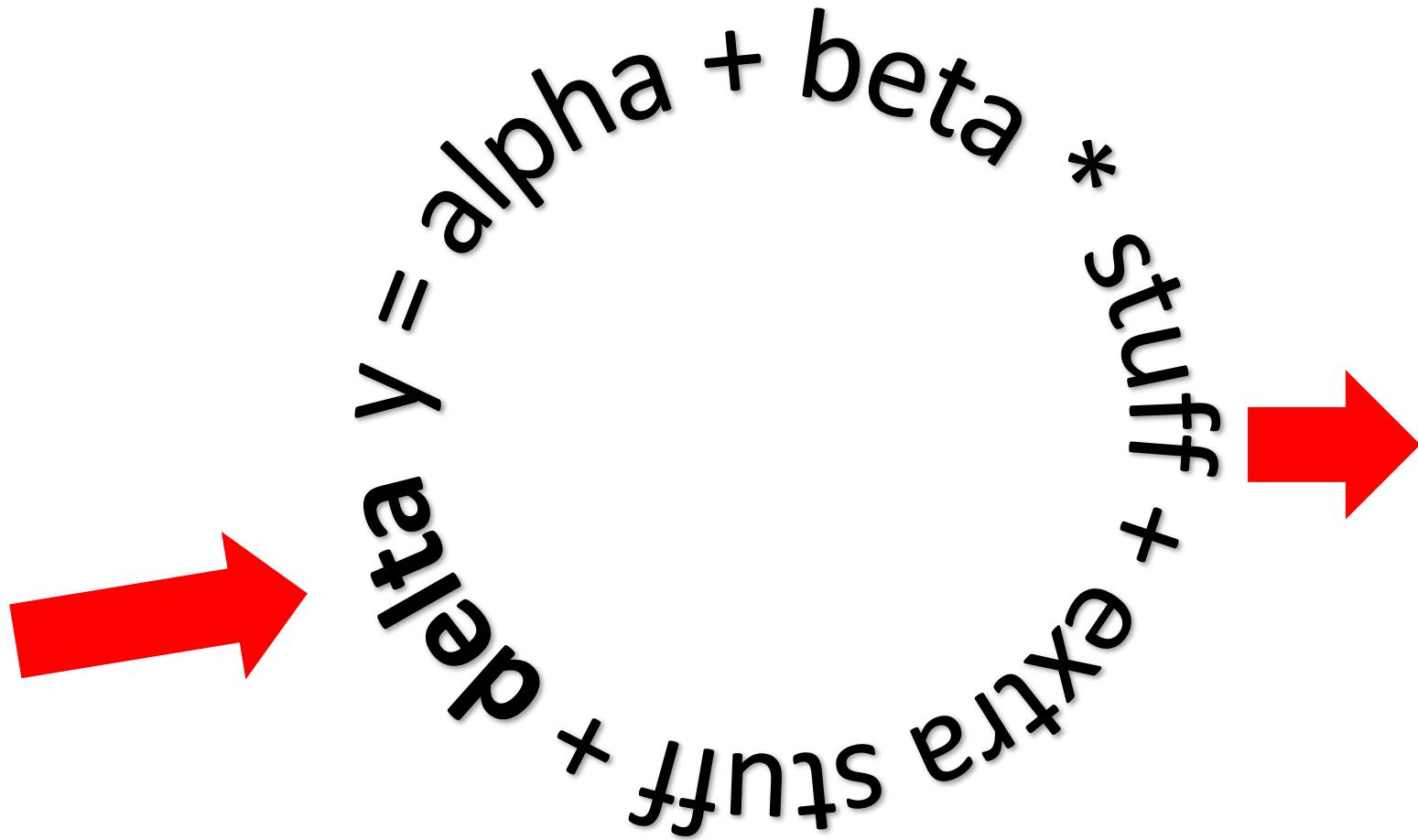
# Transformation Fill

delta y = alpha + beta \* stuff + extra stuff



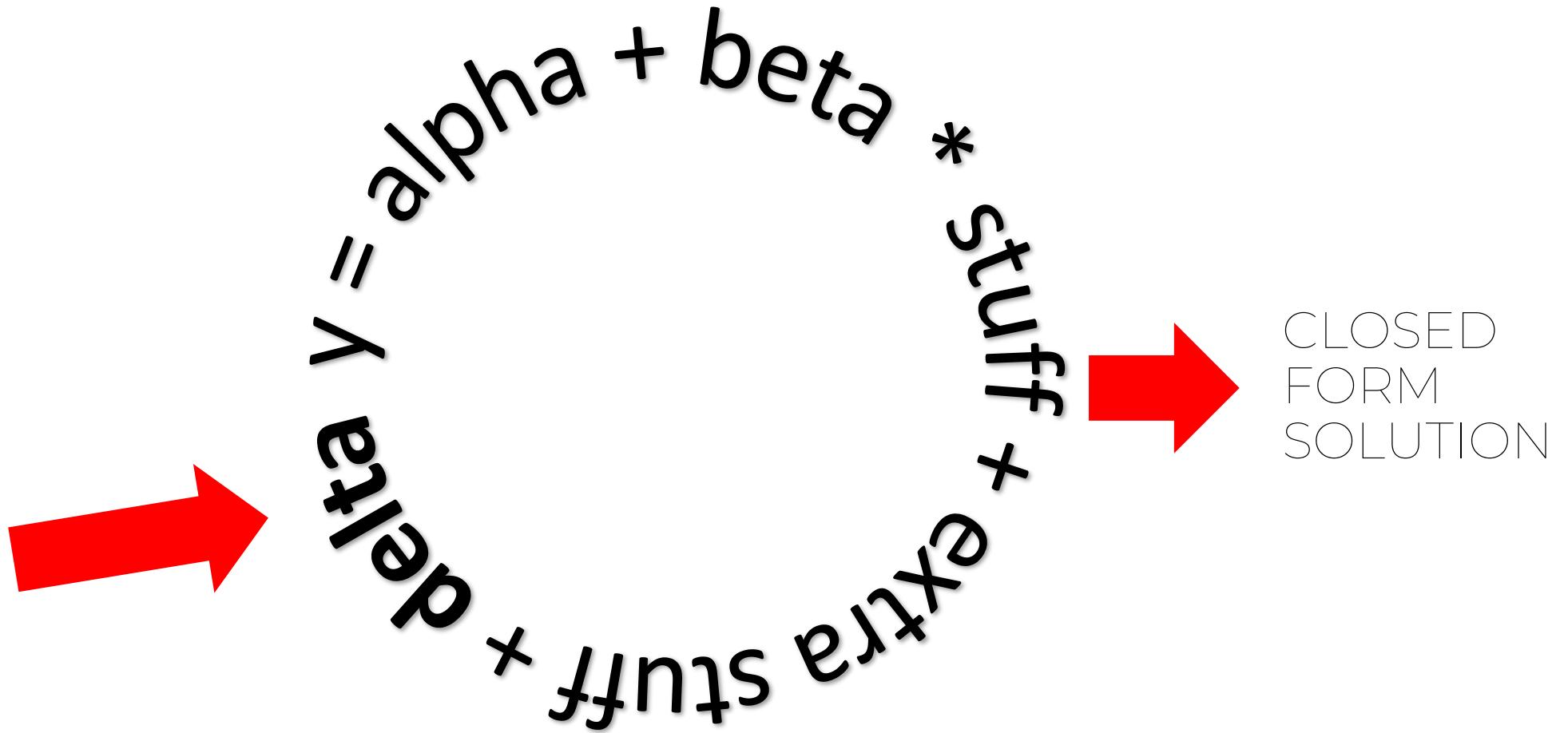
# Transformation Fill

$\delta y = \alpha + \beta * \text{stuff} + \text{extra stuff}$

A large black equation is displayed diagonally across the slide. It consists of the symbol  $\delta y$ , followed by an equals sign, and then a sum of three terms:  $\alpha$ ,  $\beta$  multiplied by the word "stuff", and the phrase "extra stuff". Two large red arrows point from the bottom left towards the words "stuff" and "extra stuff".

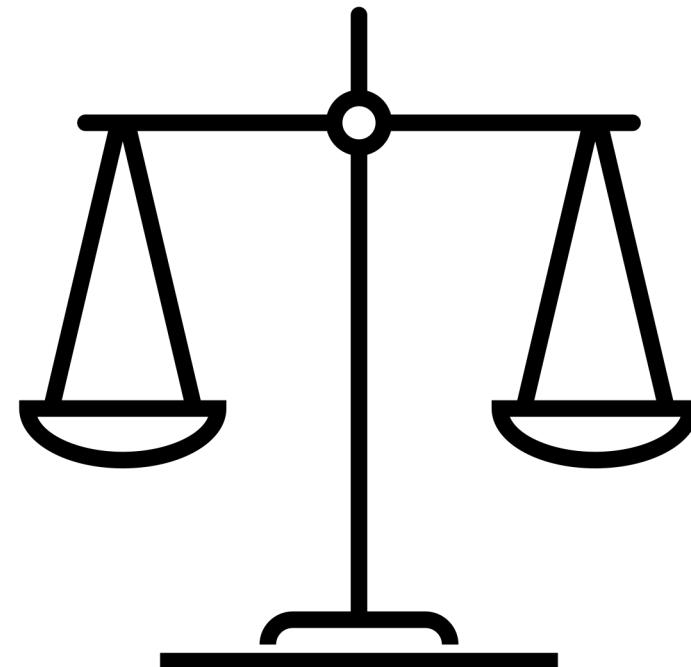
# Transformation Fill

$\delta y = \alpha + \beta * \text{stuff} + \text{extra stuff}$

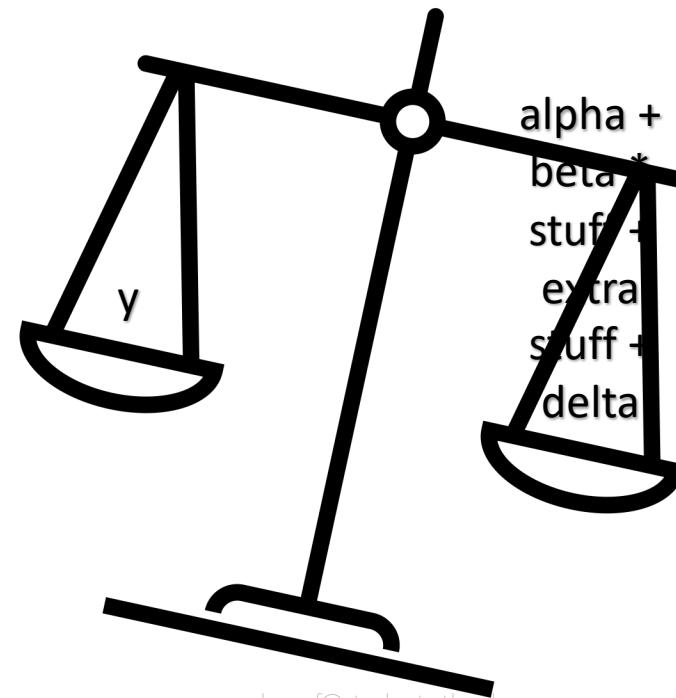


CLOSED  
FORM  
SOLUTION

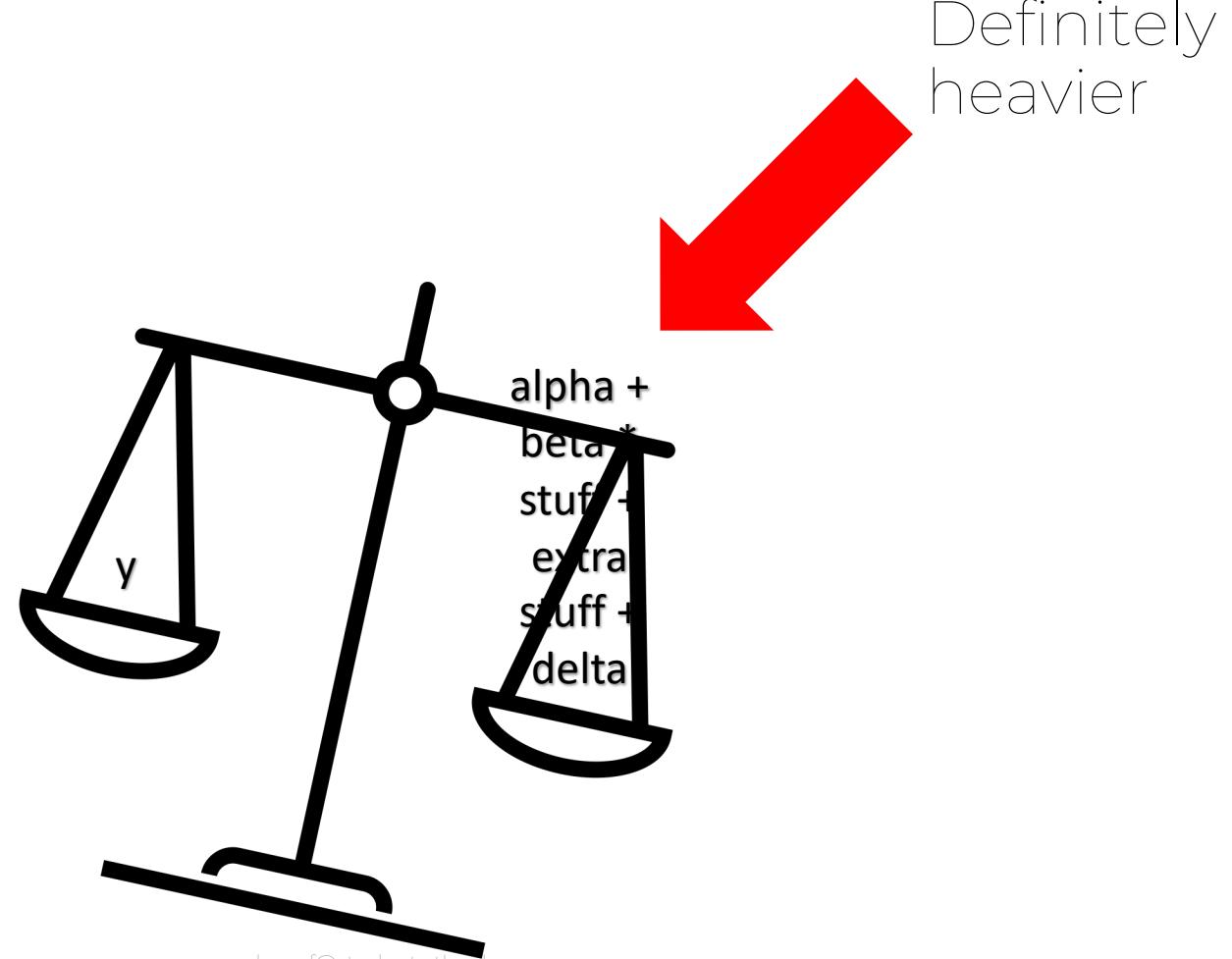
# Is our solution biased?

$$y = \text{alpha} + \text{beta} * \text{stuff} + \text{extra stuff} + \text{delta}$$


# Is our solution biased?



# Is our solution biased?



Before we end, let's look at some relevant numbers:

Before we end, let's look at some relevant numbers:

**5**

Before we end, let's look at some relevant numbers:

**5**

**7**

Before we end, let's look at some relevant numbers:

**5**

**42**

**7**

Before we end, let's look at some relevant numbers:

**5**

**42**

**3.13**

**7**

# Q&A

## Part II – Linear Regression

# Data Science

## Part III – Trees

# What are trees anyway?

# What are trees anyway?

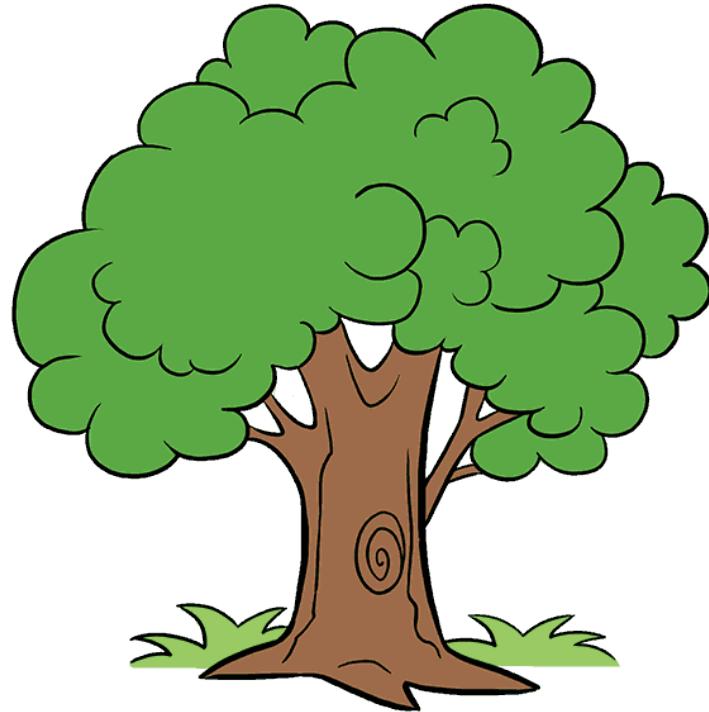


# What are trees anyway?



In botany, a tree is a perennial plant with an elongated stem, or trunk, supporting branches and leaves in most species. In some usages, the definition of a tree may be narrower, including only woody plants with secondary growth, plants that are usable as lumber or plants above a specified height. In wider definitions, the taller palms, tree ferns, bananas, and bamboos are also trees.

# What are trees anyway?



In botany, a tree is a perennial plant with an elongated stem, or trunk, supporting branches and leaves in most species. In some usages, the definition of a tree may be narrower, including only woody plants with secondary growth, plants that are usable as lumber or plants above a specified height. In wider definitions, the taller palms, tree ferns, bananas, and bamboos are also trees.

- Wikipedia

# What are trees anyway?



In botany, a tree is a perennial plant with an elongated stem, or trunk, supporting branches and leaves in most species. In some usages, the definition of a tree may be narrower, including only woody plants with secondary growth, plants that are usable as lumber or plants above a specified height. In wider definitions, the taller palms, tree ferns, bananas, and bamboos are also trees.

- Wikipedia
  - (Internet)

# Why do we care about trees?

# Why do we care about trees?



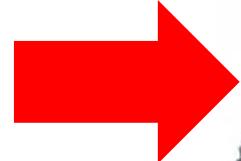
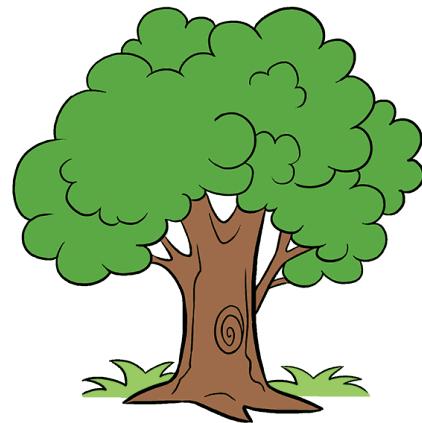
# Why do we care about trees?



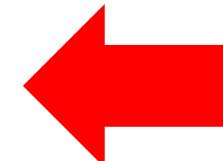
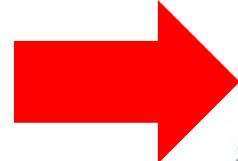
# Why do we care about trees?



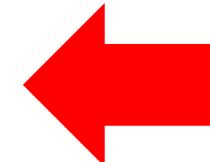
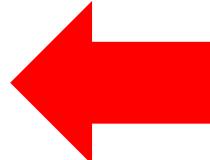
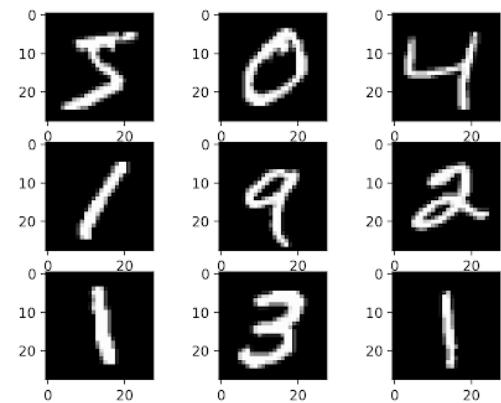
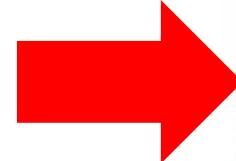
# Why do we care about trees?



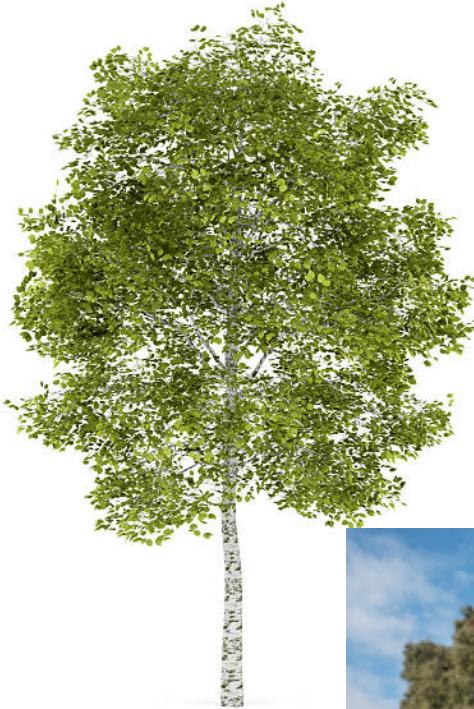
# Why do we care about trees?



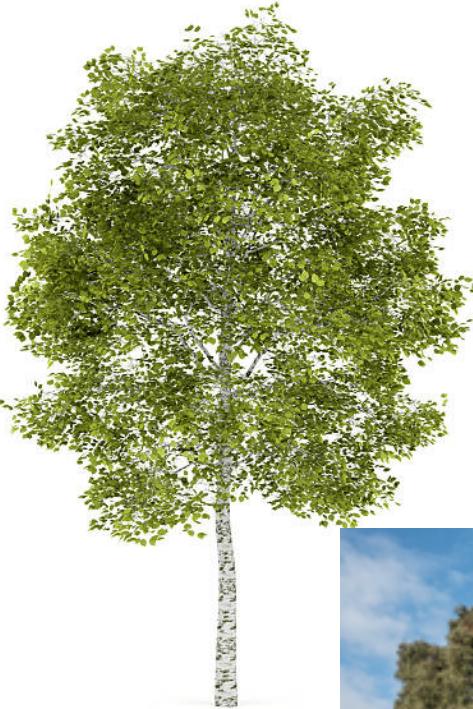
# Why do we care about trees?



# What is this? (Audience Question)

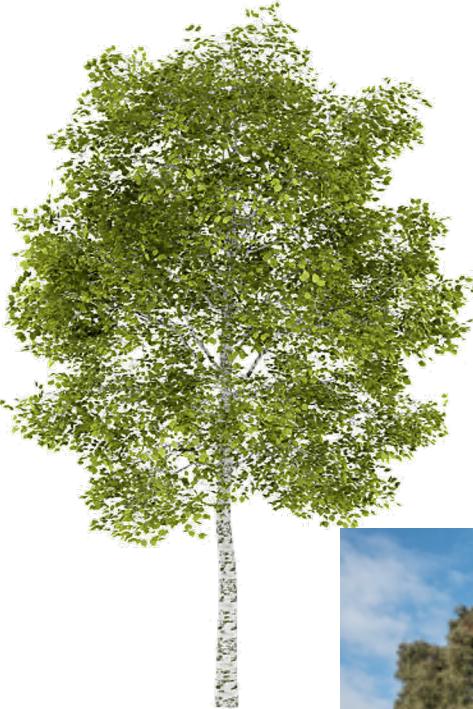


# What is this? (Audience Question)



**A random forest**

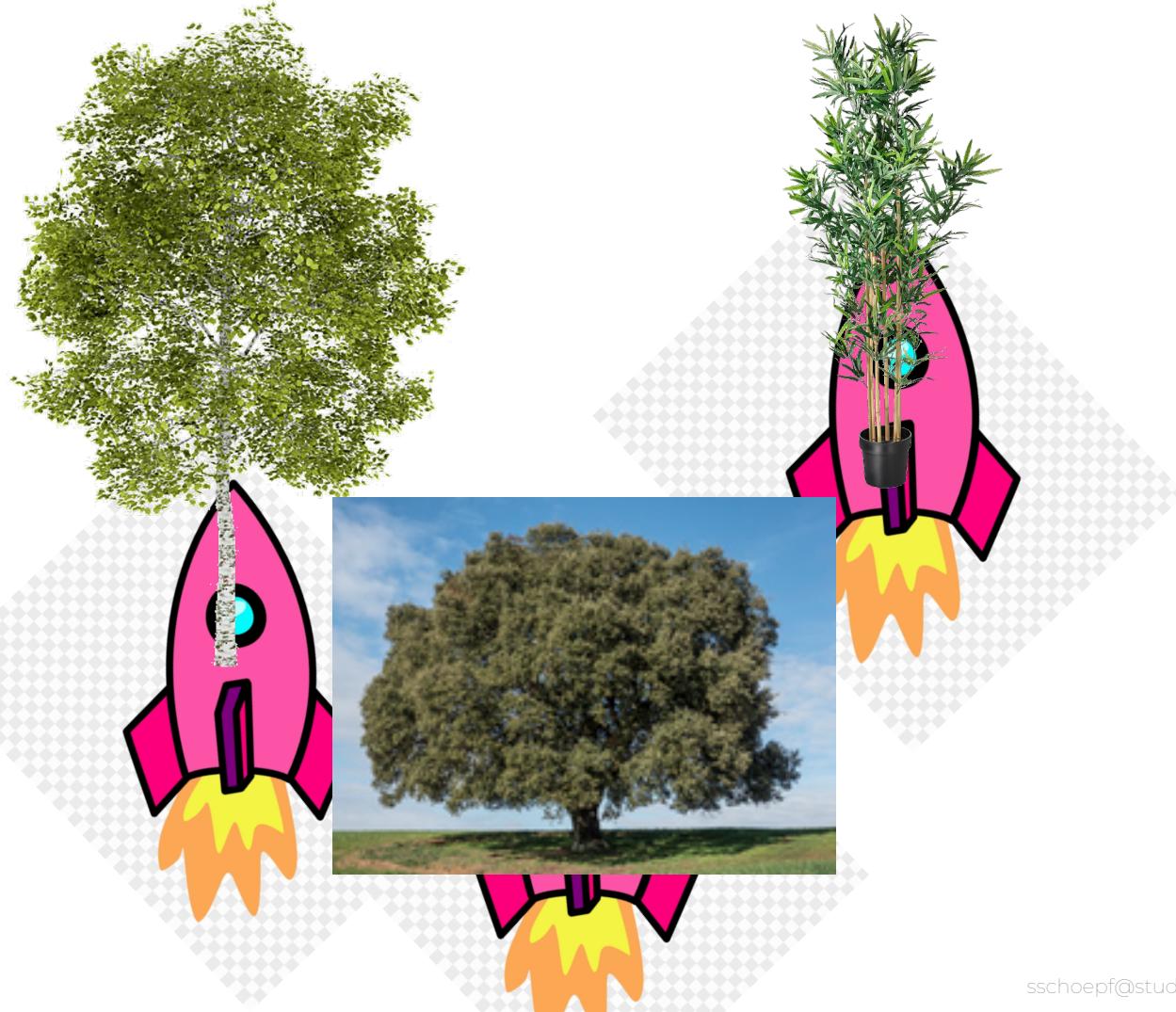
# How can we improve it?



# How can we improve it?



# How can we improve it?



## Boosting!!!

# How can we make the mathematicians happy?

# How can we make the mathematicians happy?



# How can we make the mathematicians happy?



# How can we make the mathematicians happy?



# Bootstrapping

# Q&A

Part III – Trees

# Data Science

## Part IV – Deep Learning

# What to expect from this presentation

# What to expect from this presentation



# What to expect from this presentation



# What to expect from this presentation

- No memes!



# What to expect from this presentation



- No memes!
- We are here to do real data science

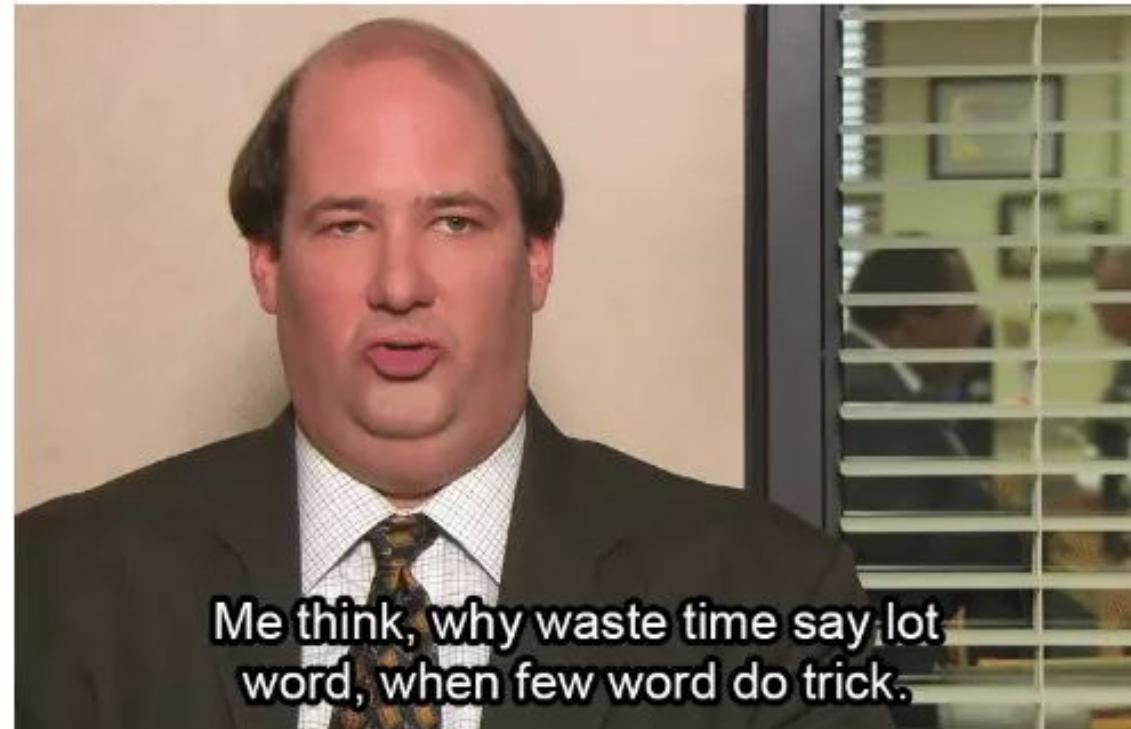
# What to expect from this presentation



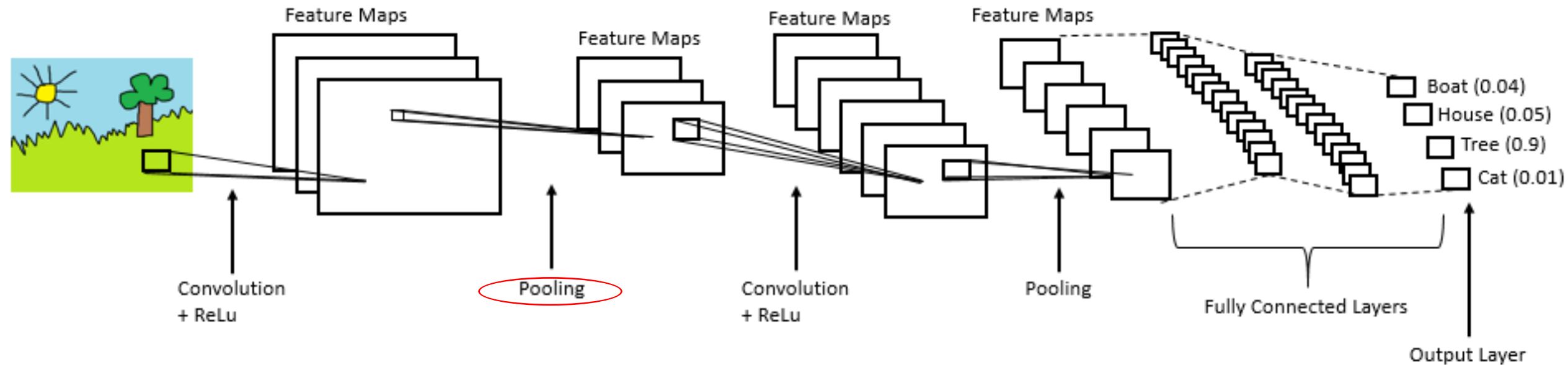
- No memes!
- We are here to do real data science
- Not like these people that presented before us

Ok, one more because it is really good

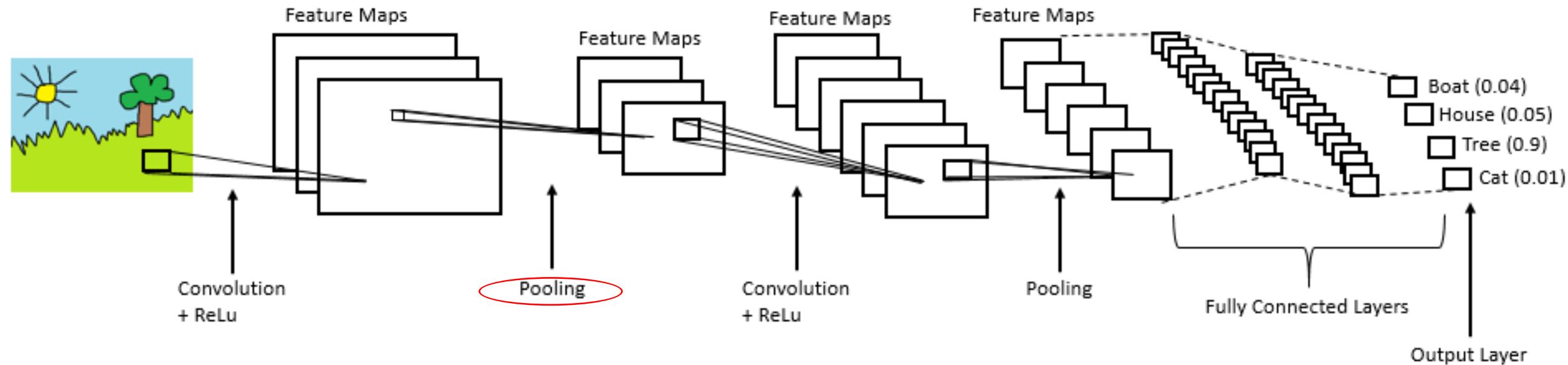
**When you penalize your Natural Language Generation model for large sentence lengths**



# CNNs

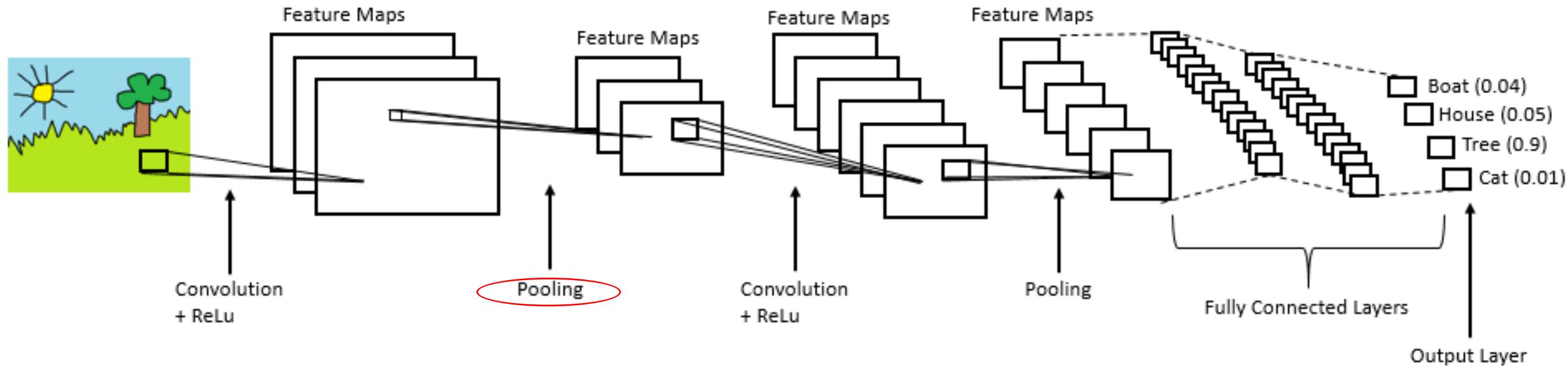


# CNNs



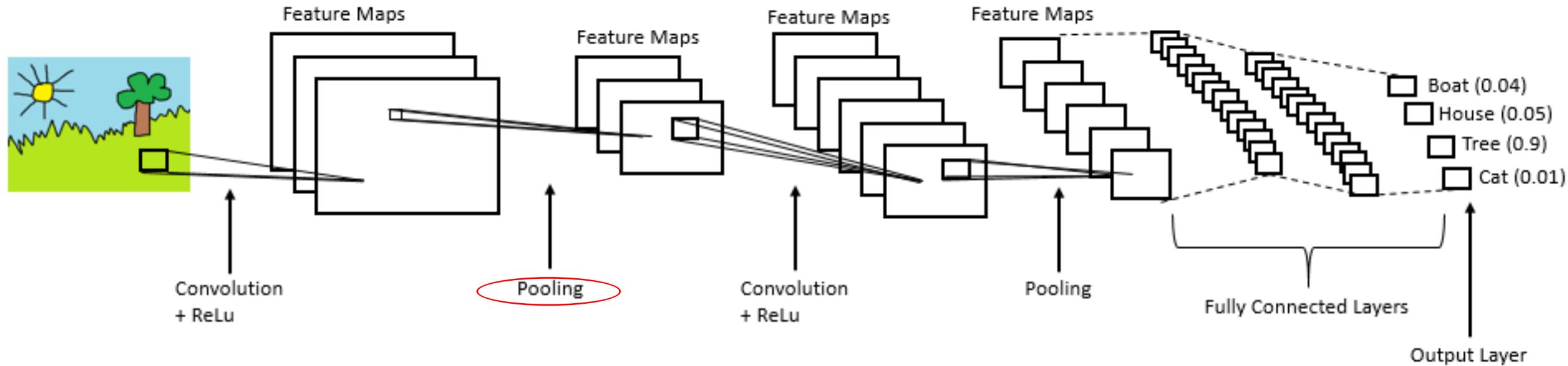
- Pooling is an area with relatively few innovations

# CNNs



- Pooling is an area with relatively few innovations
- Mostly just boring averaging or max()

# CNNs



- Pooling is an area with relatively few innovations
- Mostly just boring averaging or max()
- We have added an exiting new method!

# Entropy enhancing pooling

# Entropy enhancing pooling

- Fancy way to say random pooling

# Entropy enhancing pooling

- Fancy way to say random pooling
- Shows great potential

# Entropy enhancing pooling

- Fancy way to say random pooling
- Shows great potential
  - Solves discrimination in models -> all individuals treated equally

# Entropy enhancing pooling

- Fancy way to say random pooling
- Shows great potential
  - Solves discrimination in models -> all individuals treated equally
  - Training becomes more exciting and unpredictable

# Entropy enhancing pooling

- Fancy way to say random pooling
- Shows great potential
  - Solves discrimination in models -> all individuals treated equally
  - Training becomes more exciting and unpredictable
  - Entropy sounds great in business presentations

# Established pooling methods

Can you name the two most widely used ones? (Audience)

# Established pooling methods

Can you name the two most widely used ones? (Audience)



# Established pooling methods

Can you name the two most widely used ones? (Audience)

- Carpooling



# Established pooling methods

Can you name the two most widely used ones? (Audience)

- Carpooling



# Established pooling methods

Can you name the two most widely used ones? (Audience)

- Carpooling
- Respooling



# Q&A

Part IV – Deep Learning

# Homework

Insert one of the learned concepts in your next real meeting  
without getting caught