Announcements

Exams graded, will release solutions soon

Continue to work on your final projects

Wed 25th pre-recorded lecture

Herding Cats

How modeling turned into modern AI

CHIDATA

Recap: Model Fitting

Fitting: Observe xi, yi, **infer theta**

Regression Problems -> Continuous Y

Classification Problems -> Discrete Y

Recap: ~~Model Fitting~~ Machine Learning

Fitting: Observe xi, yi, **infer theta**

Regression Problems -> Continuous Y

Classification Problems -> Discrete Y

Recap: High Dimensional Data

"More Observations Than Individuals"

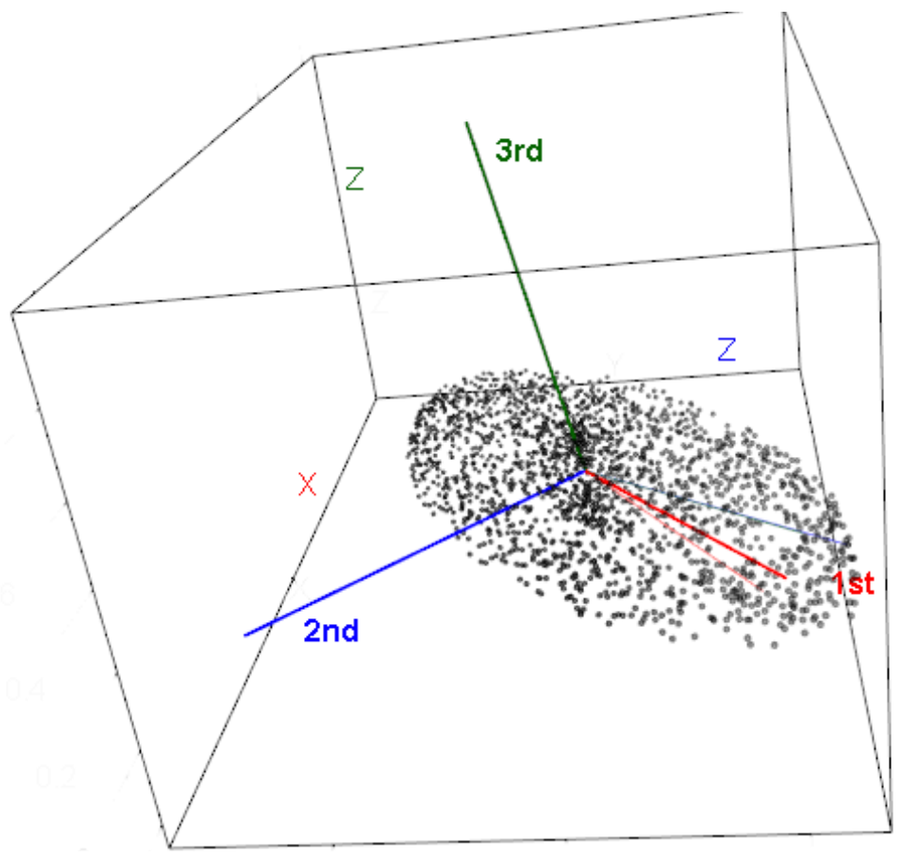
Stocks: 1000s of securities, Millions of trades

Genomics: 6.4 bn base pairs, 100s Clinical Population

Natural Language: Inf possible word sequences, Millions of Documents

Principal Component Analysis

Identify low dimensional structure in the data

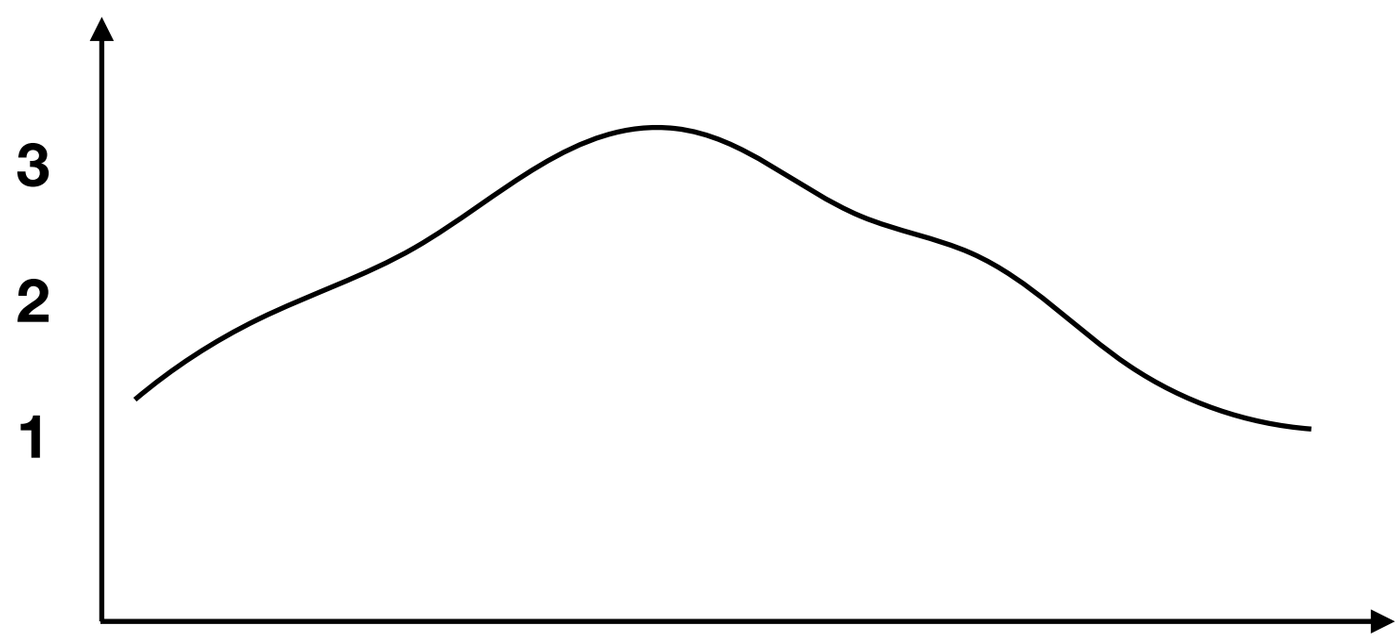


Spatial Structure

10 Dimensional Vector

1.5	2.0	2.2	2.5	3.1	2.9	2.4	2.0	1.3	0.7

Interpet as a sequence!



Spatial Structure

10 Dimensional Vector

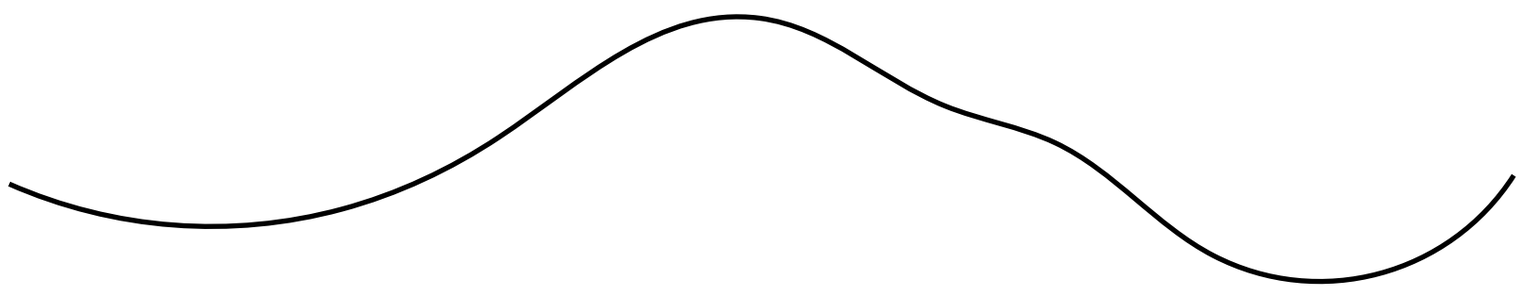
1.5	2.0	2.2	2.5	3.1	2.9	2.4	2.0	1.3	0.7

Nearby elements are similar

Spatial Structure

10 Dimensional Vector

1.5	2.0	2.2	2.5	3.1	2.9	2.4	2.0	1.3	0.7



Seasonality

Spatial Structure

10 Dimensional Vector

1.5	2.0	2.2	2.5	3.1	2.9	2.4	2.0	1.3	0.7

Effective degrees of freedom less than possible.

Stock Prices

1 element for each month, closing price

1.5	2.0	2.2	2.5	3.1	2.9	2.4	2.0	1.3	0.7

Predict stock split

“Time” series

Spatial Structure

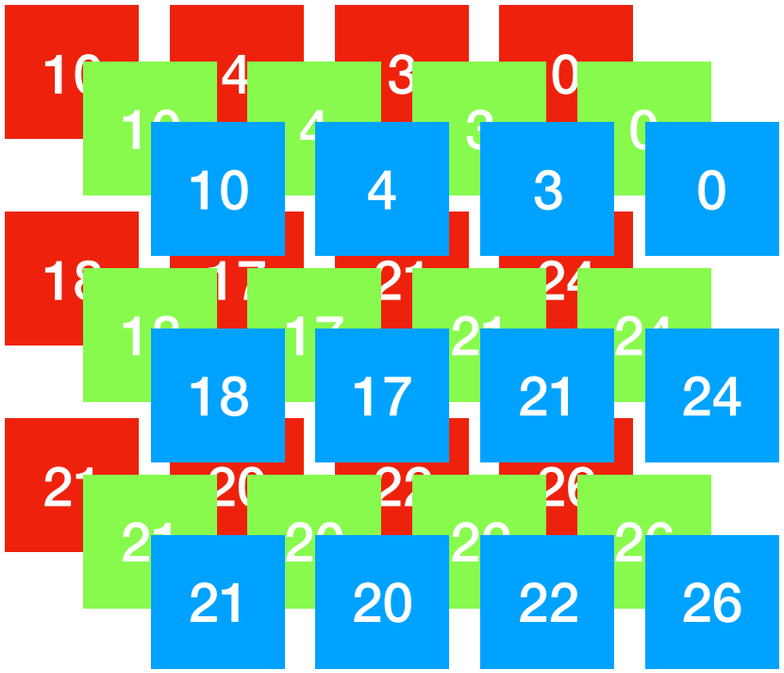
9 Dimensional Vector

1.5	2.0	2.2
2.5	3.1	2.9
2.4	2.0	1.3

Laid out as a grid

Images

3 2D Grids Representing Pixels



Values 0-255

Language

	Topic1	Topic2	Topic3
The	0.15	.55	0.3
bear	0.3	0.6	0.1
ran	0.4	0.5	0.1
fast	0.5	0.5	0.0

Recap: Linear Models

$$y = A \cdot \mathbf{x} + \epsilon$$

$$A \cdot \mathbf{x} = a_1 x_1 + a_2 x_2 + \ldots + a_d x_d$$

Positive a (means positive correlation),
negative a (negative correlation)

Difficult to apply to such data due to dimensionality.

Reduce the size of the model

Convolution

Signal

1.5	2.0	2.2	2.5	3.1	2.9	2.4	2.0	1.3	0.7

Filter size=2

A1	A2

Output: N-size = 10-2 = 8 dimensional vector

Equivalent For Sequences?

Convolution

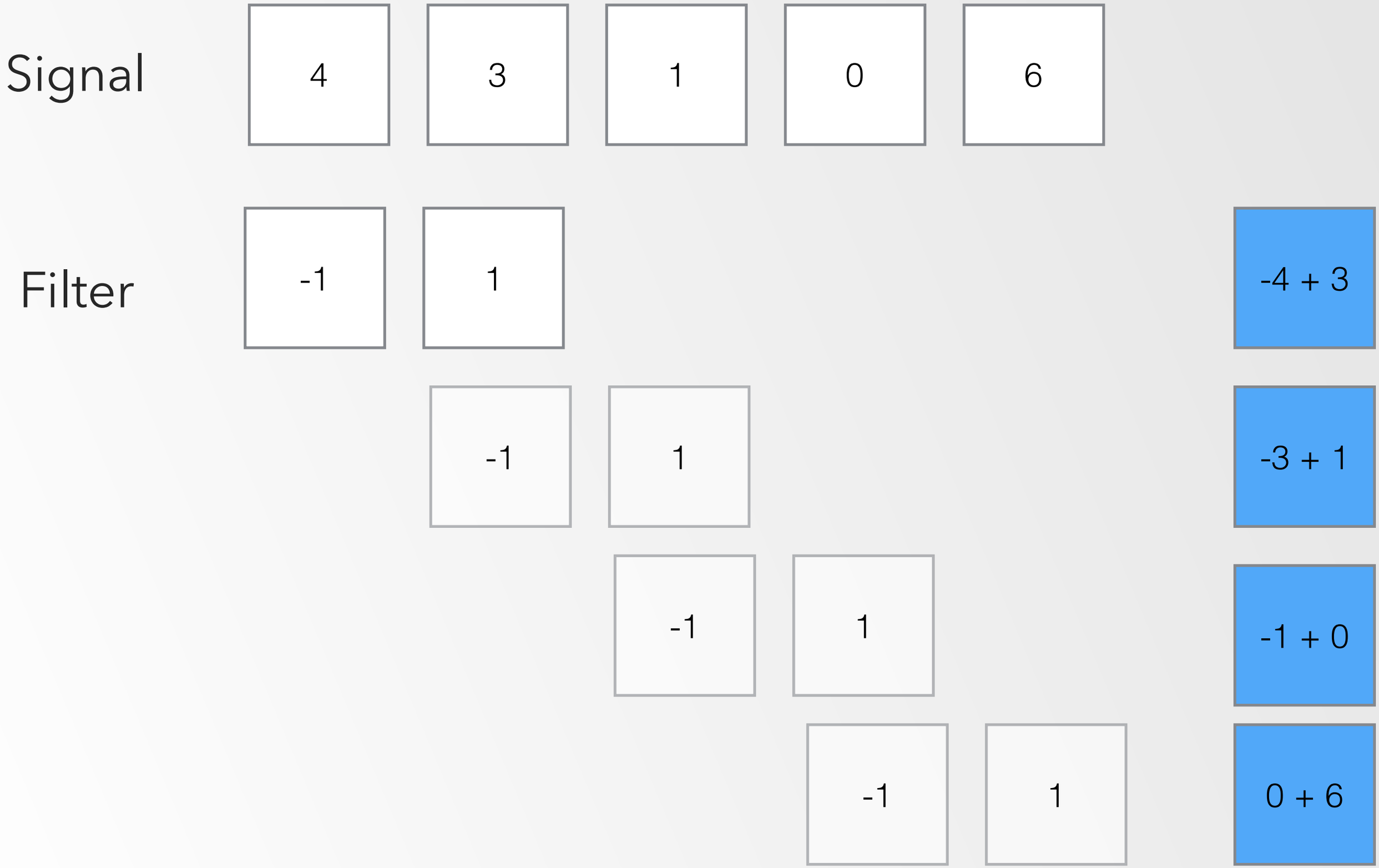
Signal

1.5	2.0	2.2	2.5	3.1	2.9	2.4	2.0	1.3	0.7

Filter size=4

A1	A2	A3	A4

Output: N-size = 10-4 = 6 dimensional vector


Signal
4
3
1
0
6
Filter
-1
1
-4 + 3
-1
1
-3 + 1
-1
1
-1 + 0
-1
1
0 + 6

Equivalent for Sequences?

Convolution

$$A * \mathbf{x}$$

Sliding dot product of A across the values of X

Interpretation?

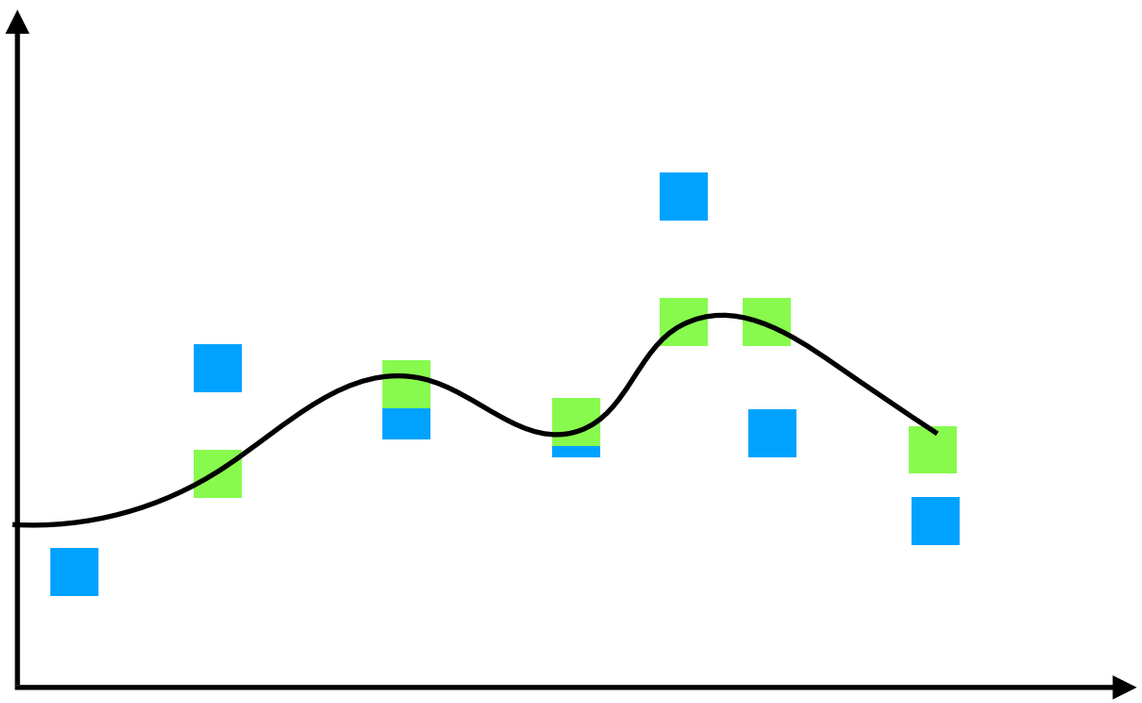
$$y = A \cdot \mathbf{x} + \epsilon$$

$$A \cdot \mathbf{x} = a_1 x_1 + a_2 x_2 + \ldots + a_d x_d$$

Positive a (means positive correlation),
negative a (negative correlation)

"Moving average"

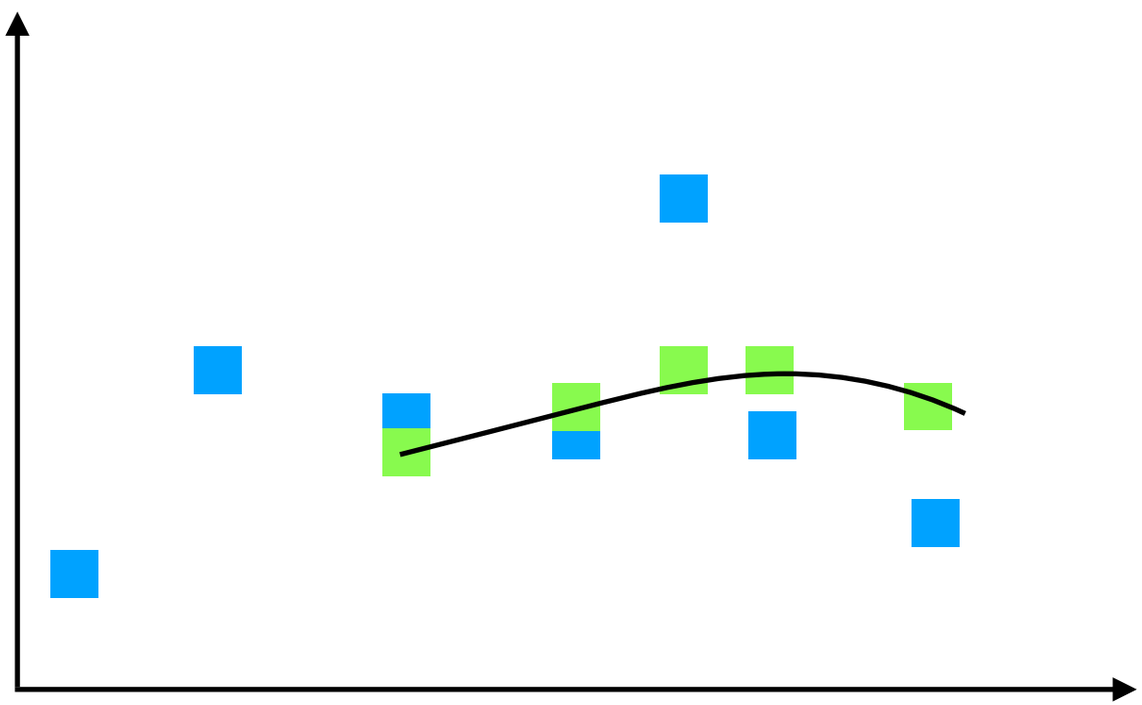
Moving Average



Smooths out the signal

"Moving average"

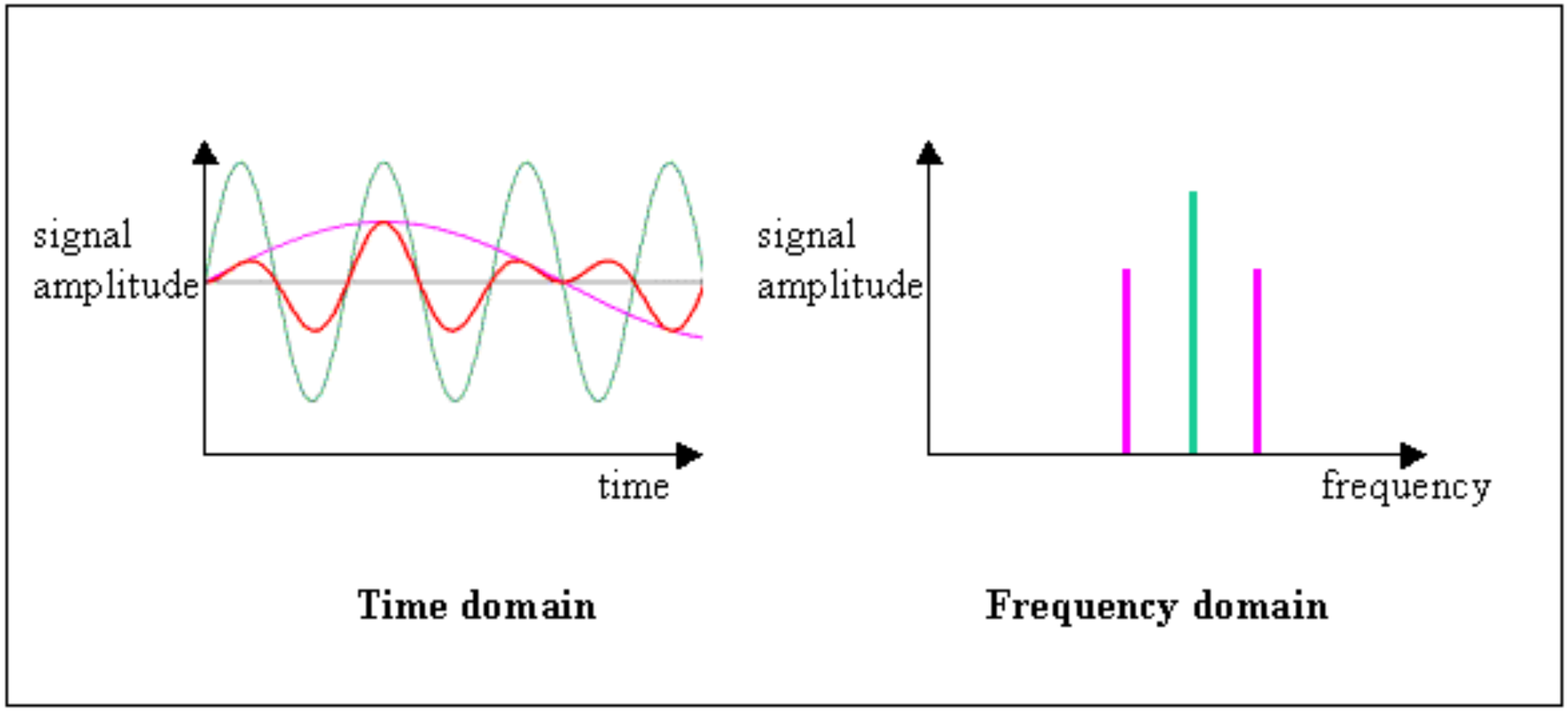
Moving Average



Smooths out the signal

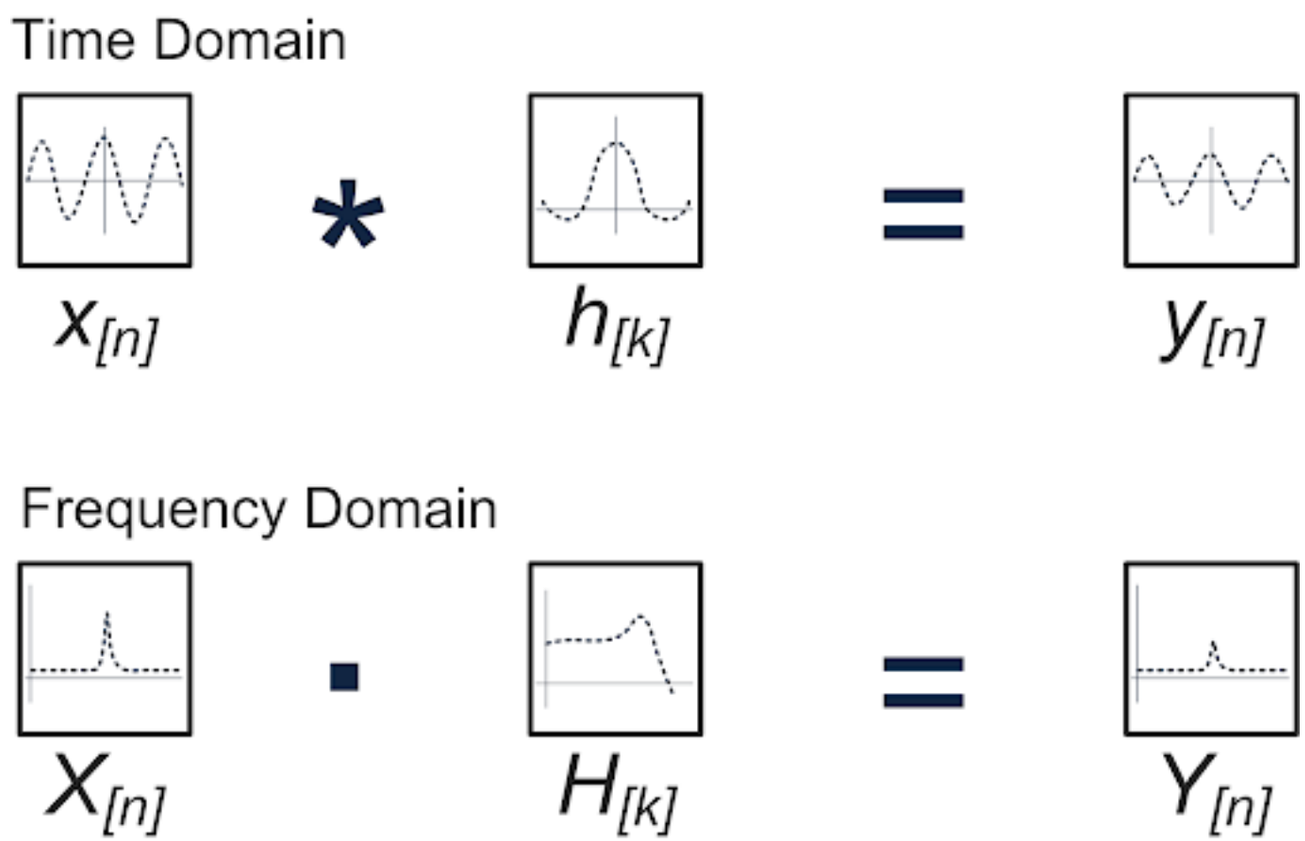
"Change Detector"

Every Signal Has A “Frequency” Domain



Every Signal Has A “Frequency” Domain

Convolutions filter certain frequencies



“Low Pass Filter”: rejects rapid spikes in the signal

“High Pass Filter”: identifies changes in the signal

Signal

4	3	1	0
6	3	0	1
6	3	0	1

Filter

-1	1
0	1

Signal

4	3	1	0
6	3	0	1
6	3	0	1

Filter

-1	1
0	1

Signal

4	3	1	0
6	3	0	1
6	3	0	1

Filter

-1	1
0	1

Output

2	0
0	2

Every Signal Has A “Frequency” Domain

Original

Low Pass

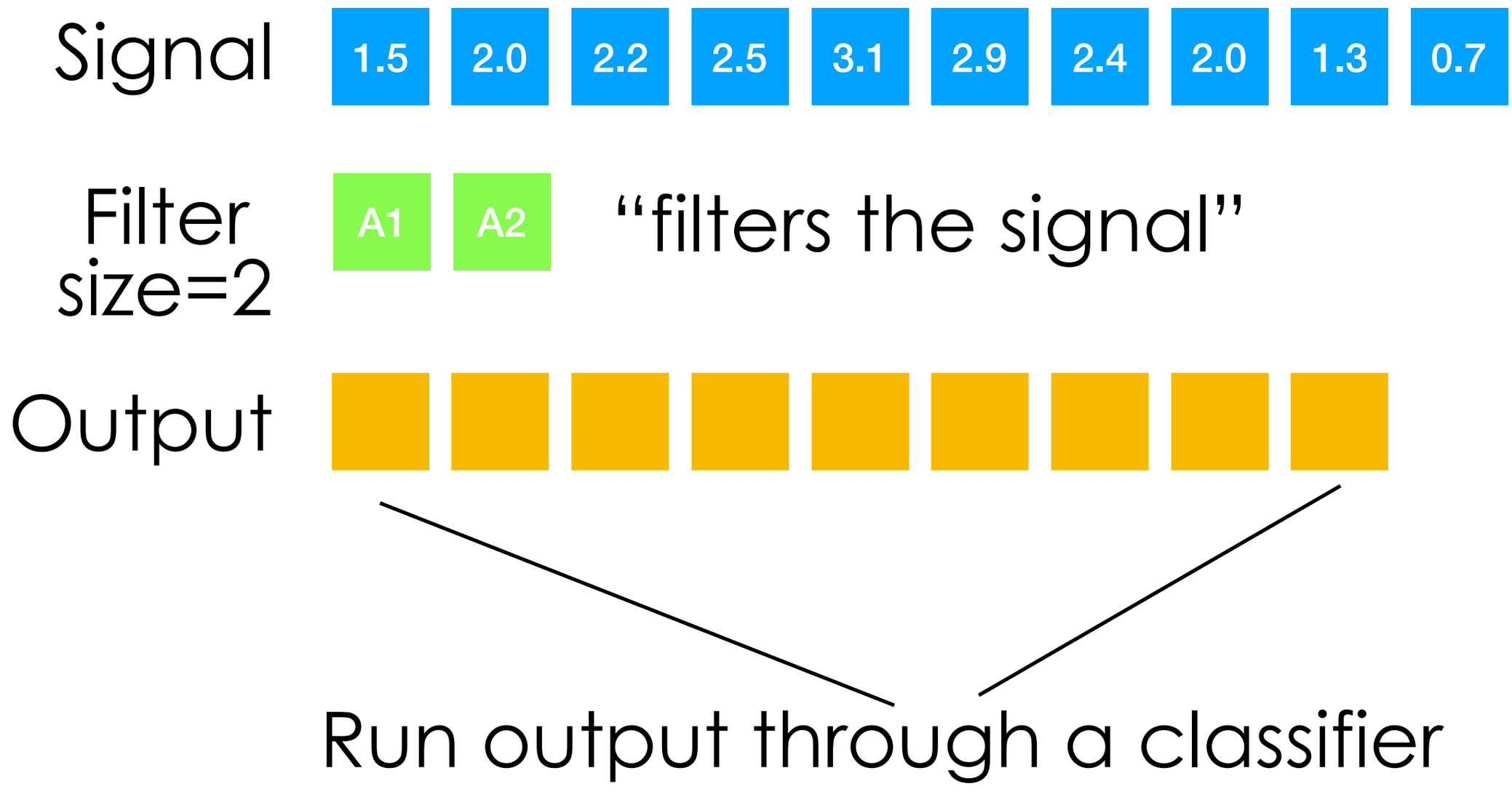
High Pass

“Low Pass Filter”: blurs the image

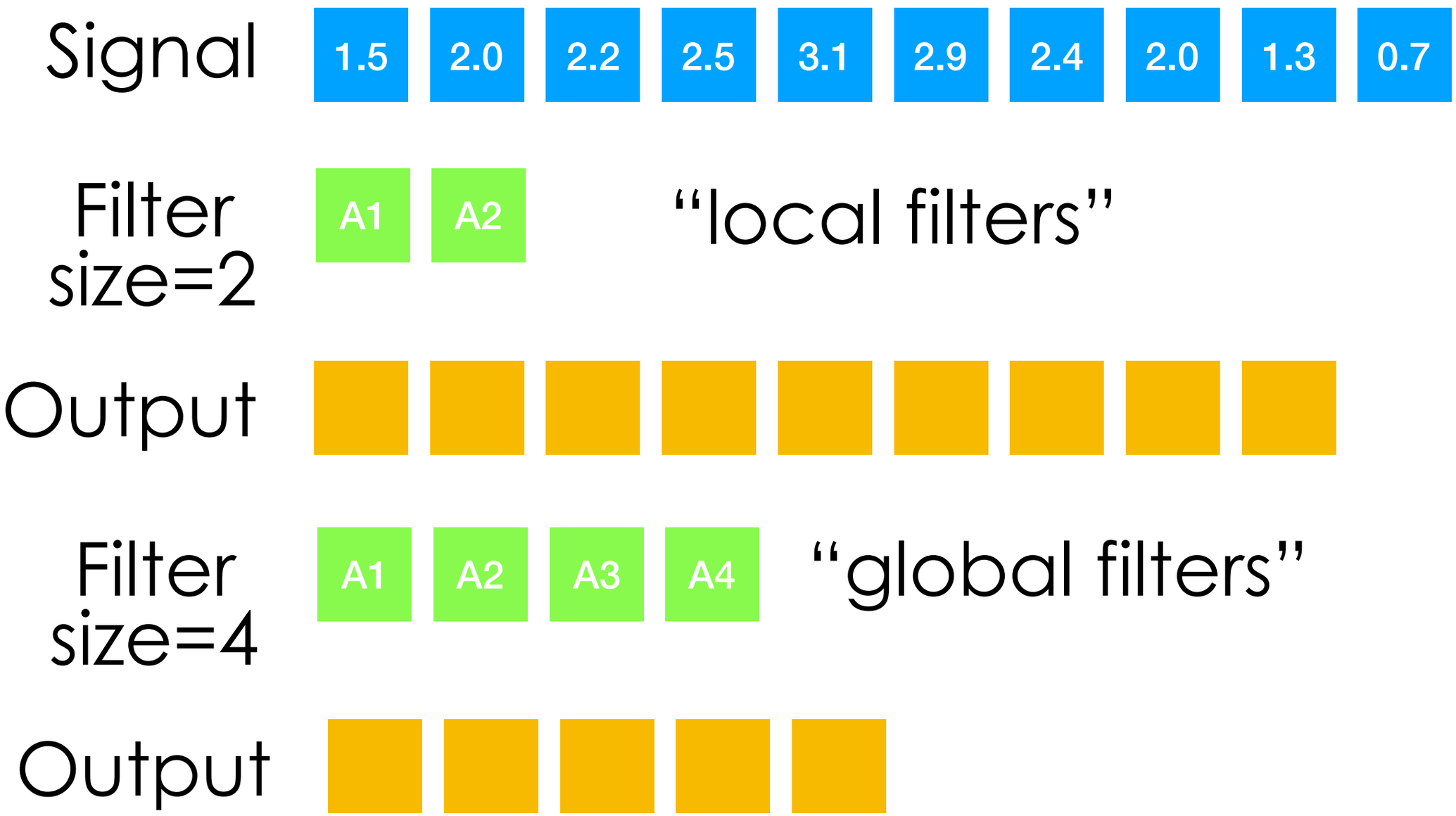
“High Pass Filter”: detects edges in the image

Anomaly Detection Example

Idea: Apply Convolution Before Classification

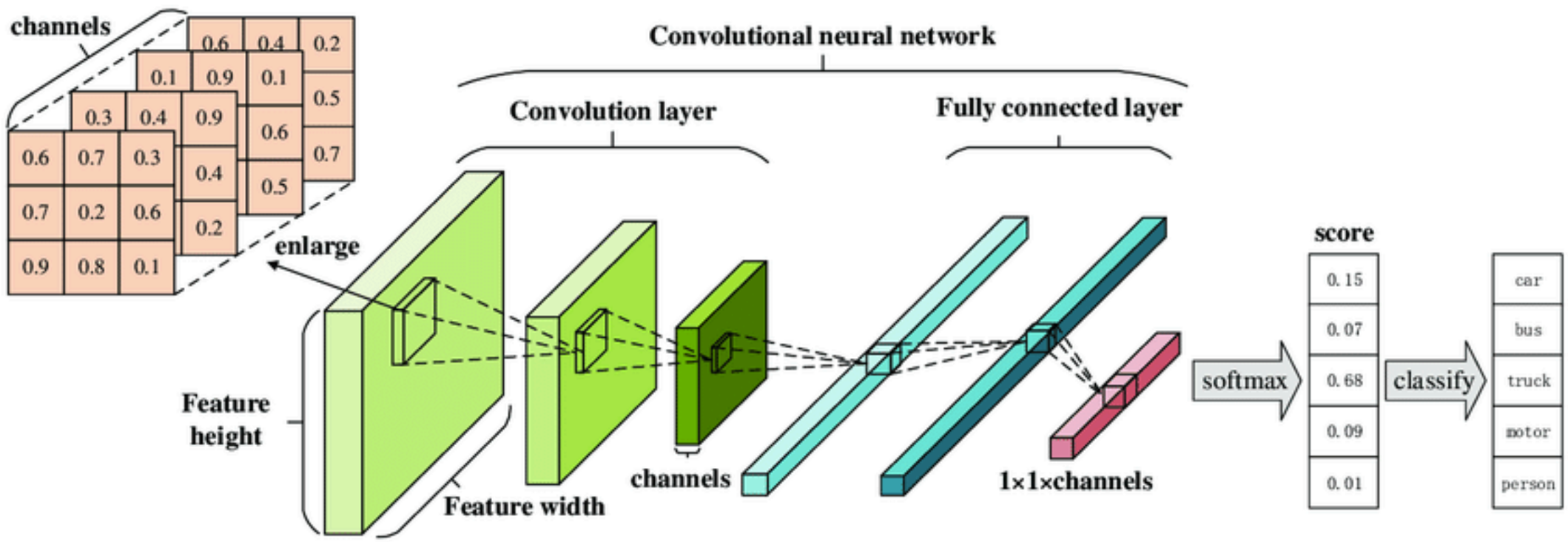


Why Only One?


Signal
1.5
2.0
2.2
2.5
3.1
2.9
2.4
2.0
1.3
0.7
Filter size=2
A1
A2
"local filters"
Output
Filter size=4
A1
A2
A3
A4
"global filters"
Output

Convolutional Neural Network

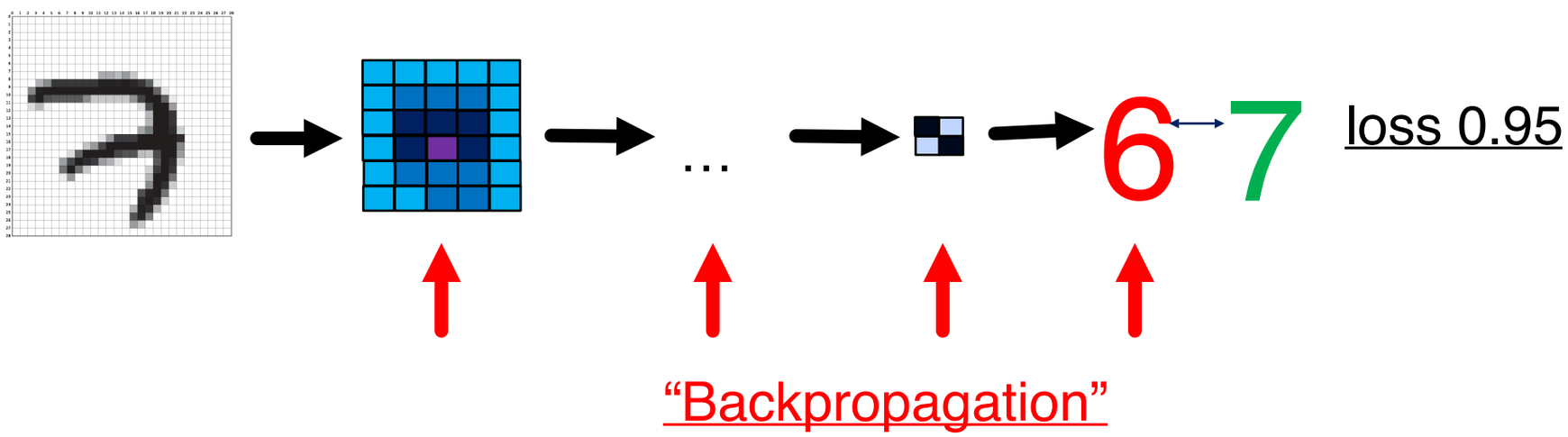
Cascade of filters that can be trained (fit) as one unit



"Layers" that consist of convolutions, classifications, regressions

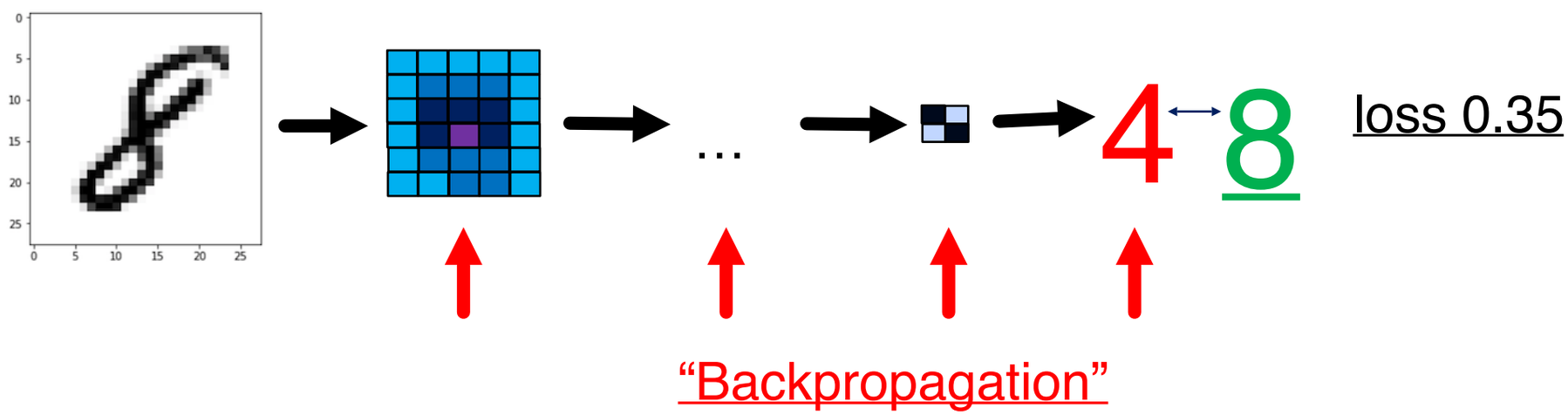
Training

“Training”: Provide images and correct values



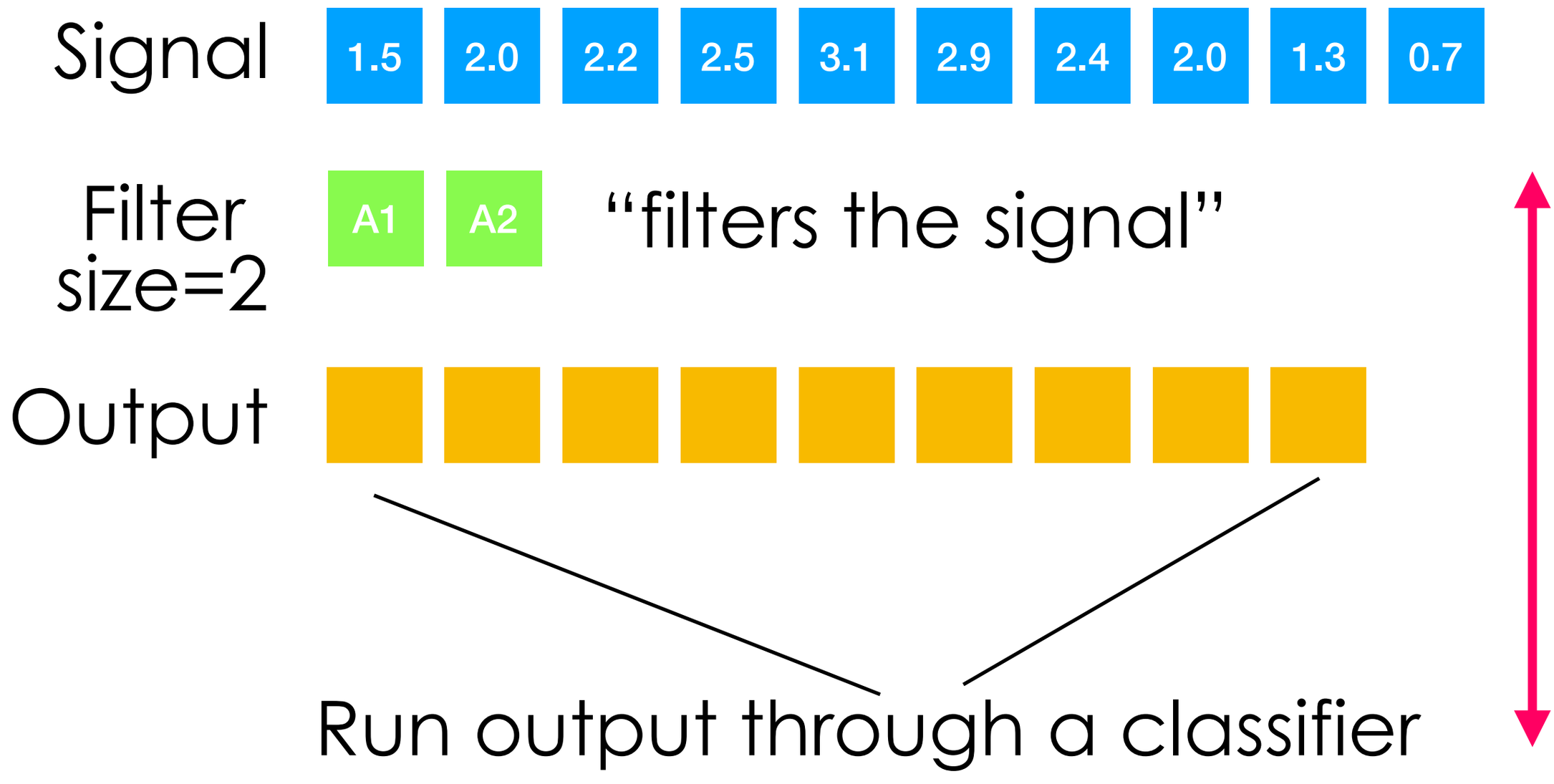
Intuition

“Training”: Provide images and correct values



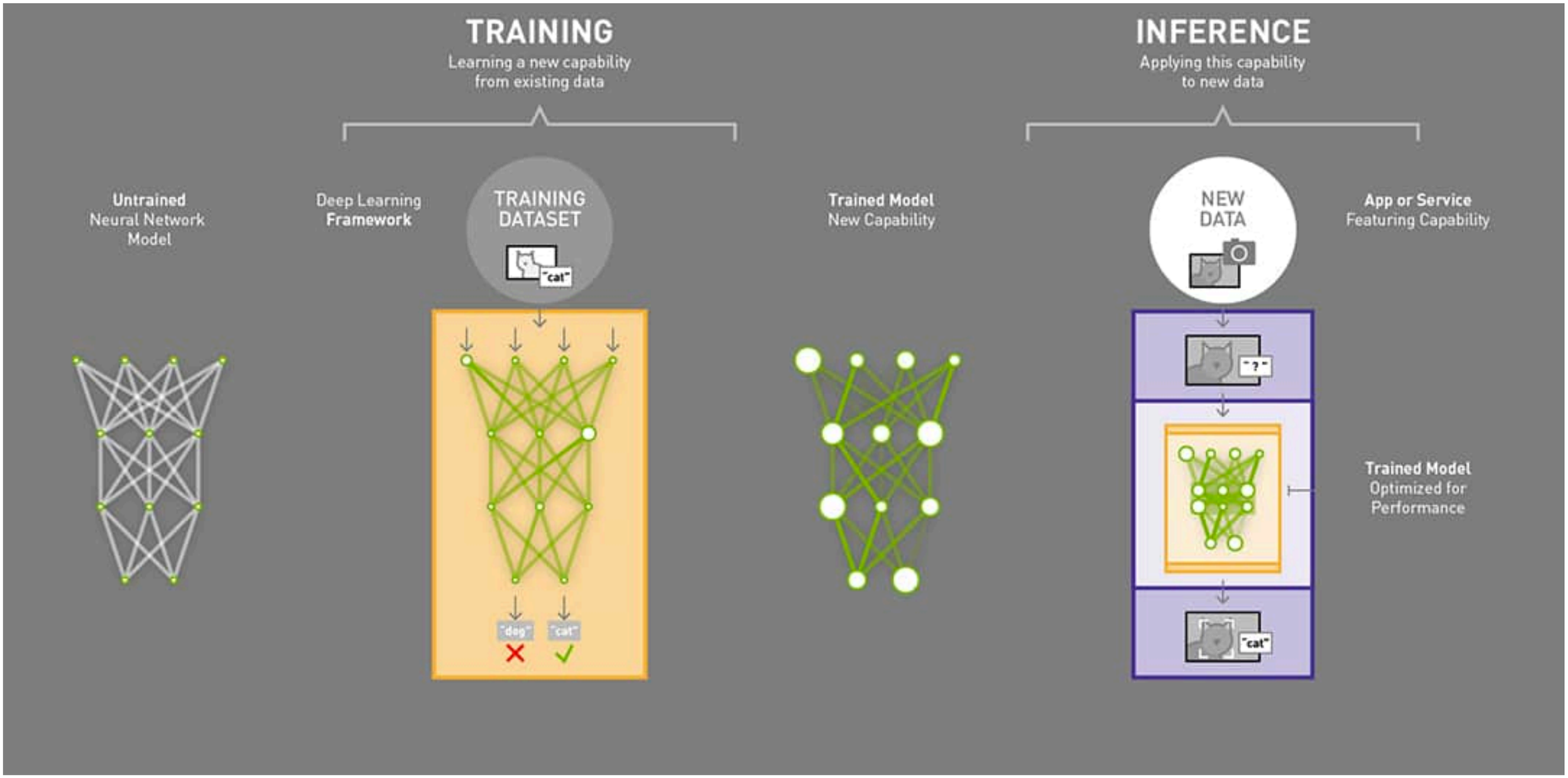
Anomaly Detection Example

Idea: Apply Convolution Before Classification



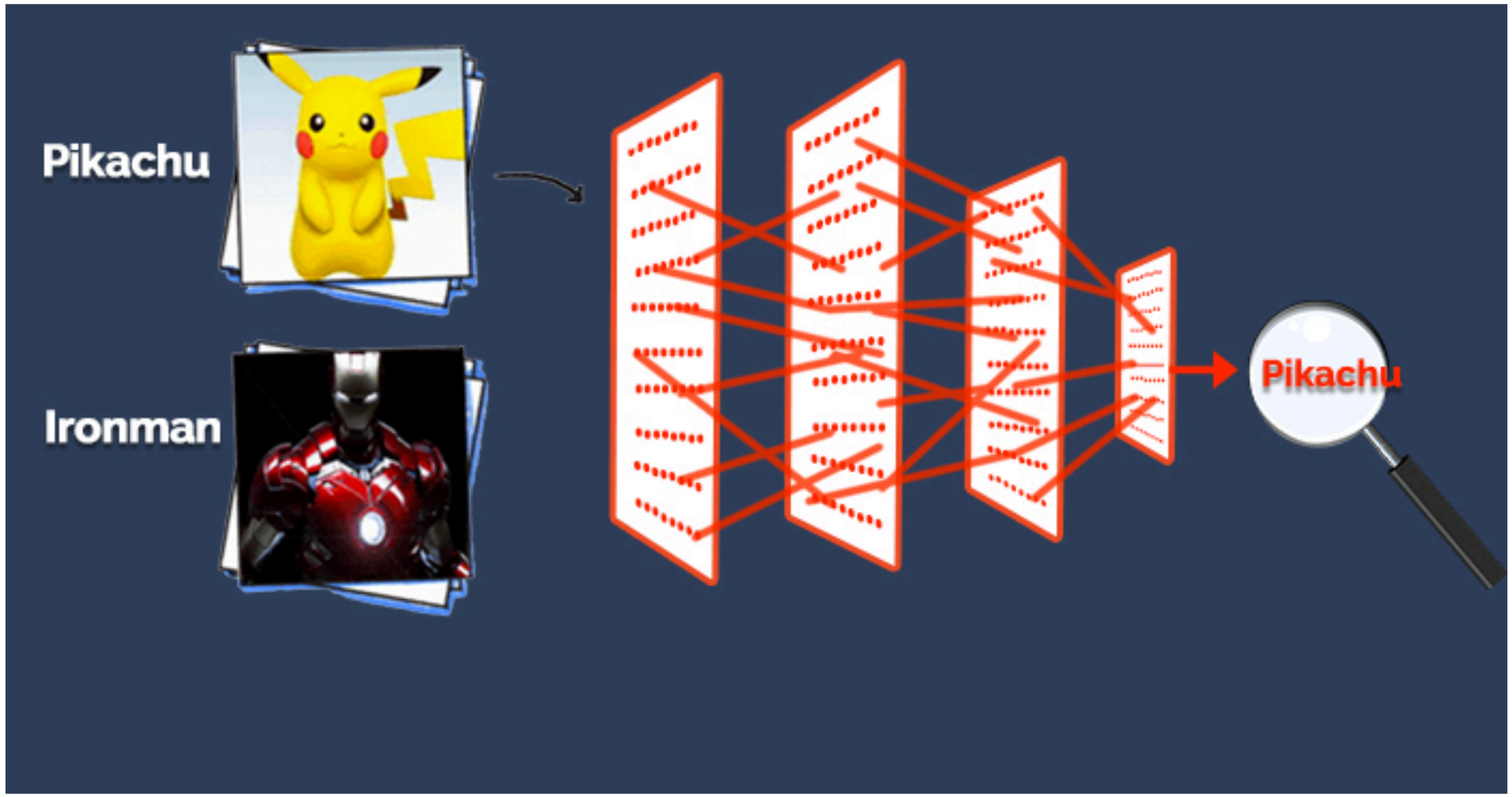
CHIDATA

How it works



Powerful and general strategy

Image classification



Powerful and general strategy

Downside? #1 Need lots of data

Image classification datasets
> 100k examples (upto millions)

Powerful and general strategy

Downside? #2 Need lots of compute

Specialized hardware to train (lots of GPUs)

Exactly why it took so long!

2008 first time CNNs actually “worked”

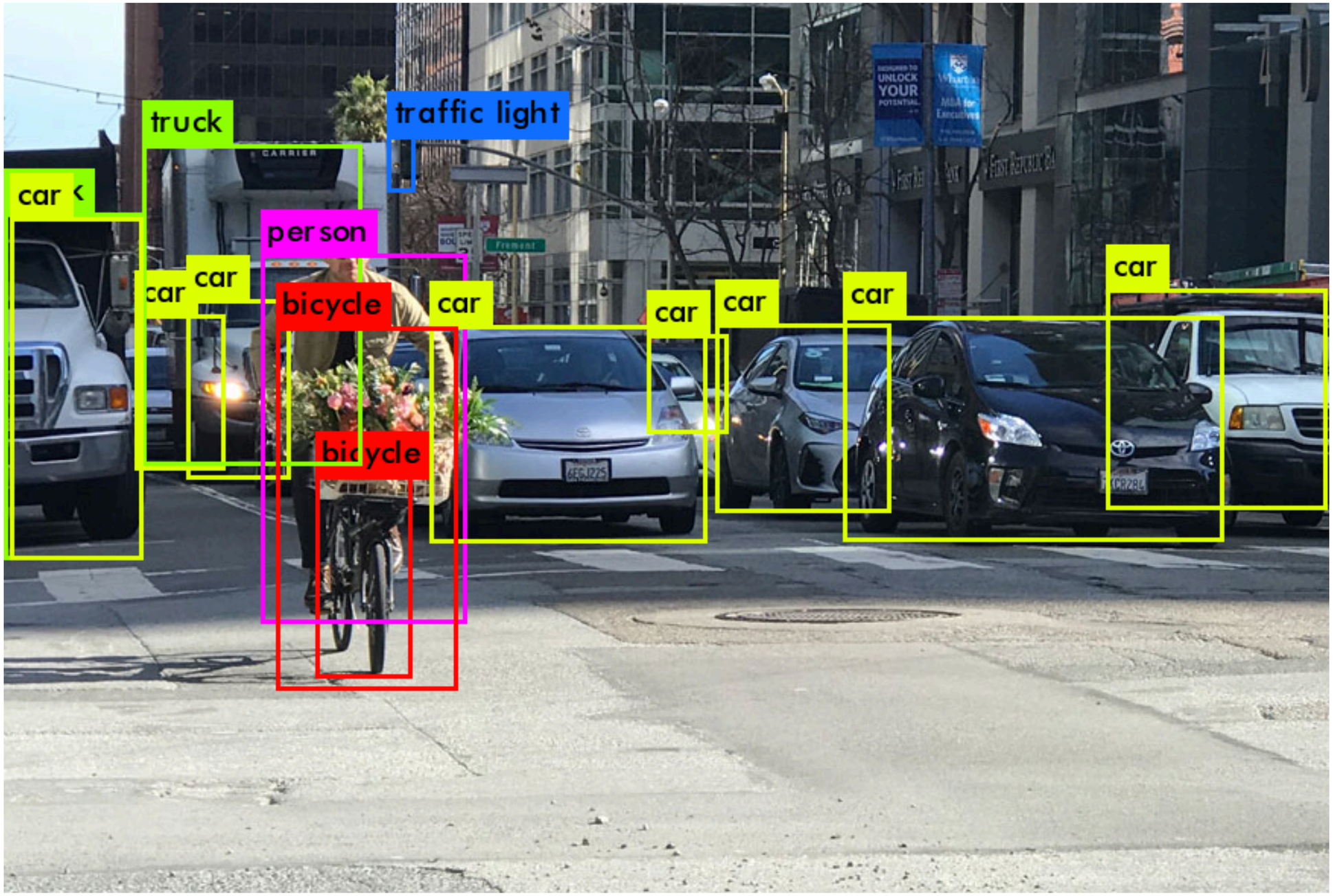
Why?

GPU programming from University of Toronto

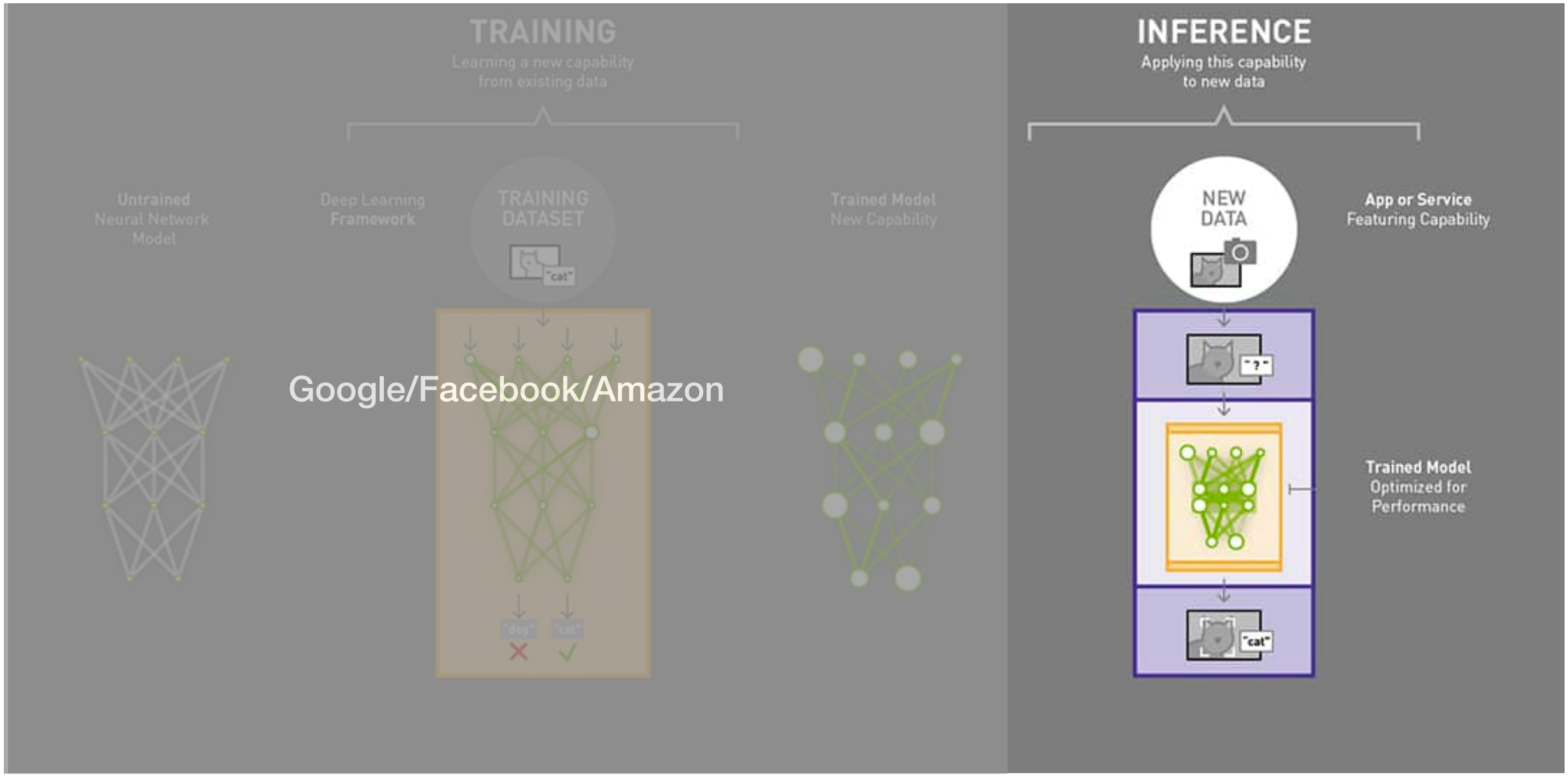
Large image dataset from Stanford

CHIDATA

Works!



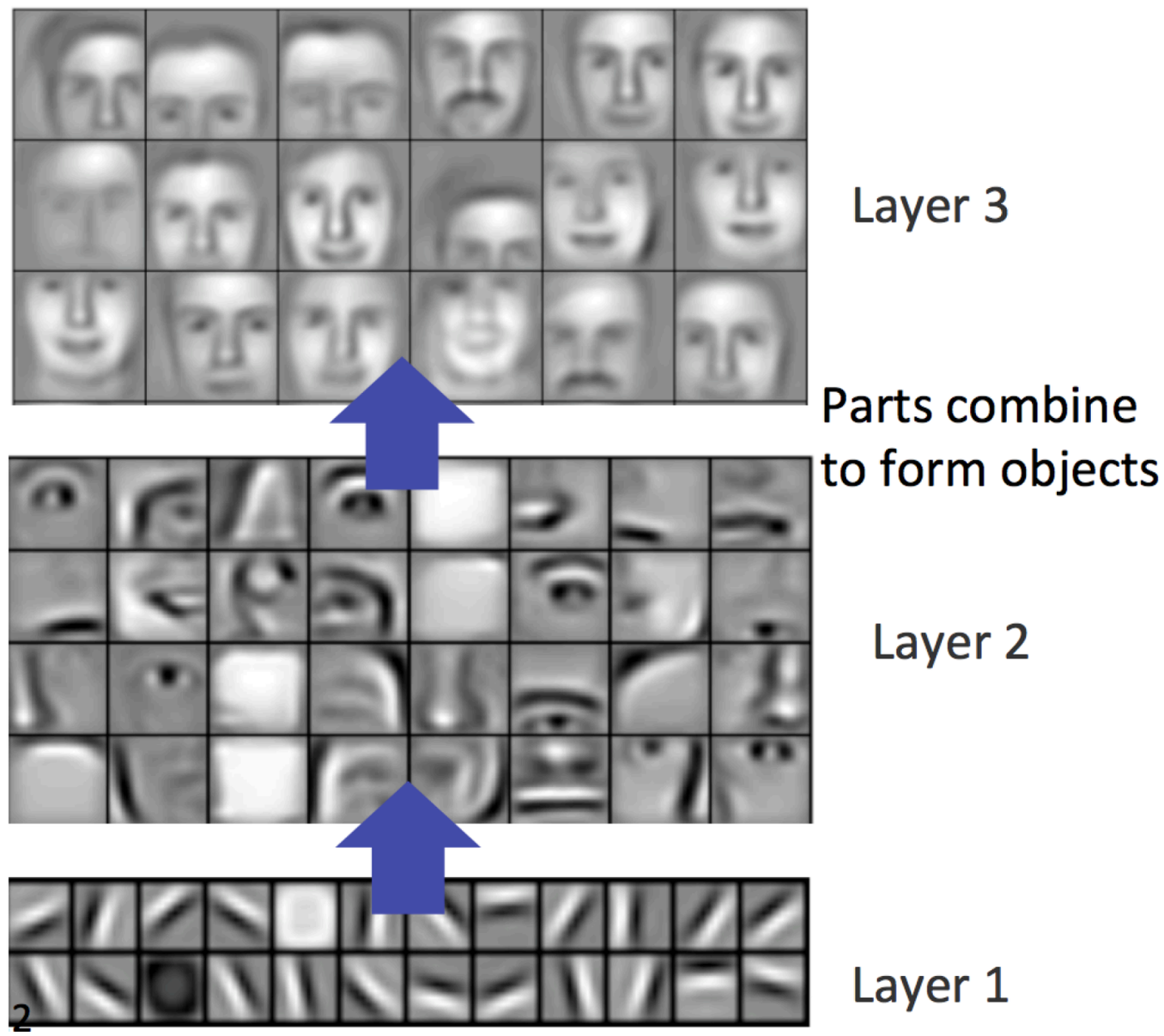
Pre-trained models



Google/Facebook/Amazon

Models Trained On Large Datasets Generalize

Filters generally apply well to new data!



Anomaly Detection Example

Idea: Apply Convolution Before Classification

Signal

1.5	2.0	2.2	2.5	3.1	2.9	2.4	2.0	1.3	0.7

Filter size=2

A1 A2

"filters the signal"

Output

Specific to your business

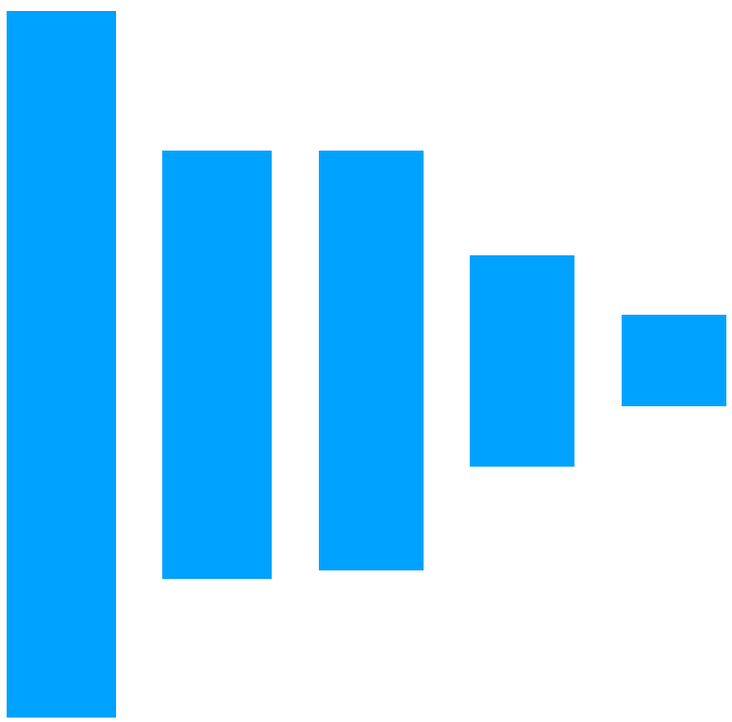
Run output through a classifier

Never Have To Implement By Hand

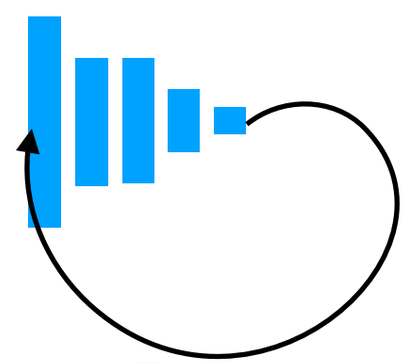
TensorFlow

Never Have To Implement By Hand

Step 1: Define Network Structure



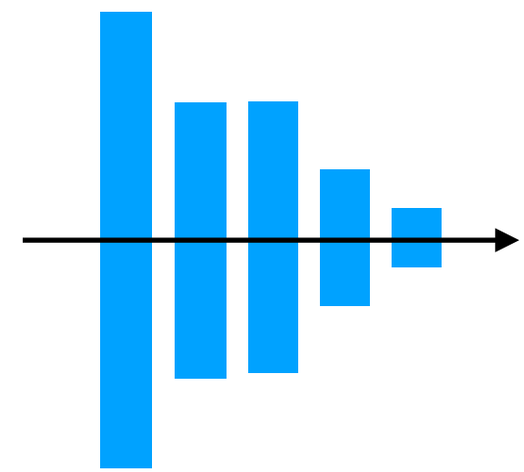
Step 2: Define Training



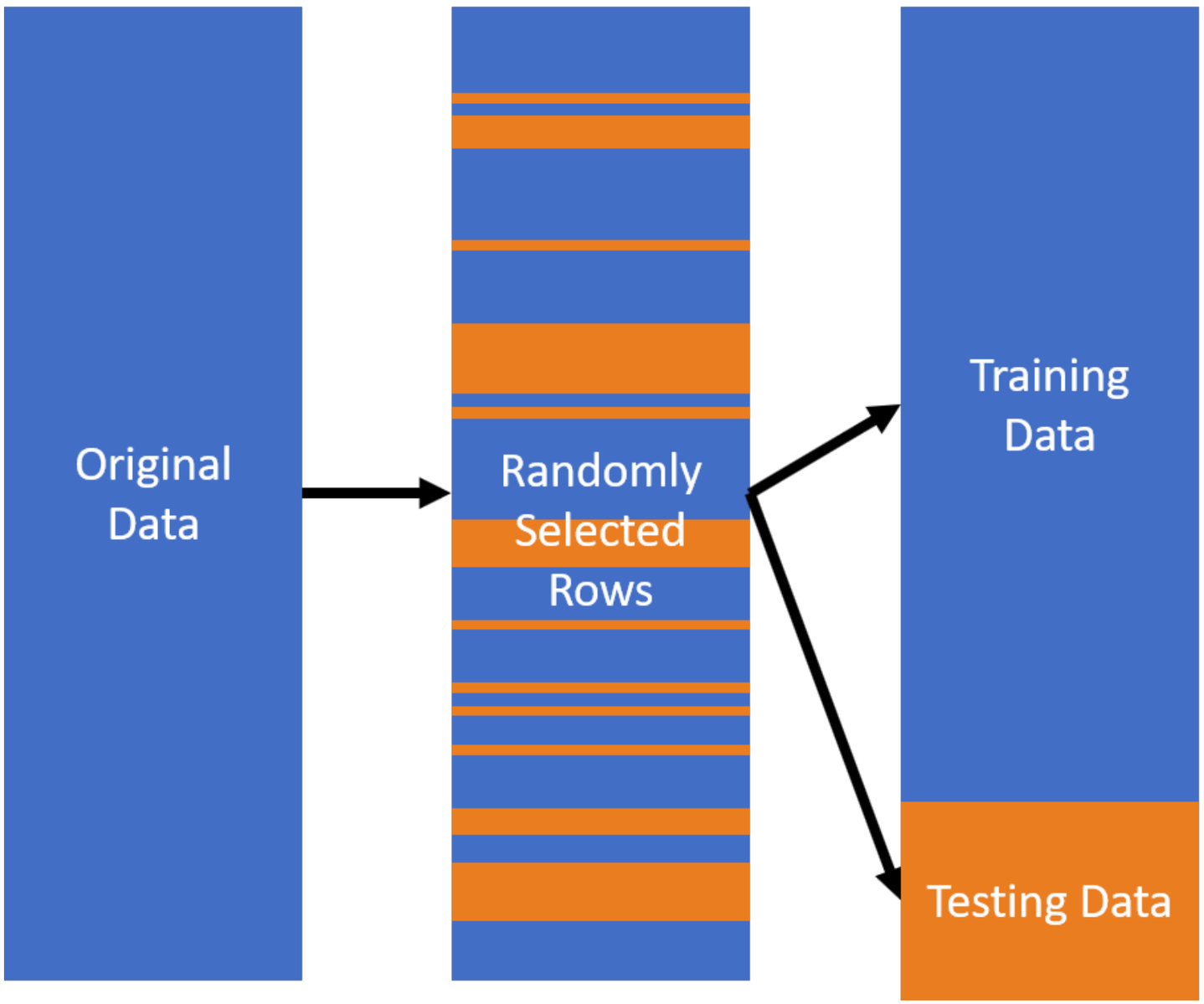
Data

How fast to update

Step 3: Predict



Training and Testing



Use this data for rule selection

Use this data for rule evaluation

Before Wed

https://www.tensorflow.org/tutorials/keras/classification

Go Through a Tutorial

Why I believe it Works

Identify low dimensional structure in the data