Carnegie Mellon University
Heinz College

Unstructured Data Analysis for Policy

Lecture 10: Hyperparameter tuning, classifier evaluation, intro to neural nets & deep learning

George Chen

But we could have chosen different proper training/validation data!

Training data point	Training data point	Training data point	Training data point	Training data point
Training data point	Training data point	Training data point	Training data point	Training data point

Train method on data in gray

Predict on data in orange

We get 5 different prediction errors… which is more accurate? 🧐

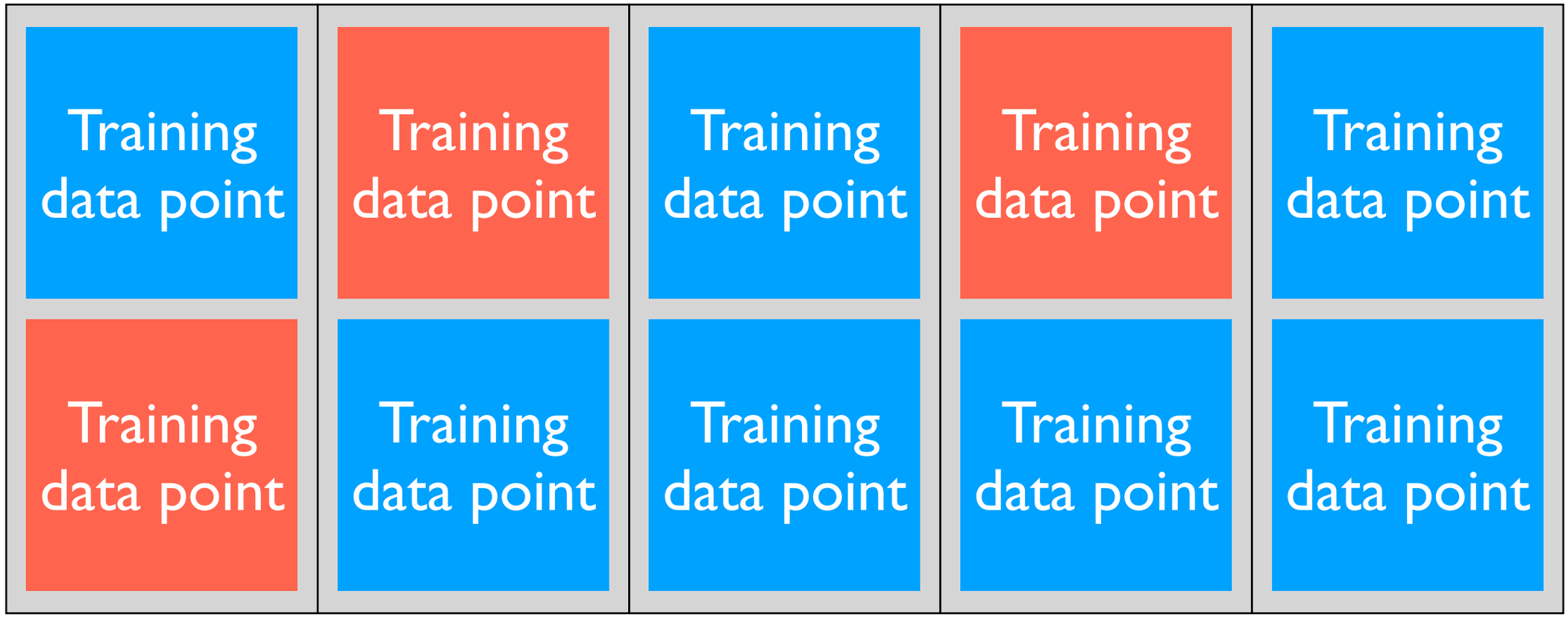
Compute prediction error

0% 0% 50% 0% 50%

Unclear which is best, so let's just average: (0+0+50+0+50)/5 = 20%

not the same ***k*** as in ***k***-means or ***k***-NN classification

k-fold Cross-Validation



1. Shuffle data and split them into 5 (roughly) equal size portions
2. For each of the equal sized portions:
   (a) Treat the current portion has the validation data and the rest as proper training data
   (b) Train on the proper training data, predict on the validation data
   (c) Compute prediction error
3. Compute average prediction error

k = 5

"cross validation score"

You need to specify how to measure prediction error!

Choosing *k* in *k*-NN Classification

For each ***k*** = 1, 2, 3, …, the maximum ***k*** you are willing to try:

Compute 5-fold cross validation score using ***k***-NN classifier as prediction method

Use whichever ***k*** has the best cross validation score

Automatic Hyperparameter Selection

Suppose the prediction algorithm you're using has hyperparameters θ

For each hyperparameter setting θ you are willing to try:

Compute 5-fold cross validation score using your algorithm with hyperparameters θ

Using whichever θ has the best cross validation score, train model on full training set

Why 5?

People have found using 10 folds or 5 folds to work well in practice but it's just empirical — there's no deep reason

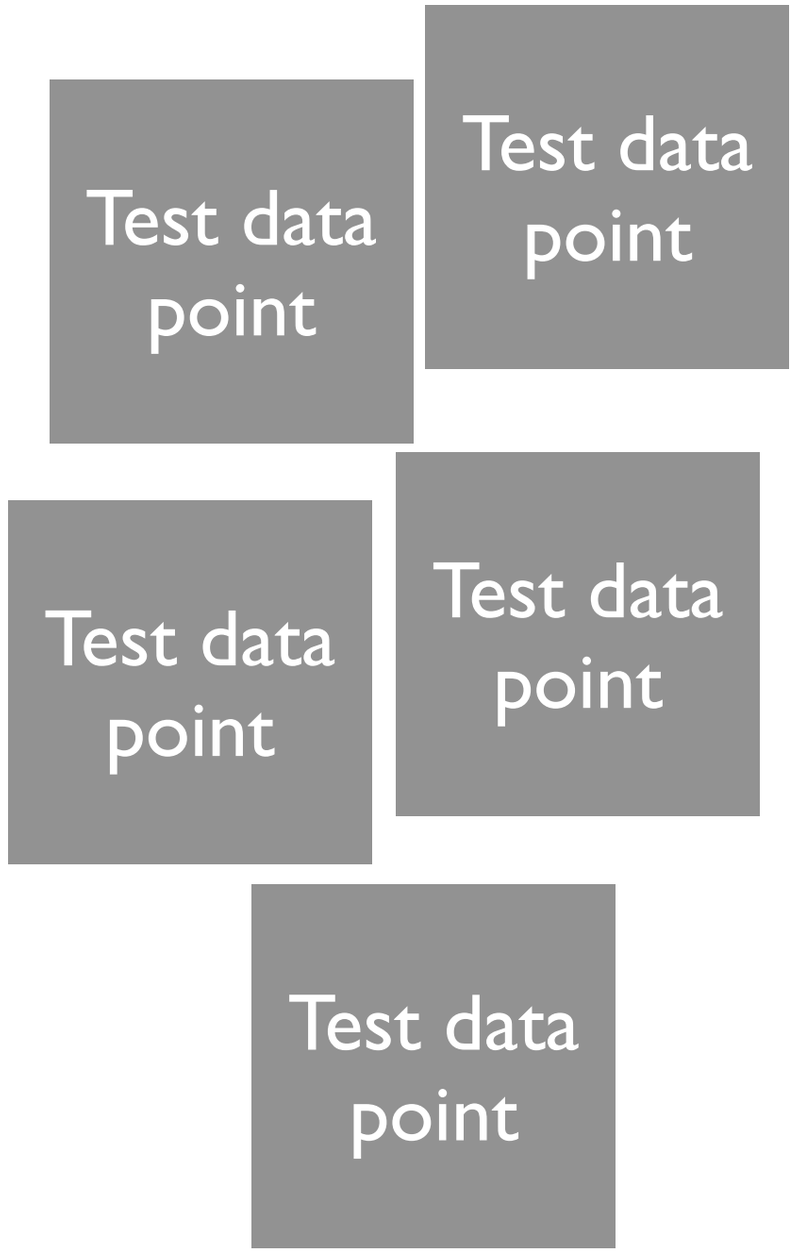
Important: the errors from data splitting and cross-validation are *estimates* of the true error on test data!

Example: earlier, we got a cross validation score of 20% error

This is a guess for the error we will get on test data

This guess is <u>not</u> always accurate!

Want to classify these points correctly



Example: future emails to classify as spam/ham

Cross-Validation Remarks

- ***k***-fold cross-validation is a <u>randomized</u> procedure

  Re-running CV results in different cross-validation scores!

- Suppose there are ***n*** training data points and ***k*** folds
  - If we are trying 10 different hyperparameter settings, how many times do we do model fitting? **$10k$**

    If this number is similar in size to ***n***, CV can overfit!
  - How many training data are used in each model fit during cross-validation? **$[(k-1)/k]n$**

    Smaller # folds typically means faster training

- If ***k* = *n***, would re-running cross-validation result in different cross-validation scores? What about ***k* = 2**?

  For deterministic training procedure: same CV result for ***k* = *n*** (since shuffling doesn't matter), different for ***k* = 2**

While we haven't covered neural nets (& deep learning) yet, as a warning: cross-validation is *not* often used with neural nets for reasons we'll discuss later

Instead, with neural nets, it's common to just use the simple train/val split we talked about earlier

Different Ways to Measure Accuracy

Simplest way:

- **Raw error rate:** fraction of predicted labels that are wrong (this was in our cross validation example earlier)

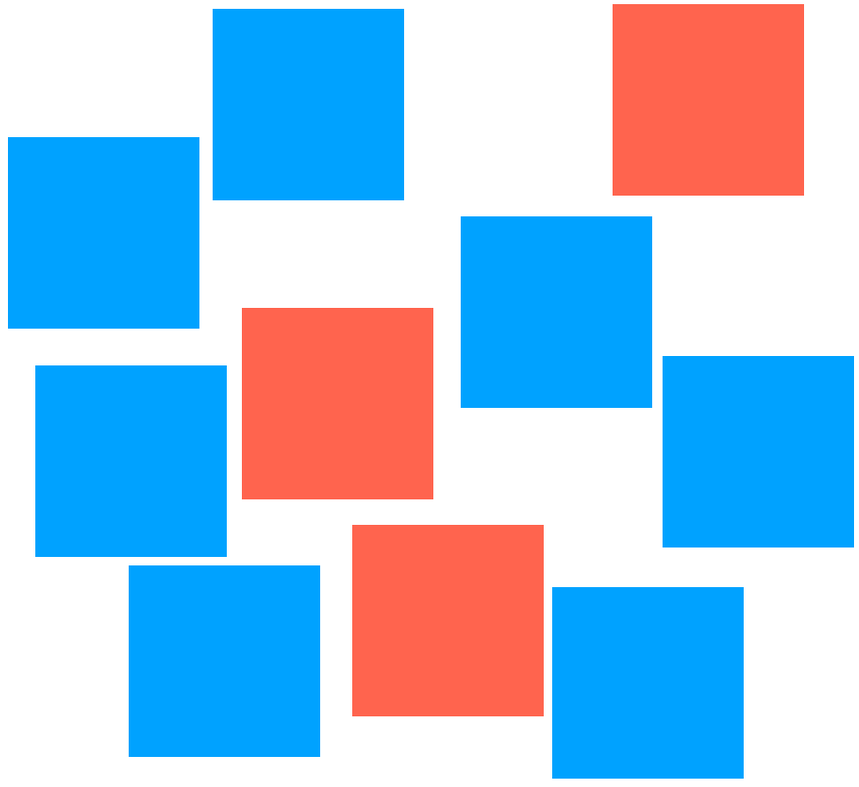
In "binary" classification (there are 2 labels such as spam/ham) when 1 label is considered "positive" and the other "negative":

Different Ways to Measure Accuracy

Simplest way:

- **Raw error rate:** fraction of predicted labels that are wrong (this was in our cross validation example earlier)

In "binary" classification (there are 2 labels such as spam/ham) when 1 label is considered "positive" and the other "negative":

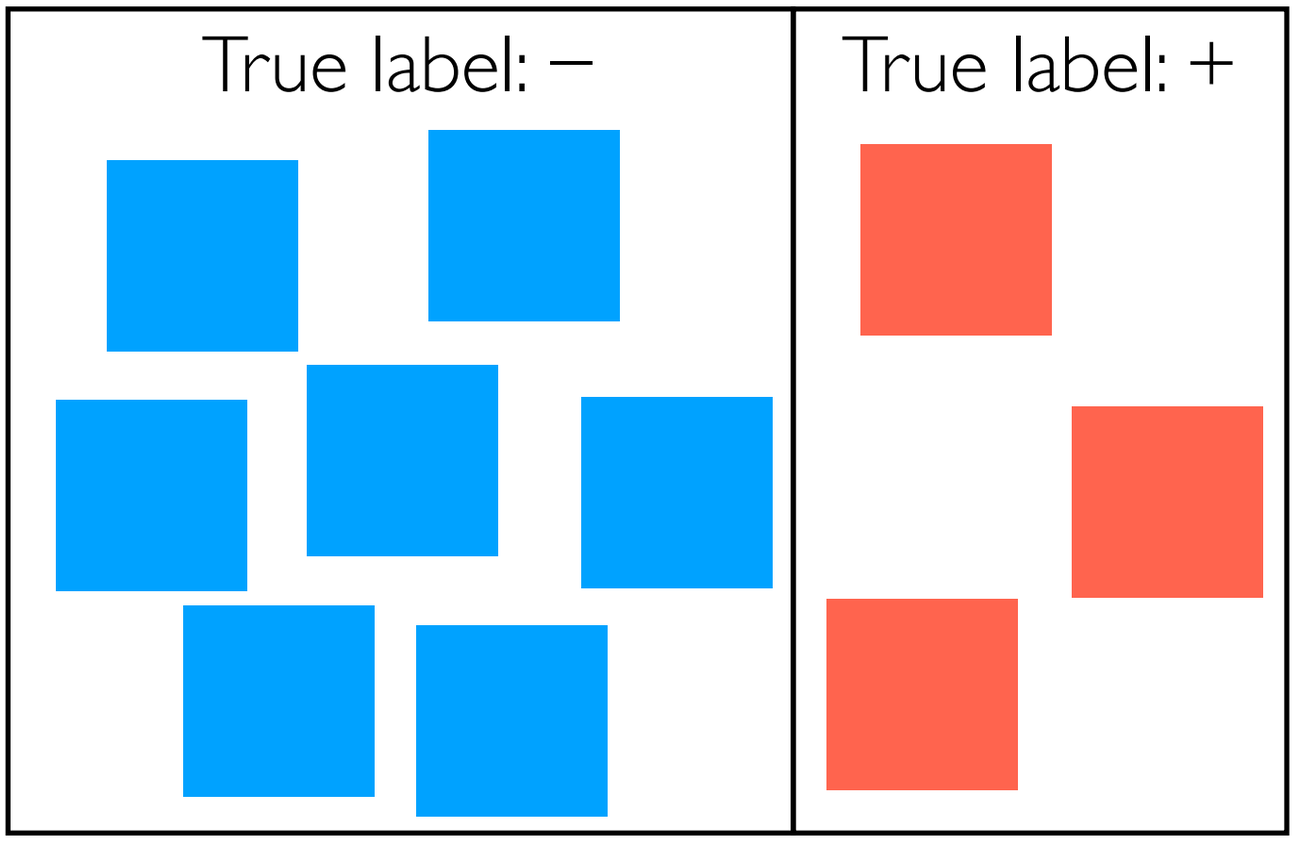


Different Ways to Measure Accuracy

Simplest way:

- **Raw error rate:** fraction of predicted labels that are wrong (this was in our cross validation example earlier)

In "binary" classification (there are 2 labels such as spam/ham) when 1 label is considered "positive" and the other "negative":



Different Ways to Measure Accuracy

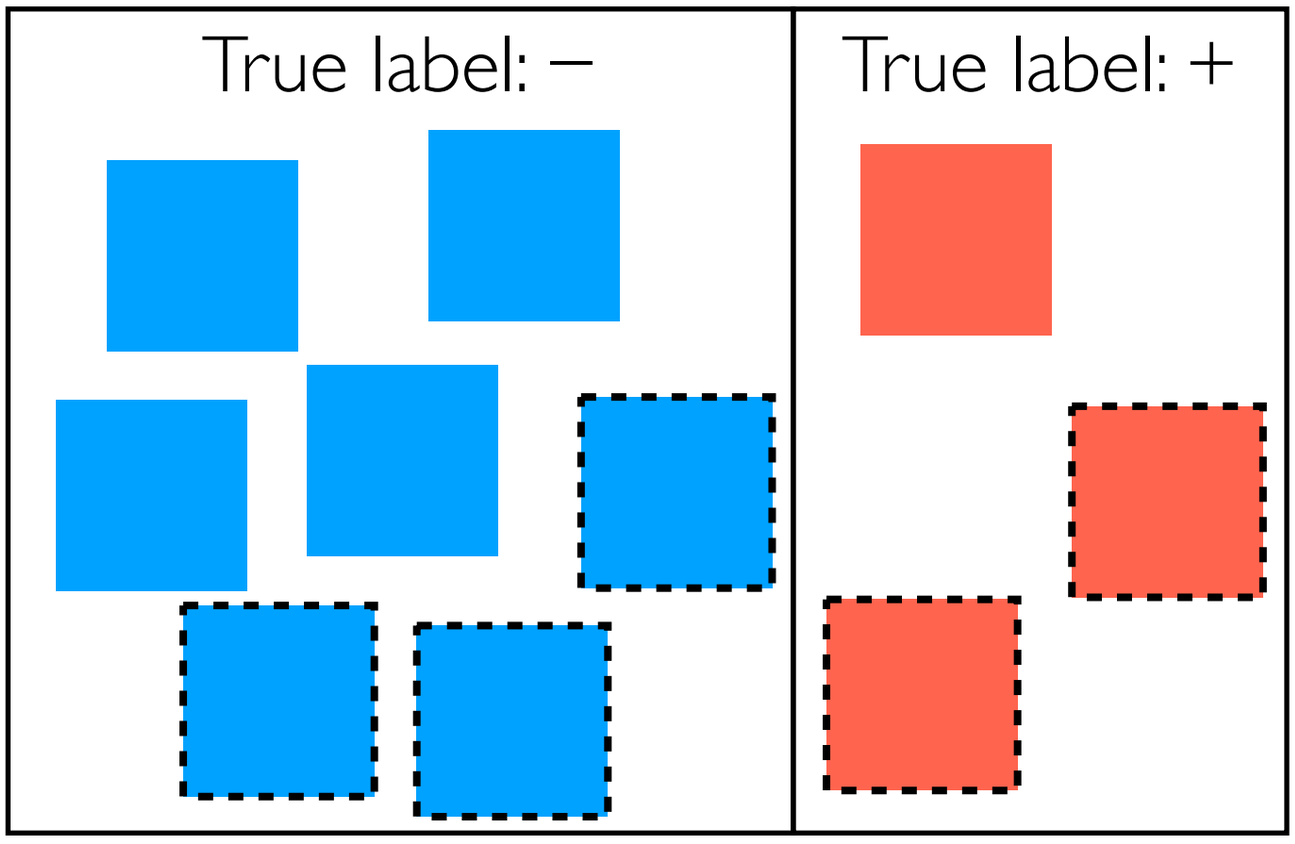
Simplest way:

- **Raw error rate:** fraction of predicted labels that are wrong (this was in our cross validation example earlier)

In "binary" classification (there are 2 labels such as spam/ham) when 1 label is considered "positive" and the other "negative":

Outlined in dotted black: predicted label +

(all other points predicted to be −)



Recall/True Positive Rate: fraction of dotted line in true label + = 2/3

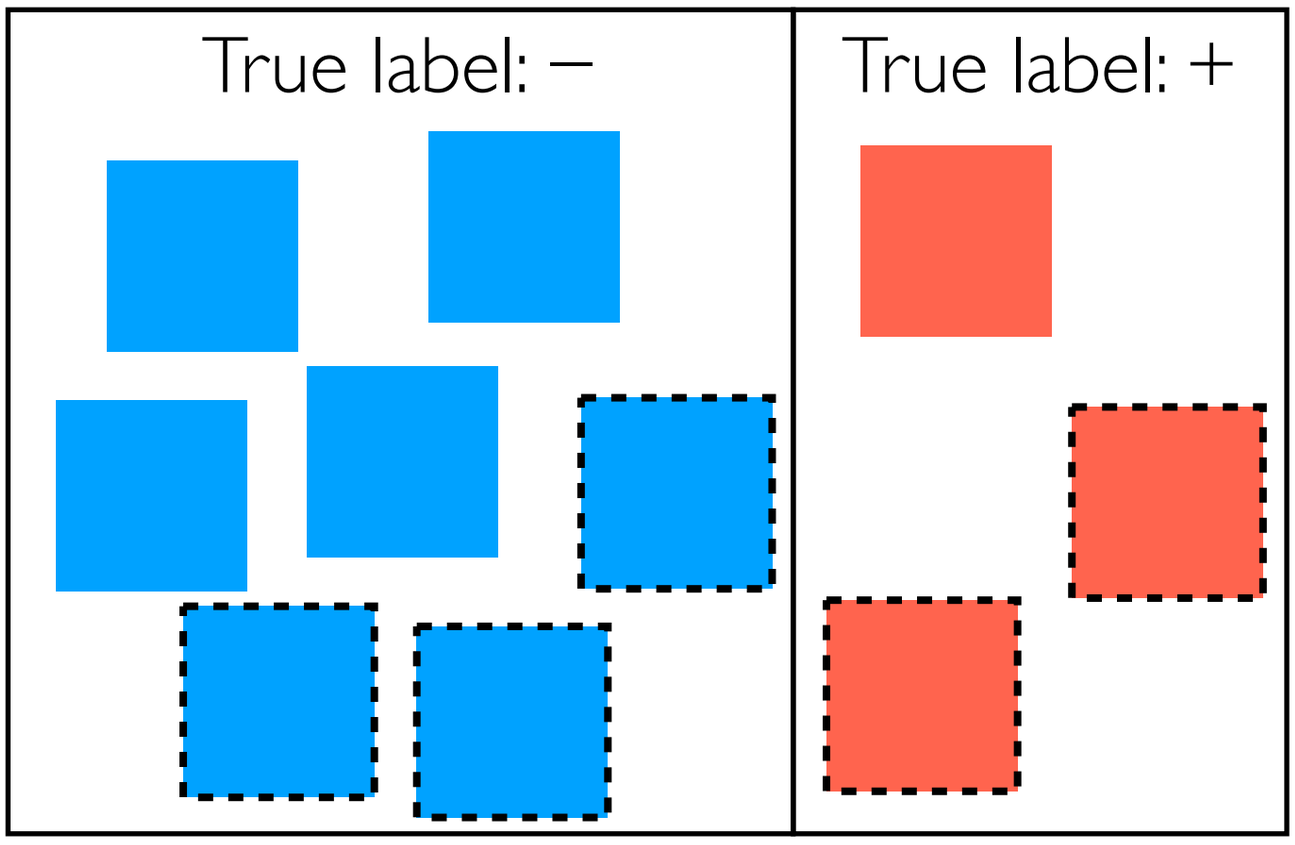
Precision: fraction of + in dotted line

- **Raw error rate:** fraction of predicted labels that are wrong (this was in our cross validation example earlier)

In "binary" classification (there are 2 labels such as spam/ham) when 1 label is considered "positive" and the other "negative":

Outlined in dotted black: predicted label +

(all other points predicted to be −)



Recall/True Positive Rate: fraction of dotted line in true label + = 2/3

Precision: fraction of + in dotted line = 2/5

False Positive Rate: fraction of dotted line in true label − = 3/7

F1 score: $\frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} = 1/2$

Basic Predictive Data Analytics Ideas

Demo

Deep Learning

Extremely useful in practice:

- Near human level image classification (including handwritten digit recognition)
- Near human level speech recognition
- Improvements in machine translation, text-to-speech
- Self-driving cars
- *Better* than humans at playing Go

Google DeepMind's AlphaGo vs Lee Sedol, 2016

GAMING TECH ARTIFICIAL INTELLIGENCE

16

DeepMind's StarCraft 2 AI is now better than 99.8 percent of all human players

AlphaStar is now grandmaster level in the real-time strategy game

By Nick Statt | @nickstatt | Oct 30, 2019, 2:00pm EDT

SHARE

Turing Award Won by 3 Pioneers in Artificial Intelligence

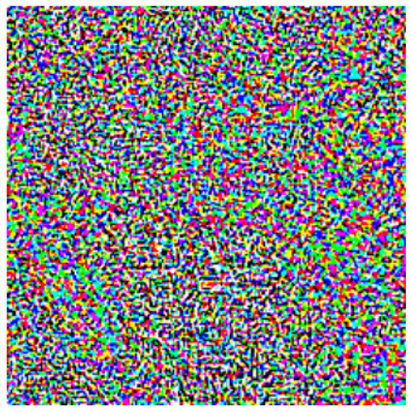
From left, Yann LeCun, Geoffrey Hinton and Yoshua Bengio. The researchers worked on key developments for neural networks, which are reshaping how computer systems are built. From left, Facebook, via Associated Press; Aaron Vincent Elkaim for The New York Times; Chad Buchanan/Getty Images

By Cade Metz

March 27, 2019

Is it all hype?

$+ .007 \times$



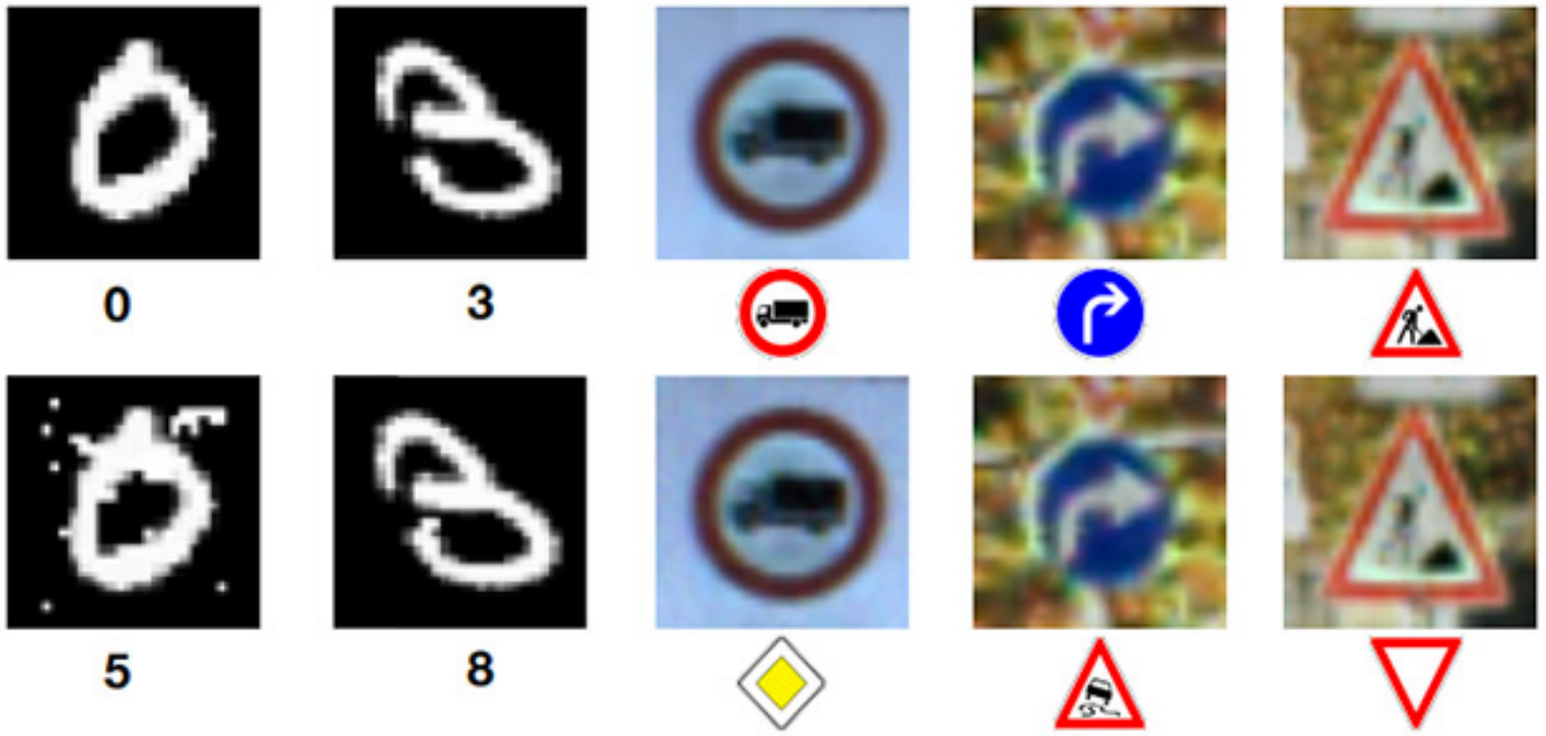
$=$

panda
~58% confidence

adversarial
noise

gibbon
~99% confidence

Source: Goodfellow, Shlens, and Szegedy. Explaining and Harnessing Adversarial Examples. ICLR 2015.



Source: Papernot et al. Practical Black-Box Attacks against Machine Learning. Asia Conference on Computer and Communications Security 2017.

Source: https://www.cc.gatech.edu/news/611783/erasing-stop-signs-shapeshifter-shows-self-driving-cars-can-still-be-manipulated

Source: Gizmodo article “This Neural Network's Hilariously Bad Image Descriptions Are Still Advanced AI”. September 16, 2015. (They’re using the NeuralTalk image-to-caption software.)

GENERAL

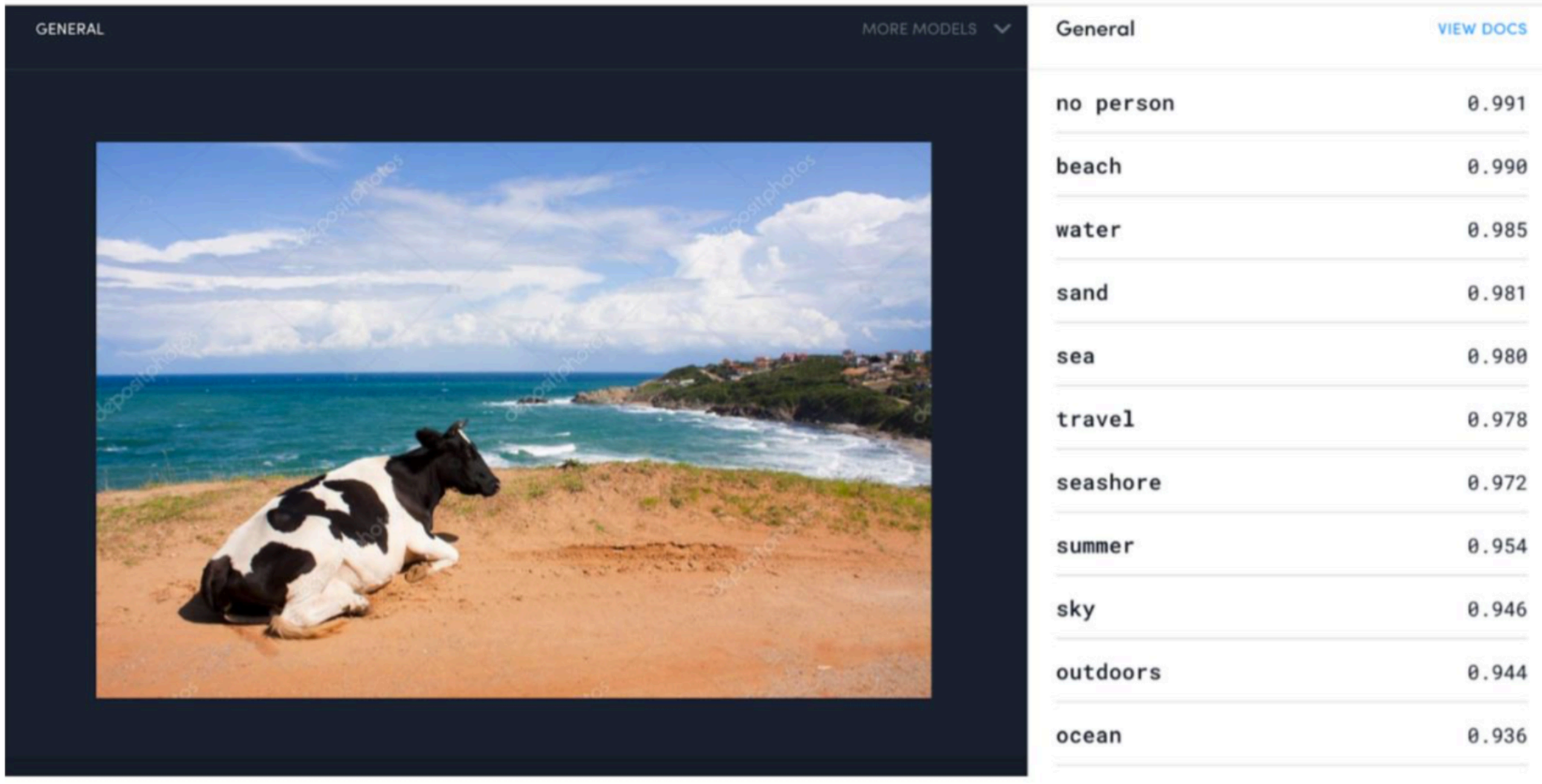
MORE MODELS

General

VIEW DOCS

cow	0.992
cattle	0.983
mammal	0.979
grass	0.978
livestock	0.966
farm	0.964
landscape	0.963
pasture	0.954
grassland	0.949
agriculture	0.948
no person	0.945

Source: Pietro Perona



cow is not among top objects found!

Source: Pietro Perona

GENERAL FACE NSFW COLOR MORE MODELS

General VIEW DOCS

PREDICTED CONCEPT	PROBABILITY
group	0.979
adult	0.977
people	0.976
furniture	0.960
room	0.957
business	0.903
indoors	0.901
man	0.896
seat	0.895

elephant is not among top objects found!

Source: David Lopez-Paz

Michael Jordan Follow

Michael I. Jordan is a Professor in the Department of Electrical Engineering and Computer Sciences and the Department of Statistics at UC Berkeley.

Apr 18 · 16 min read

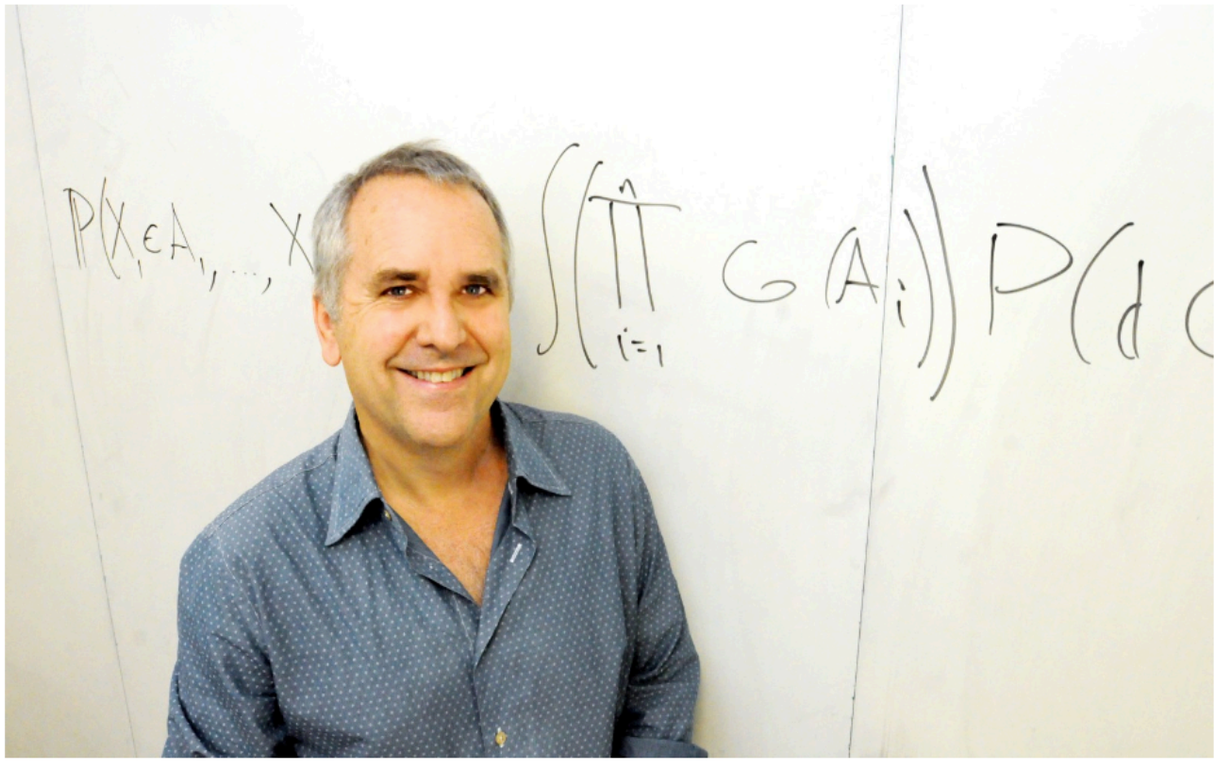


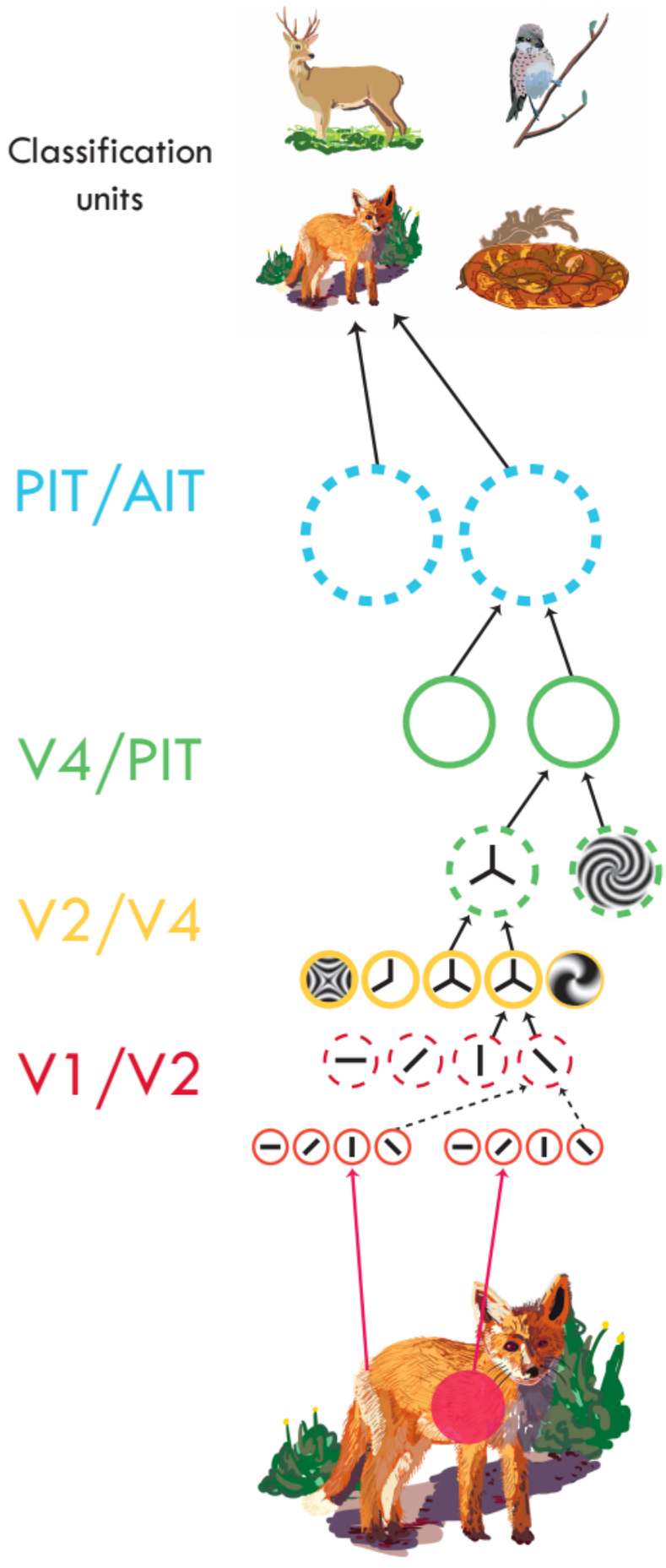
Photo credit: Peg Skorpinski

Artificial Intelligence—The Revolution Hasn't Happened Yet

Artificial Intelligence (AI) is the mantra of the current era. The phrase is intoned by technologists, academicians, journalists and venture capitalists

https://medium.com/@mijordan3/artificial-intelligence-the-revolution-hasnt-happened-yet-5e1d5812e1e7

What is deep learning?



Slide by Phillip Isola

Serre, 2014

Basic Idea

Brain/Machine

"clown fish"

Slide by Phillip Isola

Object Recognition

Edges

Texture

Colors

Segments

Parts

"clown fish"

Feature extractors

Classifier

Slide by Phillip Isola

Object Recognition

Learned

Edges

Texture

Colors

Segments

Parts

"clown fish"

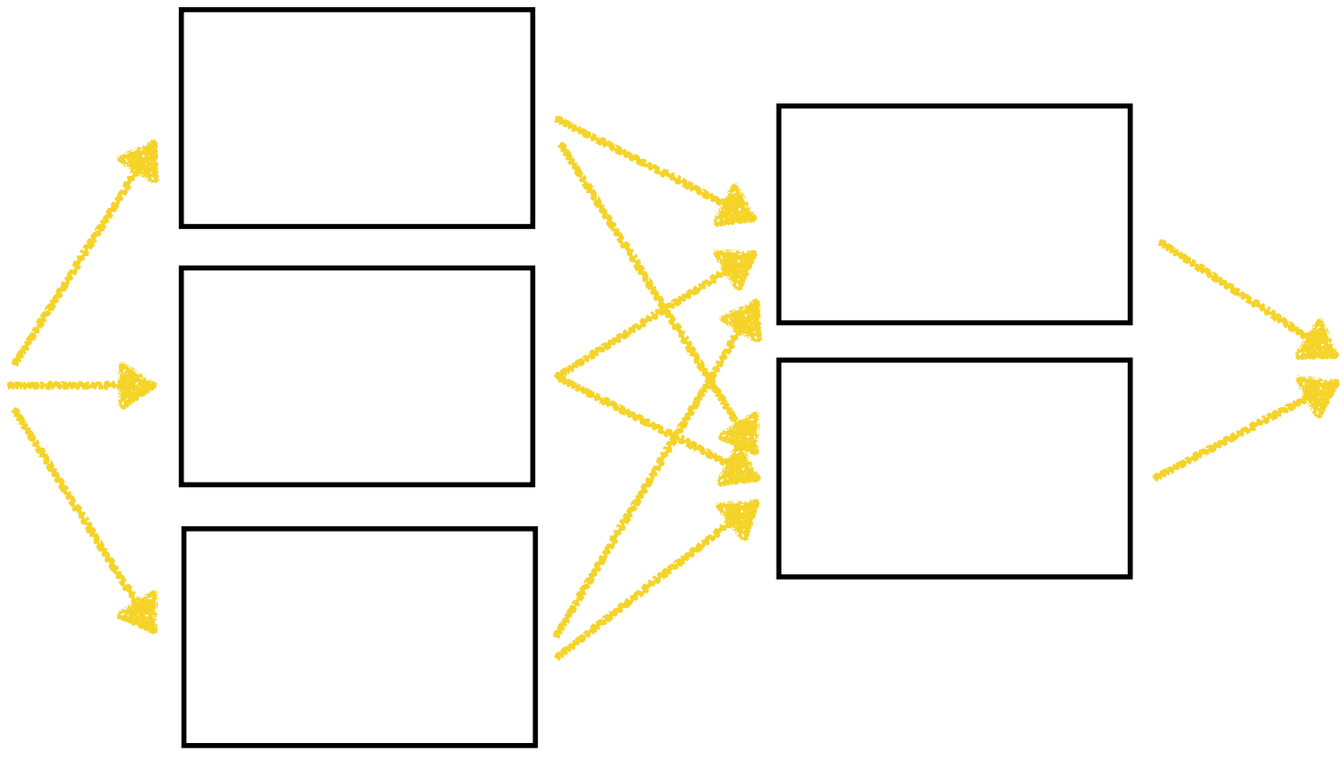
Feature extractors

Classifier

Slide by Phillip Isola

Neural Network

Learned

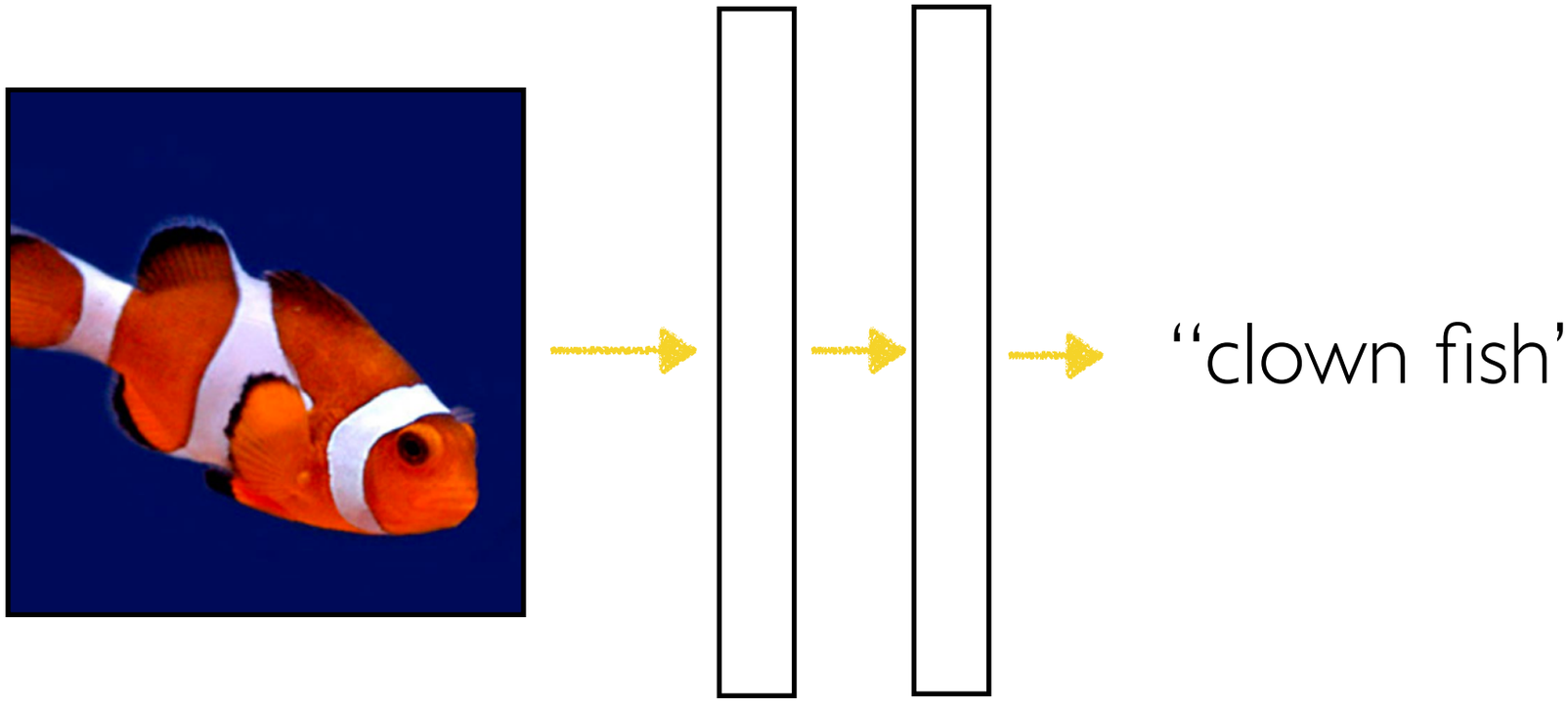


"clown fish"

Slide by Phillip Isola

Neural Network

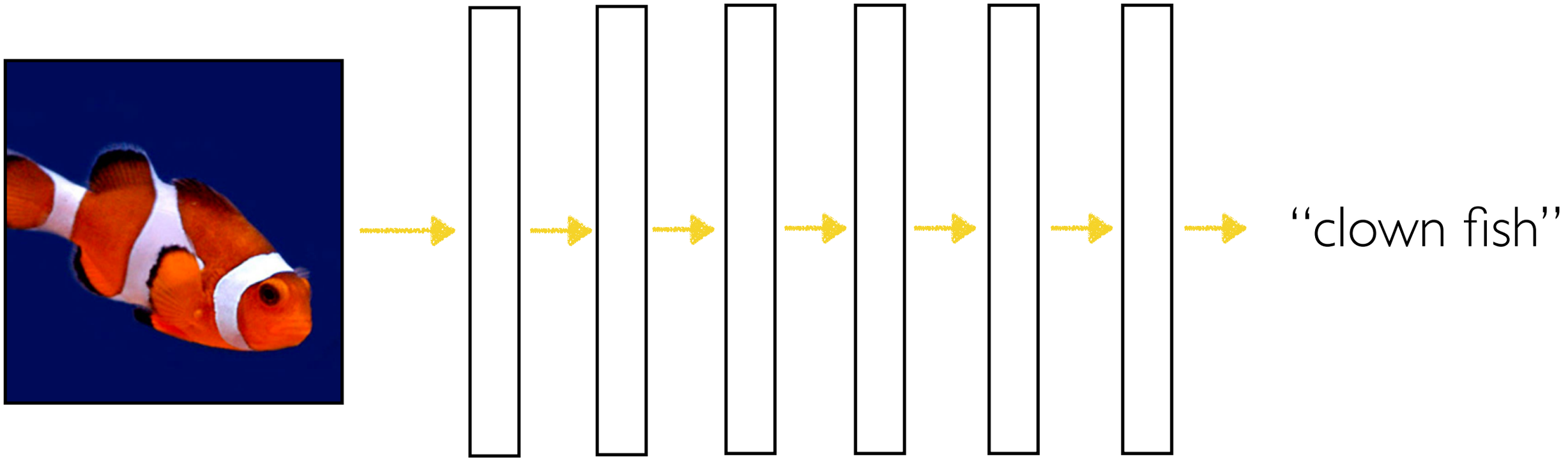
Learned



Slide by Phillip Isola

Deep Neural Network

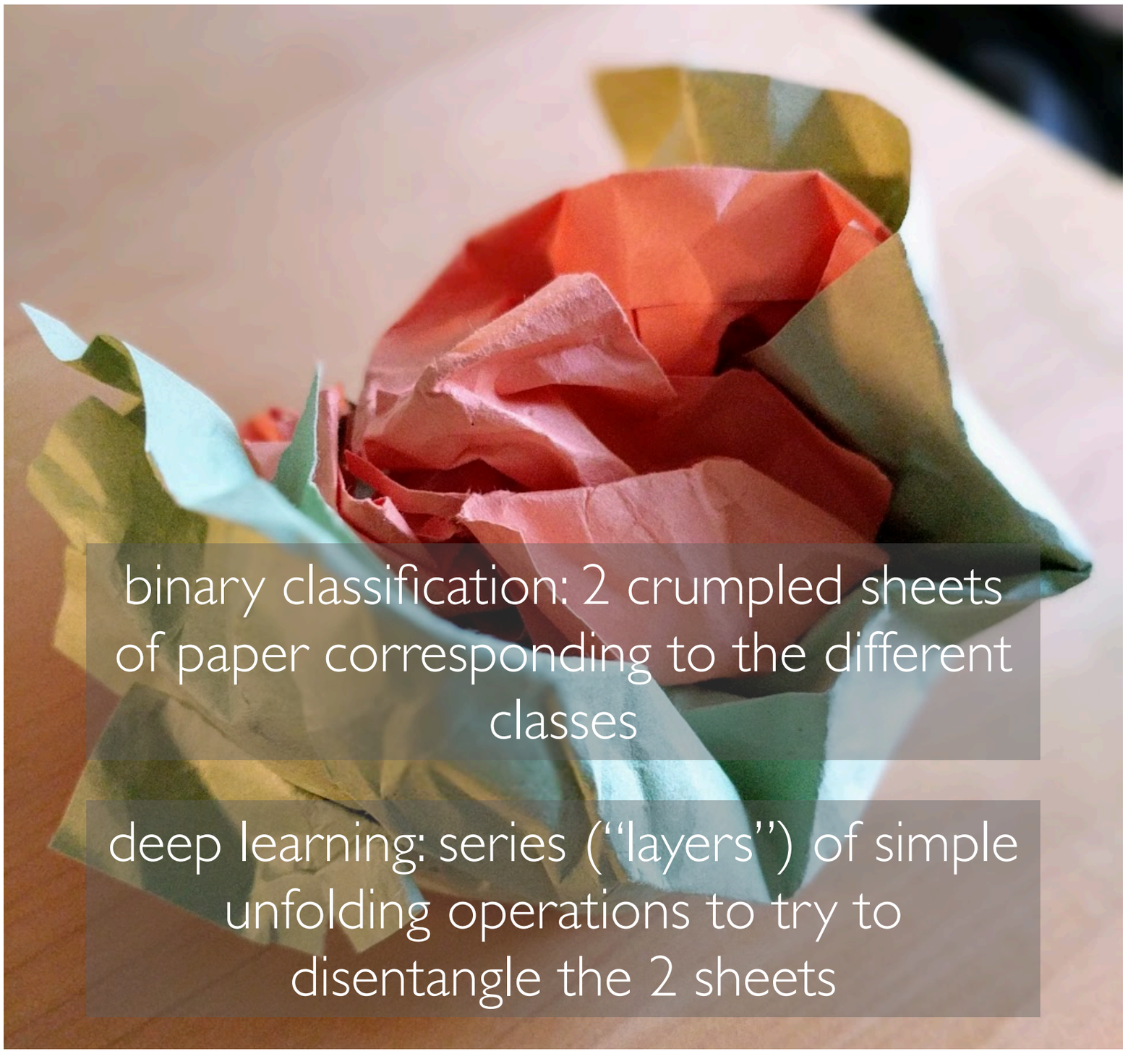
Learned



Deep learning just refers to learning deep neural nets

Slide by Phillip Isola

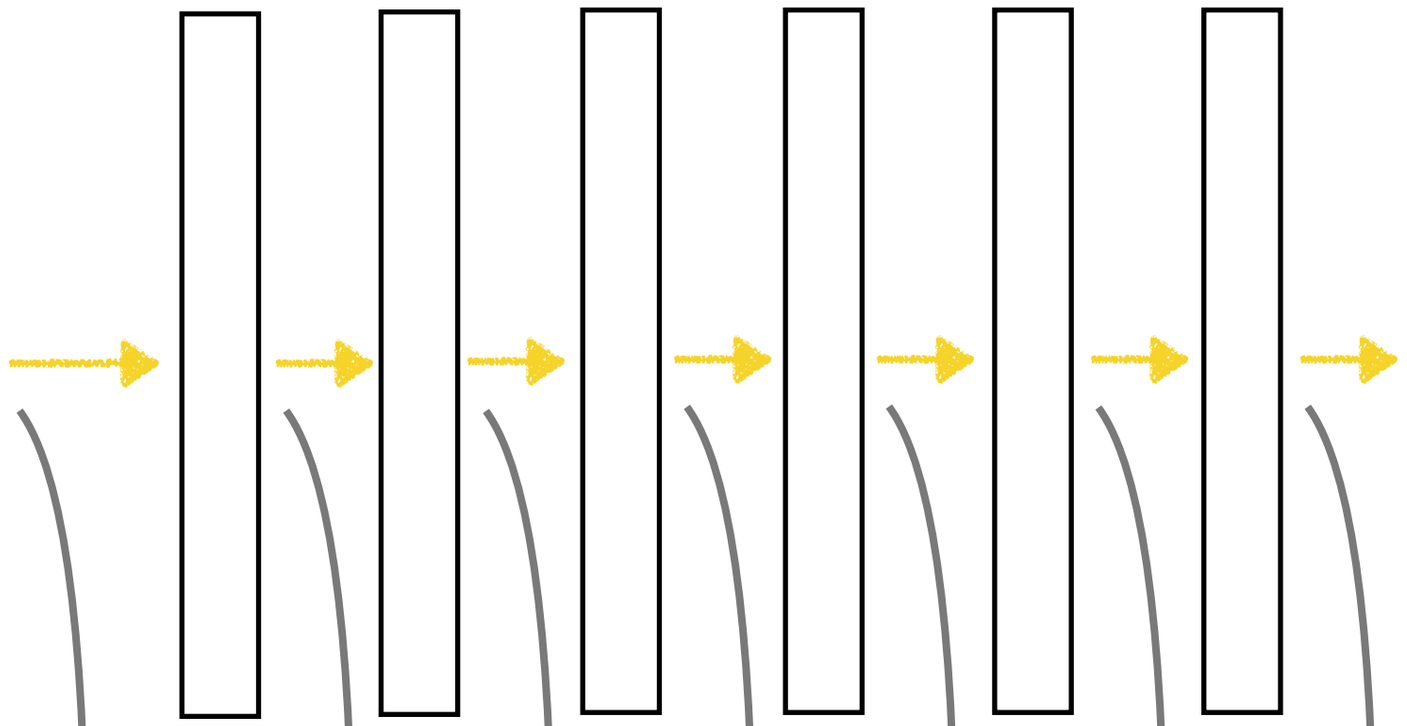
Crumpled Paper Analogy



Analogy: Francois Chollet, photo: George Chen

Representation Learning

Each layer's output is *another way we could represent the input data*



Learned

"clown fish"

Visualize (e.g., t-SNE)

Visualize

Visualize

Visualize

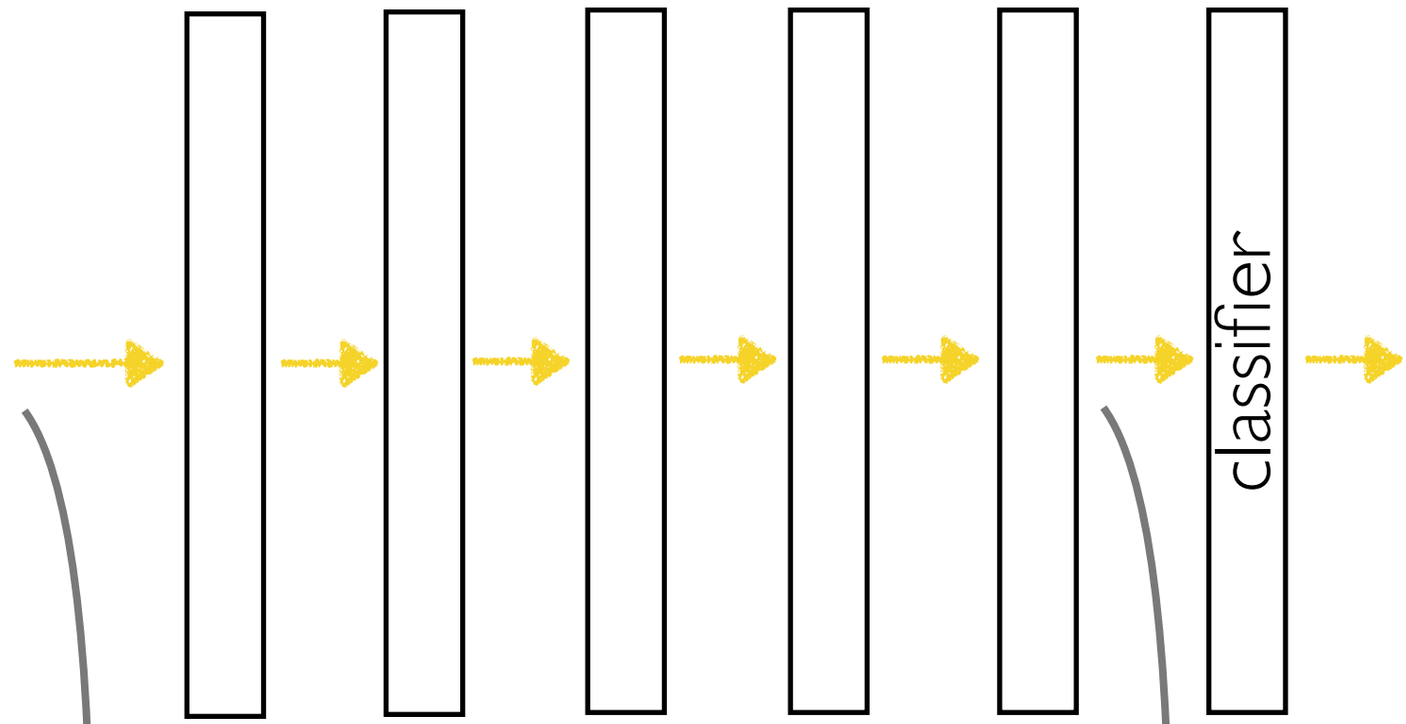
Visualize

Visualize

Visualize

Representation Learning

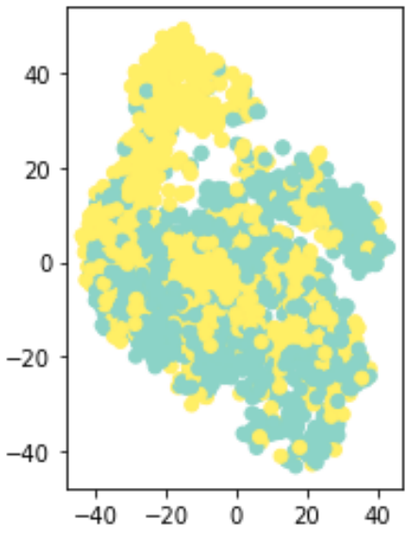
Each layer's output is *another way we could represent the input data*



Learned

"clown fish"

Visualize (e.g., t-SNE)



Visualize