

is greater

than theta?

Yes/No

Output

SUM

Score

https://github.com/L-ashwin

Agenda

• Introduction

• The idea of Learning

• Intuition for deep network

• Curse of Dimensionality

• Why we need it?

F2 x Weightage of F2

F3 x Weightage of F3

F4 x Weightage of F4

F5 x Weightage of F5

Natural Signals data

• Locality

• Stationarity

• Compositionality

• Transfer learning

We utilize the properties to

• Improve the architecture

Properties of Natural signals

• 1MP image - 3 million feature points.

• 1sec of audio at standard sampling rate - 44.1 thousand feature points

Intuitions for deep network, Different Representation of Input at each stage

• Why it works?

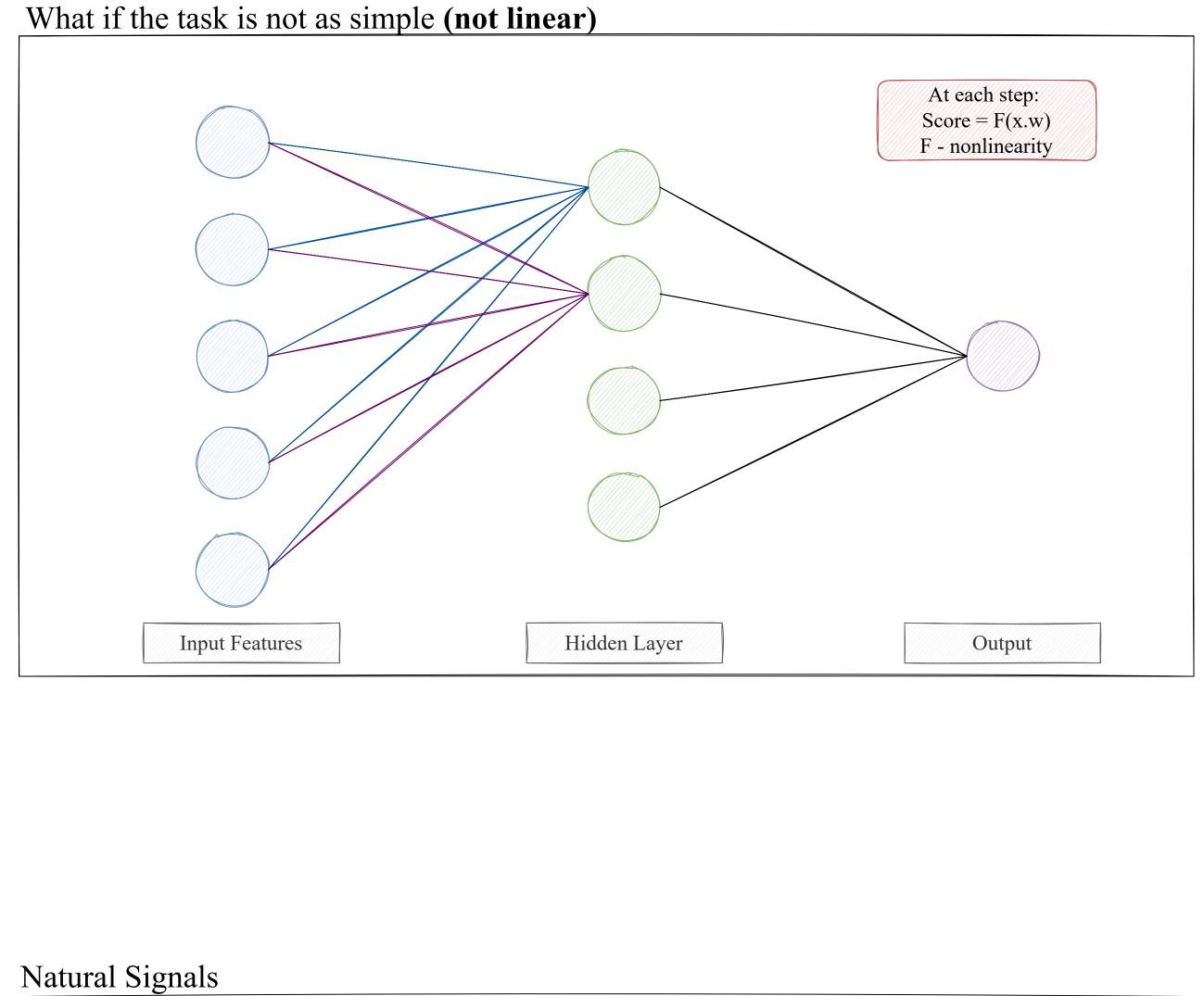
• Blessings of Domain Knowledge

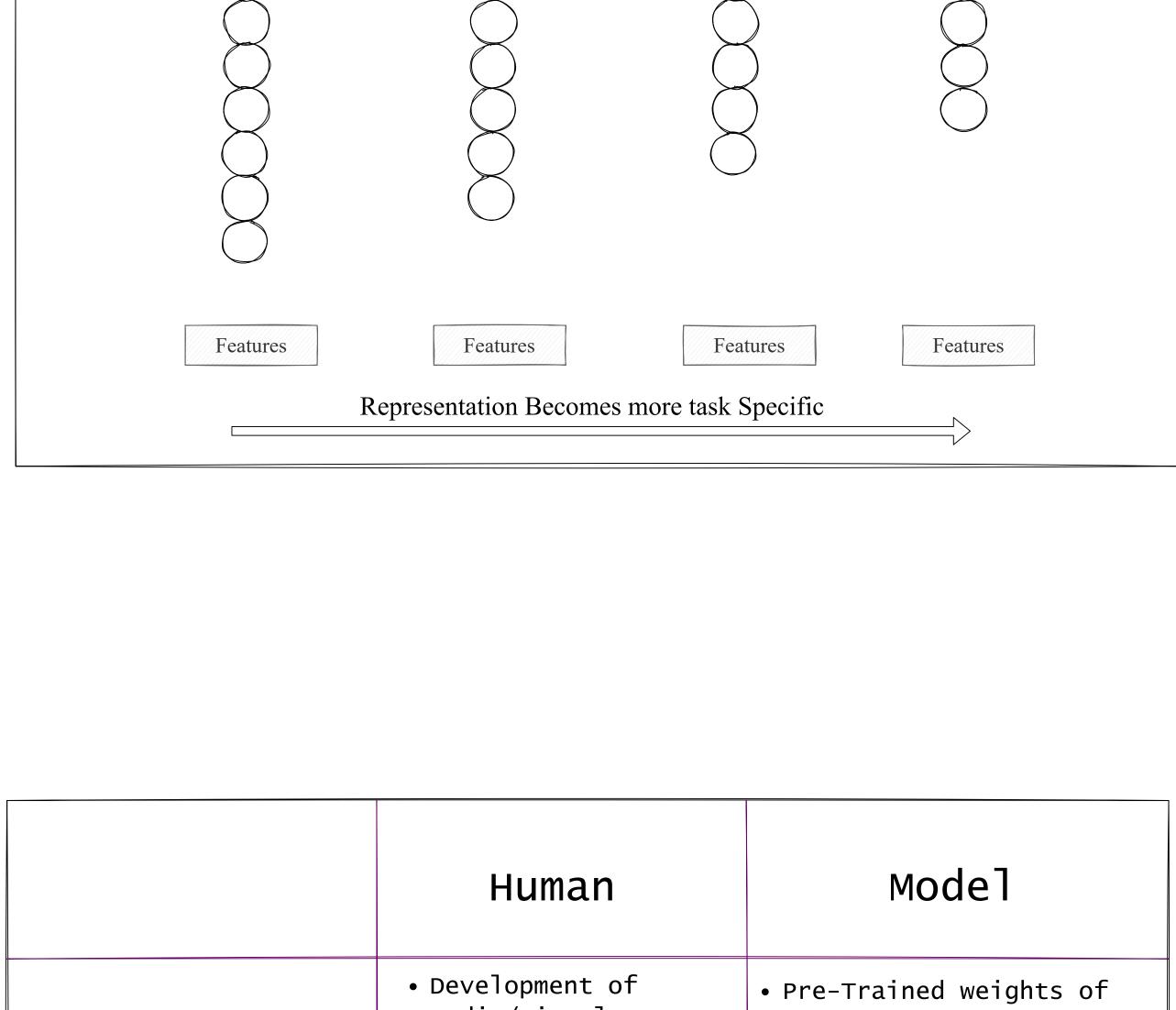
• The Idea of Transfer Learning

• Natural Signals - Visual, Audio & Textual

• How to make it work (Detectron2 - demo)

• Deep Learning





	Human	Model
Learning the Representation	 Development of audio/visual understanding with experience Learning language 	 Pre-Trained weights of models Learning word Embeddings
Learning the Task	• Reasoning, Understanding	• Fine Tuning specific task on top of features.
Task 1 Pre-Training Training all the layers. Model has to learn representation as well as the task. Large number of trainable parameters & requires large number of training data.		

Task 2 Fine-Tuning

task 1 as is and keep them fix.

learned in task 1.

requires less training data.

We use few of the pre-trained layers from

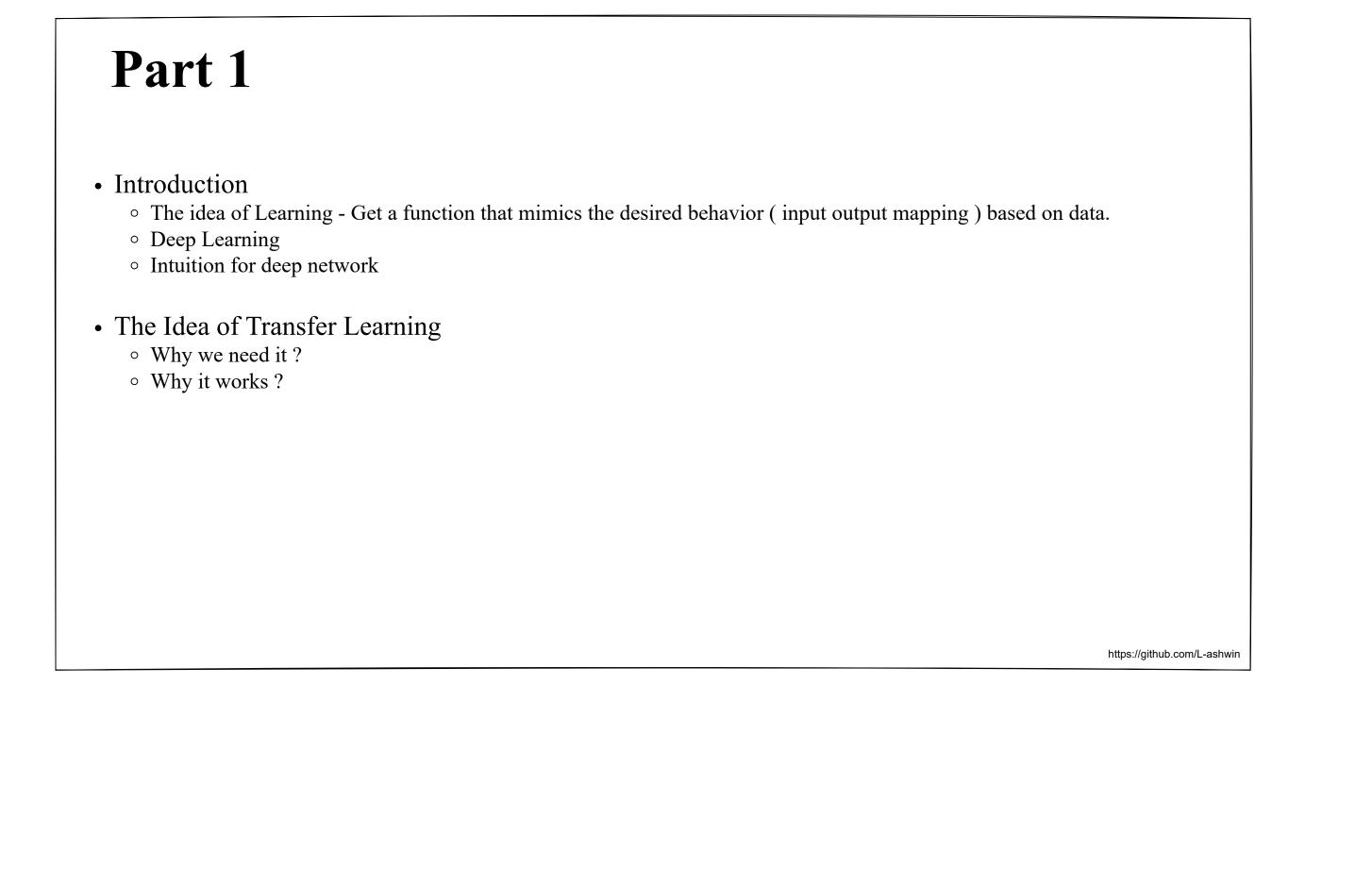
Model has to learn task specific things more while building on representation

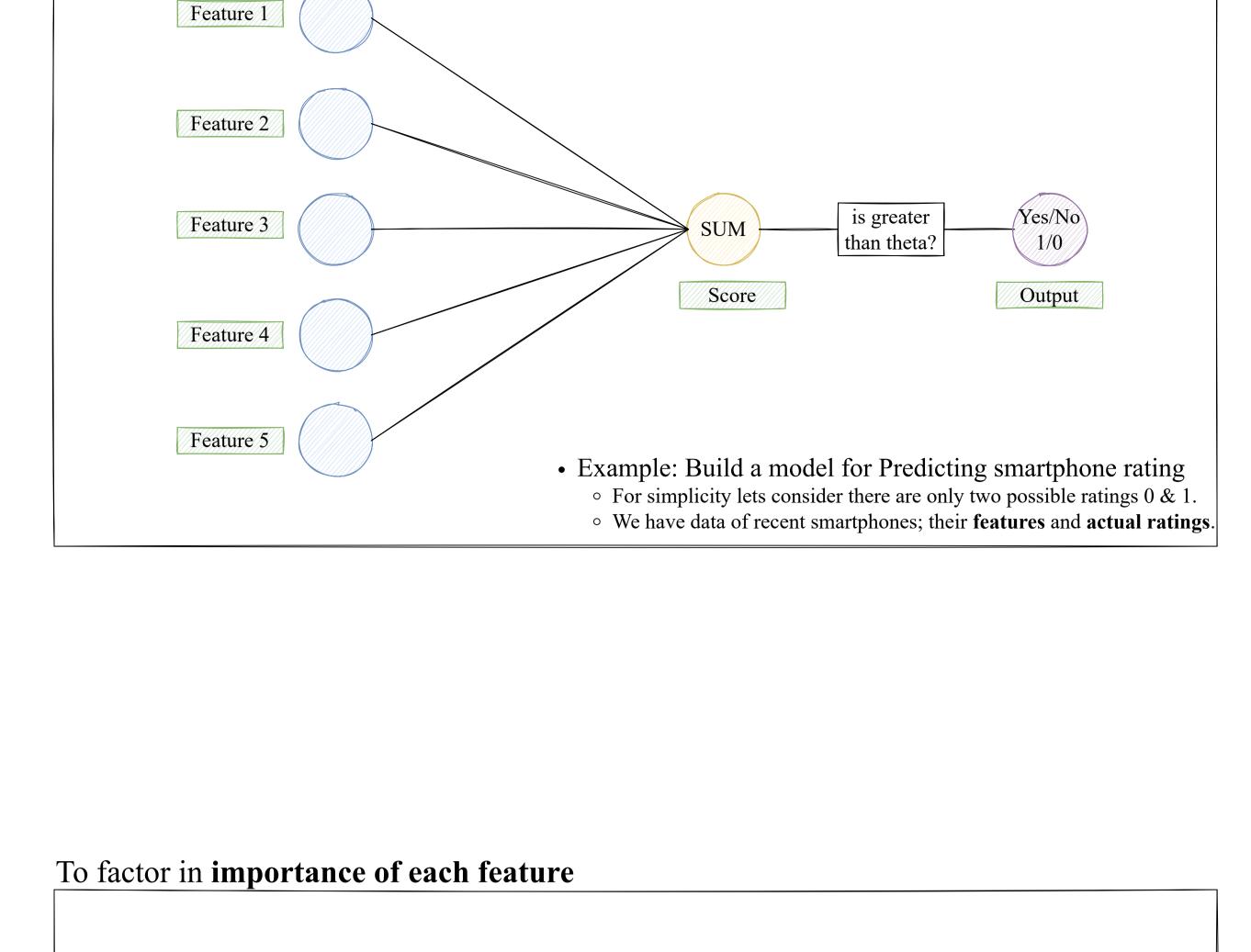
Relatively less trainable parameters &

Backprop

Backprop







Yes/No

Output

Output

is greater

than theta?

SUM

Score

Simplest, yet logical thing to do

F1 x Weightage of F1

F2 x Weightage of F2

F3 x Weightage of F3

F4 x Weightage of F4

F5 x Weightage of F5

Input Features

Task 1 Pre-Training

Model has to learn representation as well

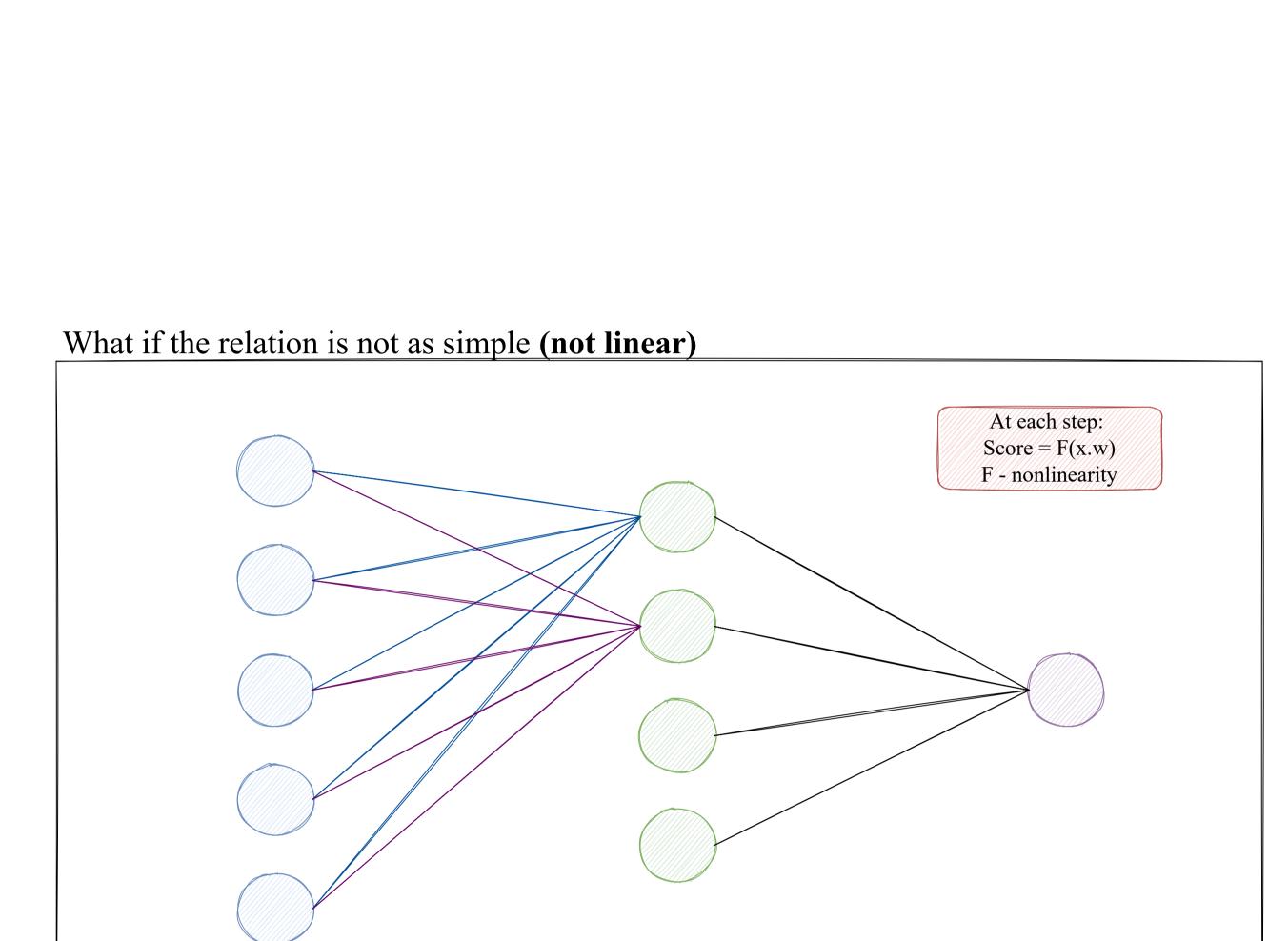
Learning the

Representation

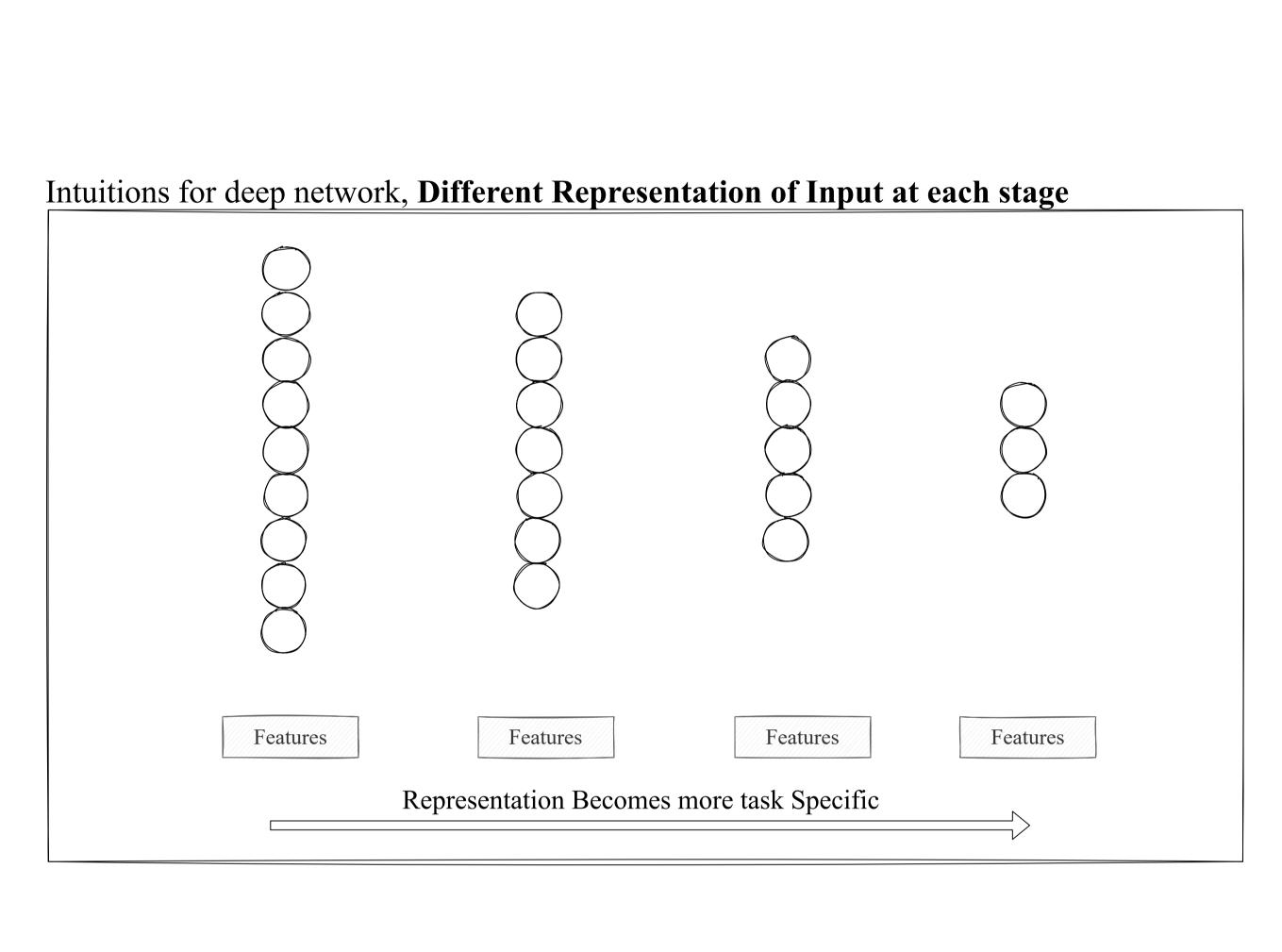
Learning the

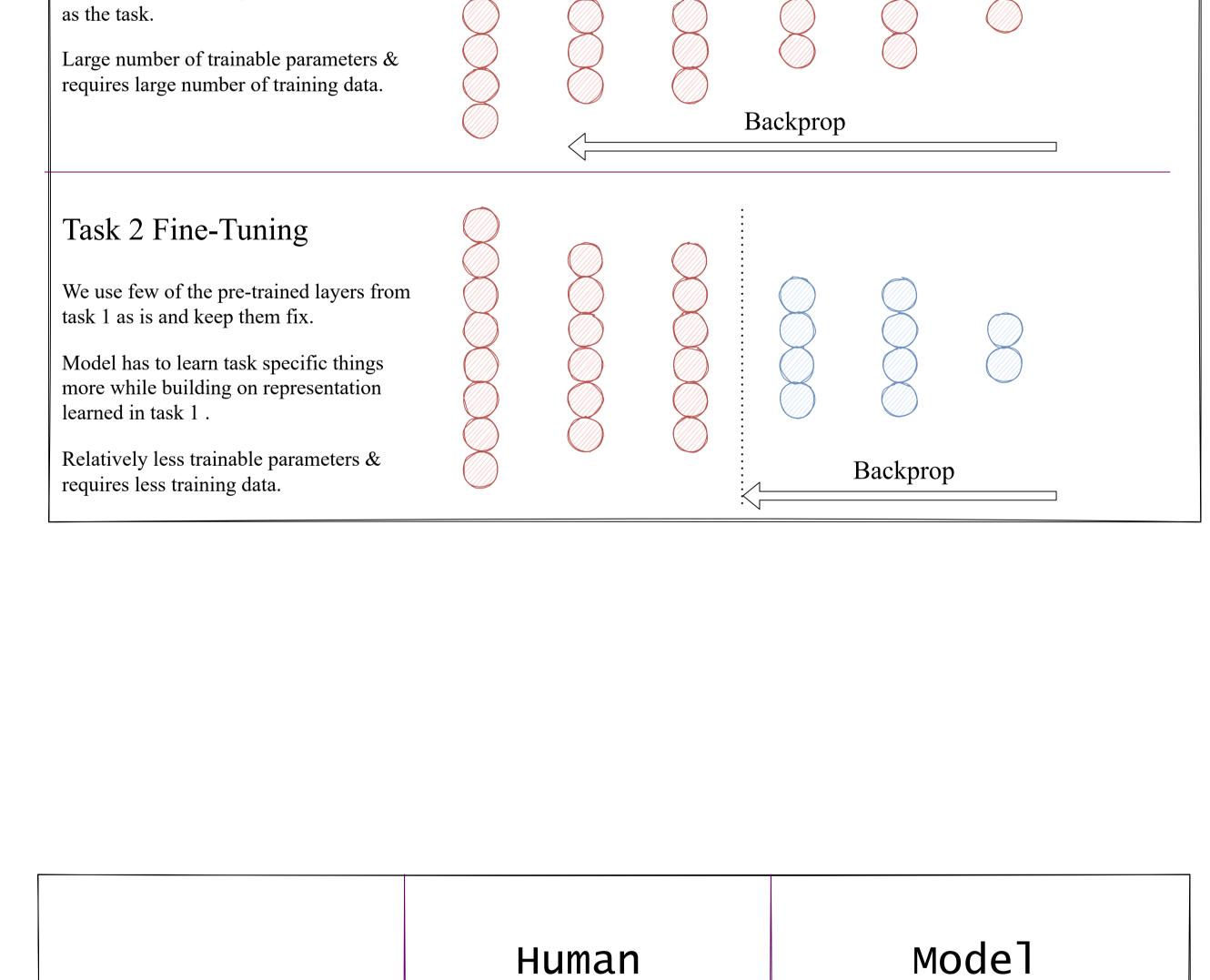
Task

Training all the layers.



Hidden Layer





Development of

audio/visual

experience

Reasoning,

understanding with

Learning language

Understanding...

Pre-Trained weights of

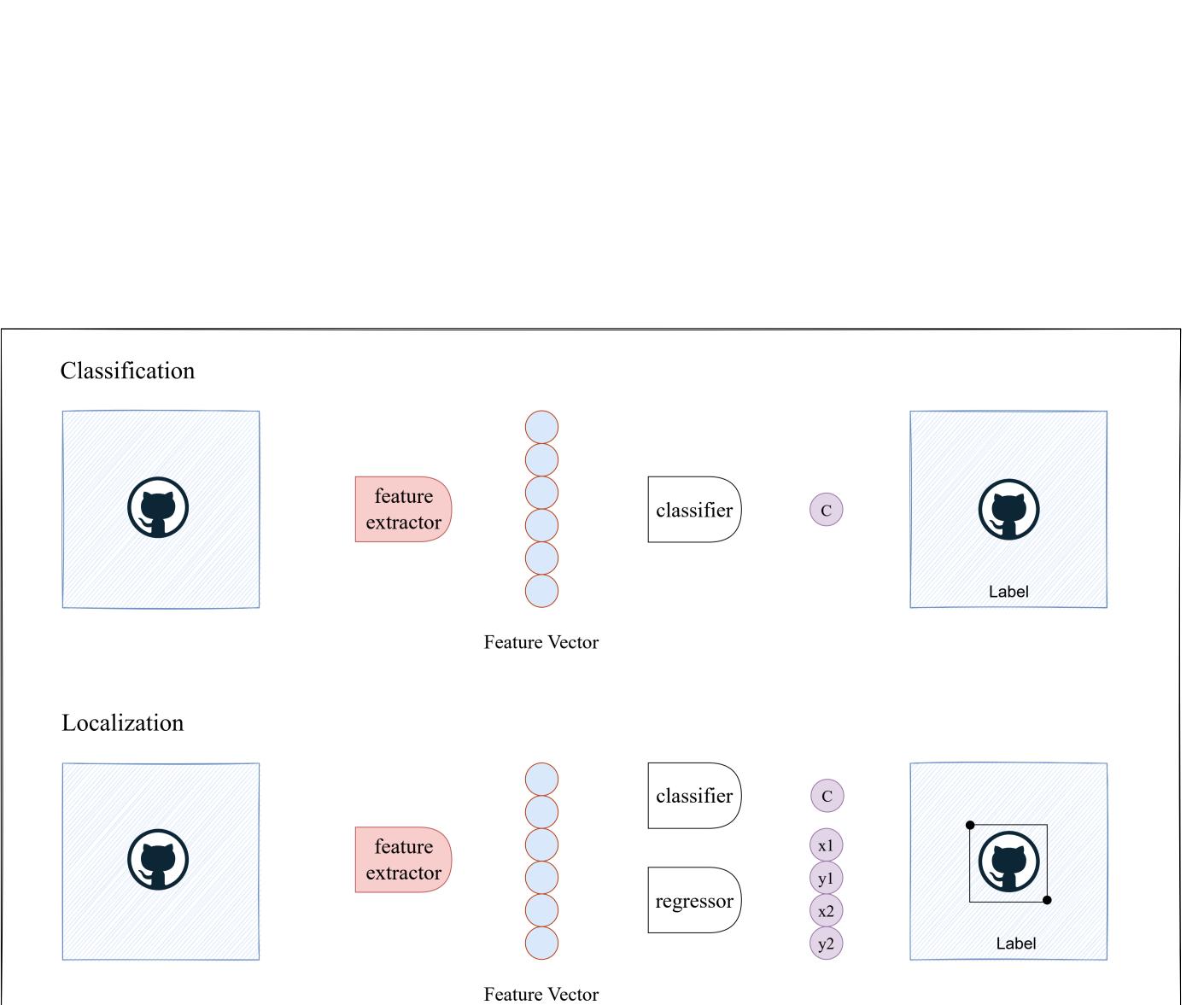
Learning word Embeddings

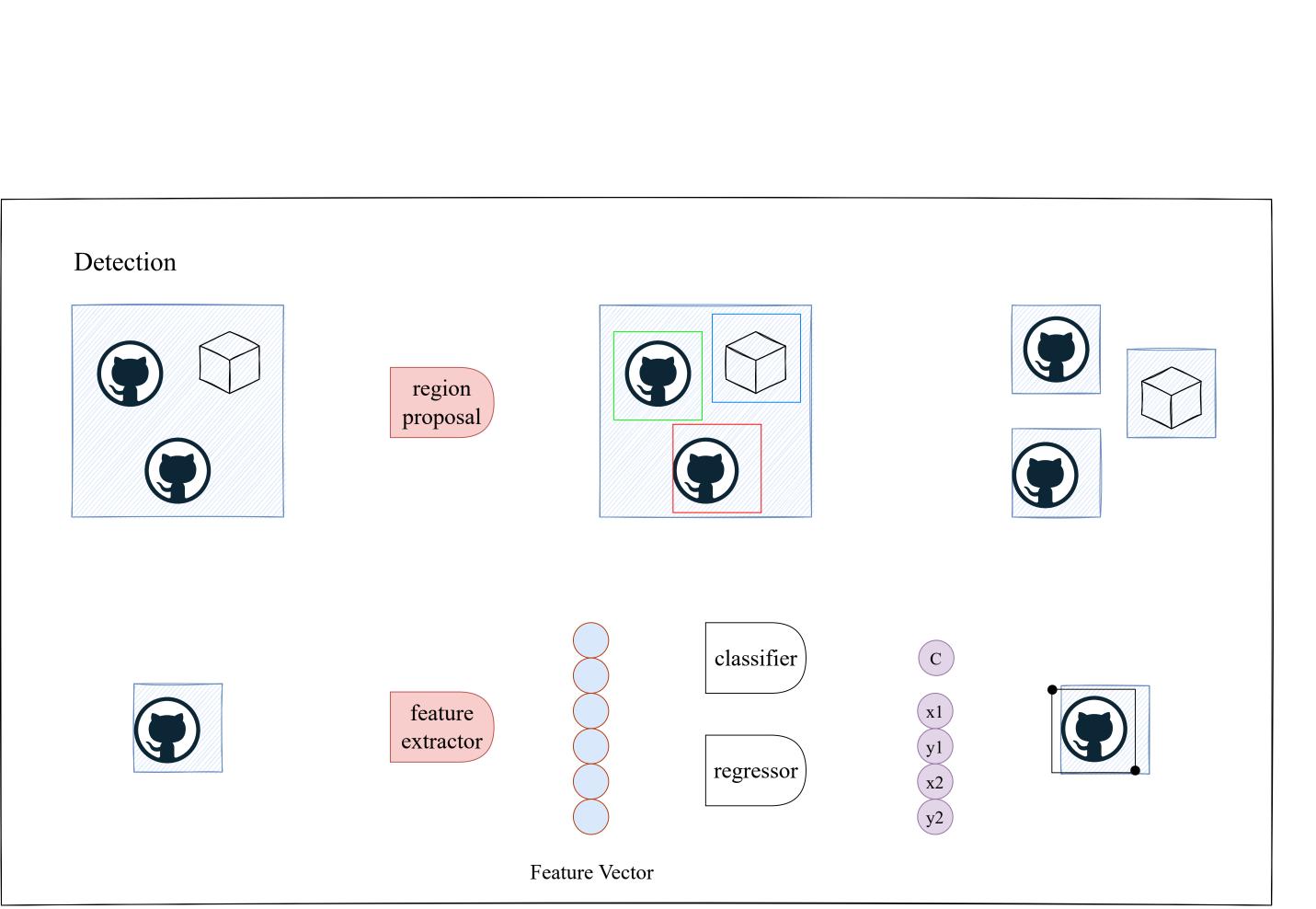
specific task on top of

models

Fine Tuning

features.





Natural Signals Natural Signals data IMP image - 3 million feature points. Isec of audio at standard sampling rate - 44.1 thousand feature points Properties of Natural signals Locality Stationarity Compositionality We utilize the properties to Improve the architecture Transfer learning