Computer Vision

INTRODUCTION

Jison Hsu
Artificial Vision Laboratory
Taiwan Tech

Contents

How to Make an Image Classifier - Intro to Deep Learning #6

https://www.youtube.com/watch?v=cAICT4Al5Ow[8:44]

MIT Deep Learning Basics: Introduction and Overview

- OUTLINE:
- 0:00 - Introduction
- 0:53 - Deep learning in one slide
- 4:55 - History of ideas and tools
- 9:43 - Simple example in TensorFlow
- 11:36 - TensorFlow in one slide
- 13:32 - Deep learning is representation learning
- 16:02 - Why deep learning (and why not)
- 22:00 - Challenges for supervised learning
- 38:27 - Key low-level concepts
- 46:15 - Higher-level methods
- 1:06:00 - Toward artificial general intelligence

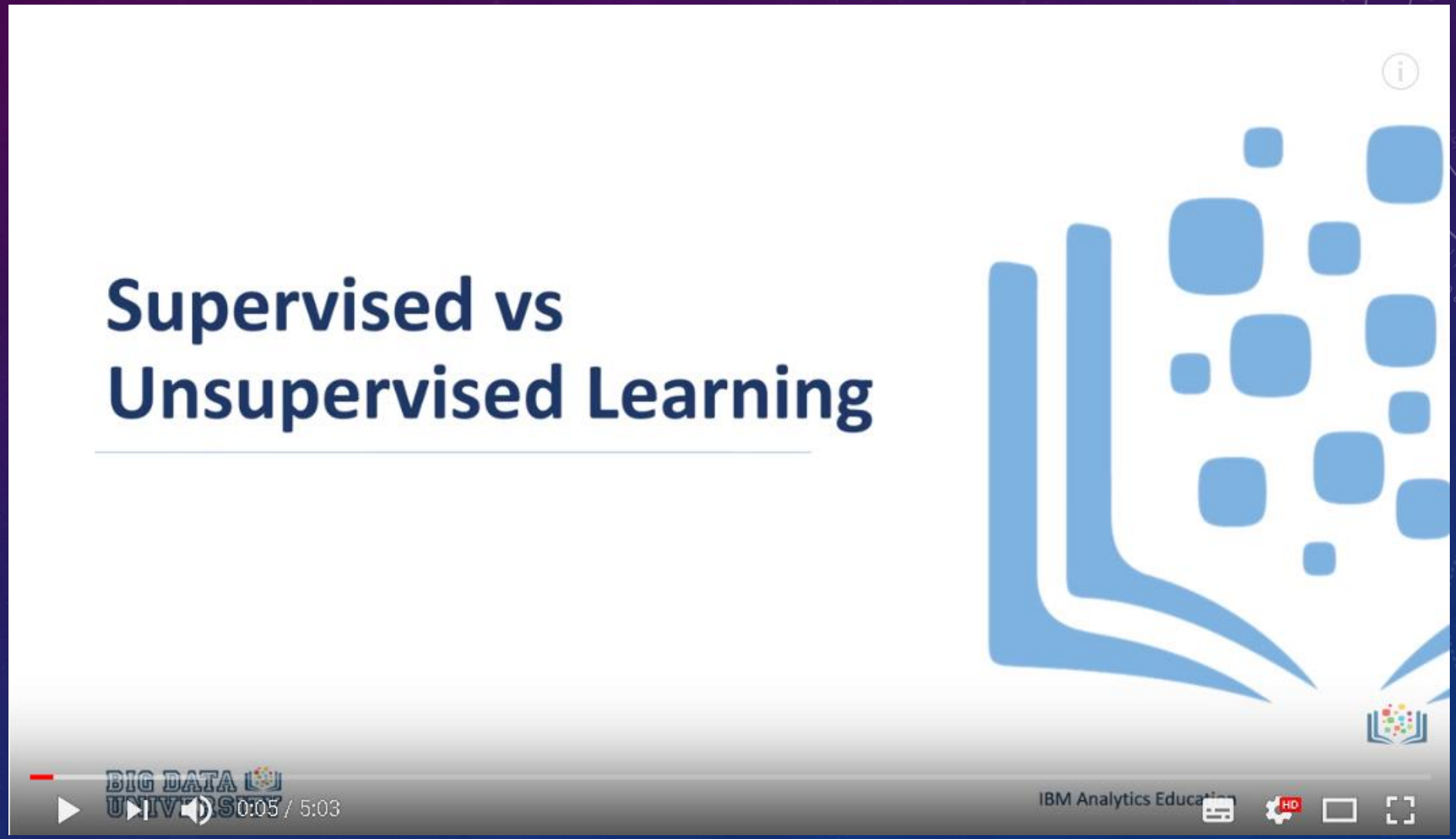
MIT Deep Learning Basics: Introduction and Overview

https://youtu.be/O5xeyoRL95U [1:08:05]

Deep Learning: Practice and Trends (NIPS 2017 Tutorial, parts I & II)

https://youtu.be/YJnddoa8sHk [1:44:24]

Supervised and Unsupervised Learning



https://www.youtube.com/watch?v=cfj6yaYE86U [5:03]

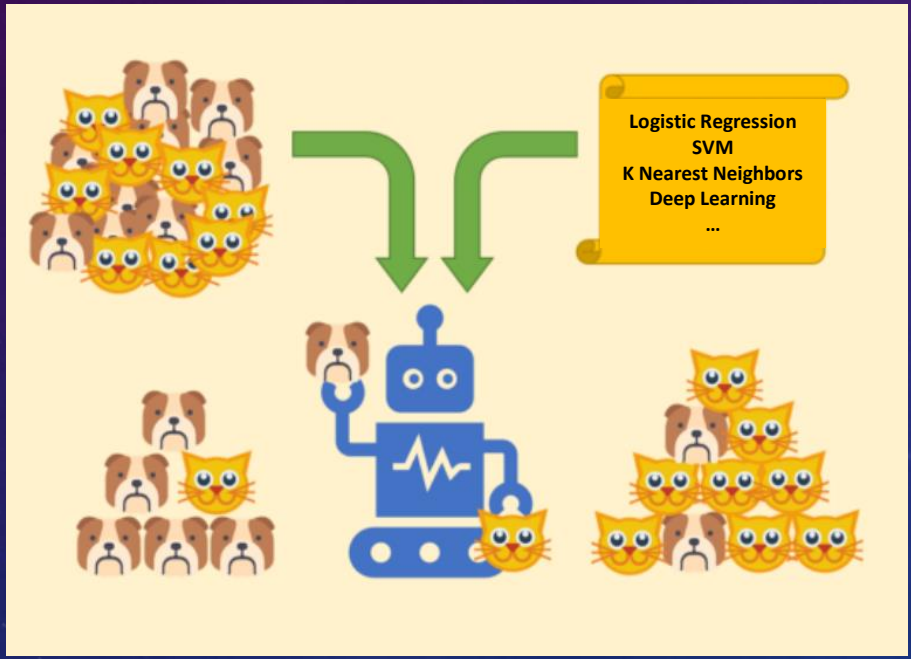
Classification Models

- Collection of training data
- Model built upon features (or more precisely, on the feature space)
  - Happy face = f (f refers to the features extracted from a face)
- Model-based prediction when new data is given

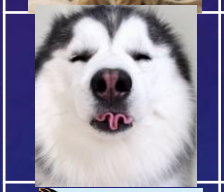
High Dimensional Feature Space

Binary Classification

Binary classification is the task of classifying the elements of a set into two groups on the basis of a classification rule.



Common Binary Classification Task: dogs and cats

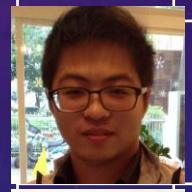
	Prob. (Dog)	Prob. (Cat)	Label (G.T)
	0.2311	0.7689	1
	0.7842	0.2158	0
	0.8913	0.1087	0
	0.0182	0.9818	1
...	...	...	...

Ref: https://towardsdatascience.com/analytics-building-blocks-binary-classification-d205890314fc

Multiclass Classification

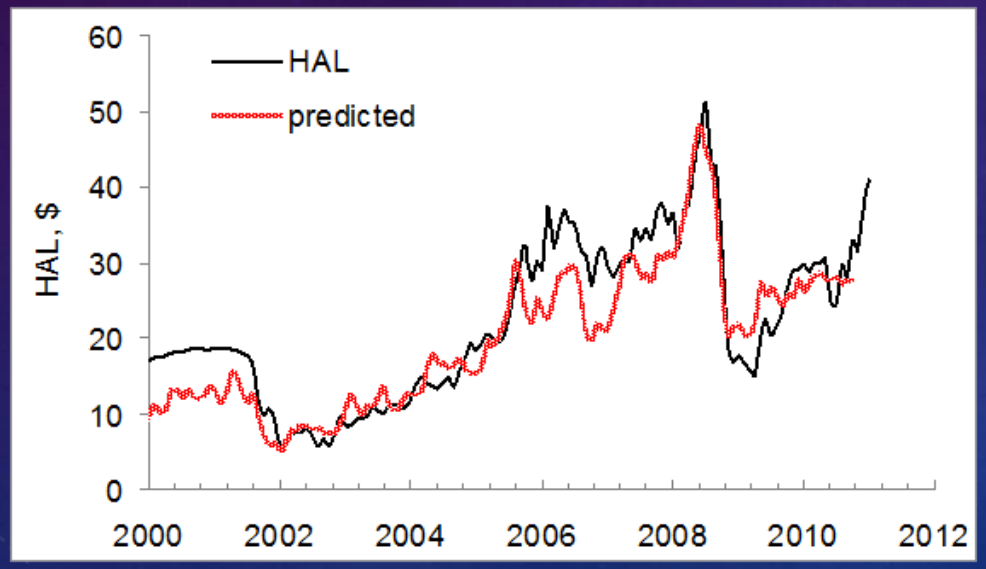
"Multinomial classification is the problem of classifying instances into one of three or more classes. (classifying instances into one of two classes is called binary classification)"

Multiclass Classification: more than two labels

	Prob. (Dog)	Prob. (Cat)	Prob. (Bird)	…	Pred.	Label
	0.2311	0.3271	0.4312	…	Fish	1
	0.7842	0.2158	0.2277	…	Dog	0
	0.1913	0.4487	0.3752	…	Cat	3
	0.2182	0.9818	0.4312	…	Bird	6
…	…	…	…	…	…	…

Regression Models

- Collection of training data;
- Regression model built upon feature space
  - Stock value = ***G*** (previous closing, financial indices, profits, revenues, ….)
- Make a prediction given the known features.



Sample of the stock prediction

TSMC 2330

Linear Regression

The simplest form of a linear regression problem can be defined by the following sample equation:

$$\hat{Y} = aX + b + \mathrm{e}$$

where,

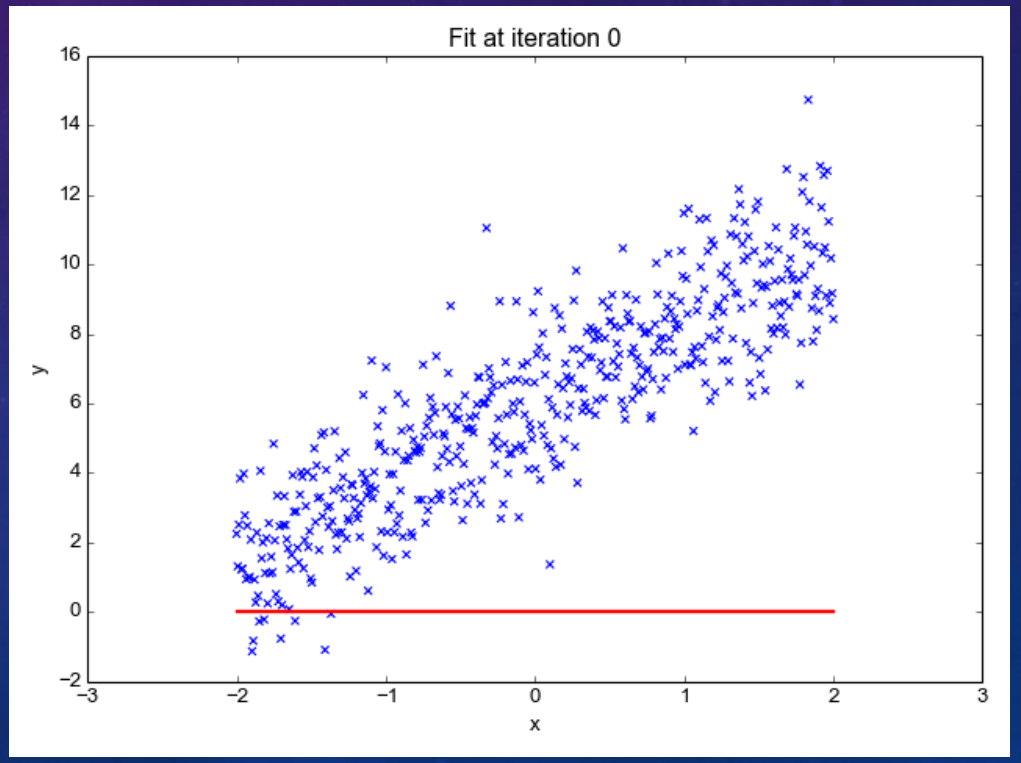
$\hat{Y}$ = Predicted value of Y

X = Independent variable

a = Slope coefficient based on best-fitting line

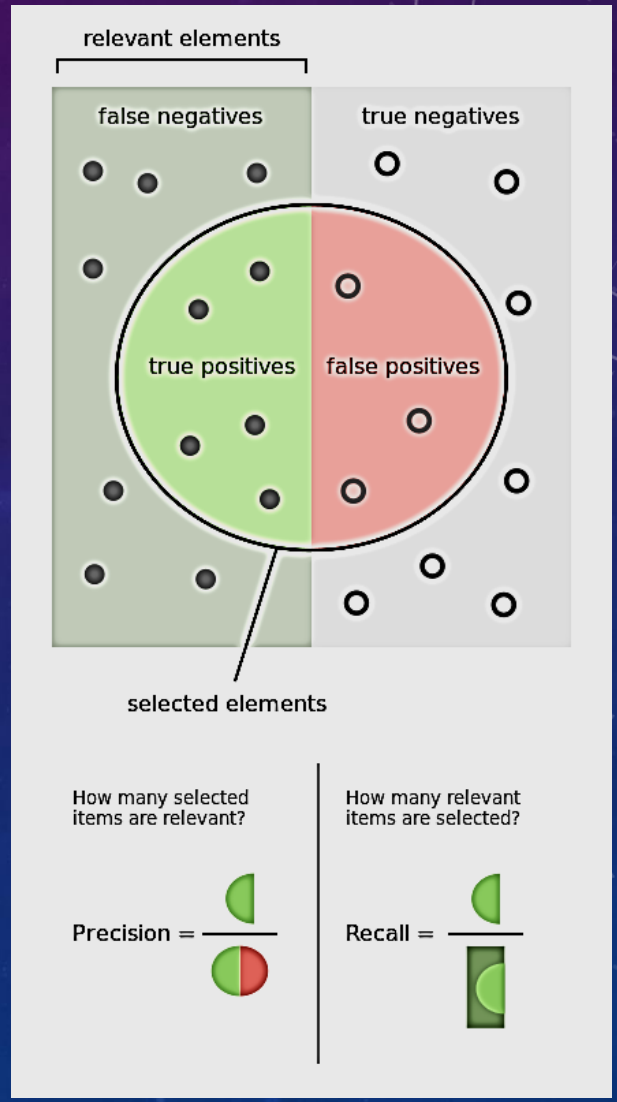
b = Y-axis intercept

e = Noise



Precision and recall

- **Precision** is the fraction of retrieved instances that are relevant, while **Recall** is the fraction of relevant instances that are retrieved.
  - True positive (TP) = correctly identified
  - False positive (FP) = incorrectly identified
  - True negative (TN) = correctly rejected
  - False negative (FN) = incorrectly rejected



From Wikipedia Precision and Recall

Samples of Face Detection

TP = 5 FP = 1 FN = 3

Nine faces in a single image.

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{Precision} = \frac{5}{5+1} = 0.833$$

$$\text{Recall} = \frac{5}{5+3} = 0.625$$

Training and Testing Sets

- Training Set
  - A set in which data are known to a system for building classification/regression model.
  - For example, in a face recognition neural network, the face images used to train the network.
- Testing Set
  - A set in which data are unknown to a system for recognition.
  - For example, the face images to be recognized by the trained face recognition network.

But what is a Neural Network? | Deep learning, chapter 1

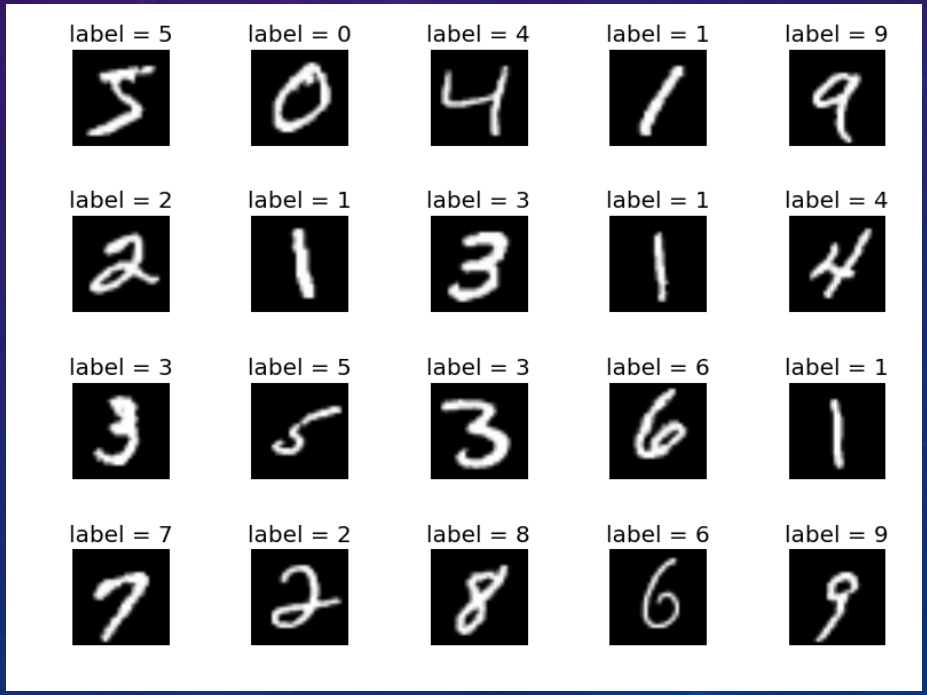
https://youtu.be/aircAruvnKk [19:13]

Gradient descent, how neural networks learn | Deep learning, chapter 2

https://youtu.be/IHZwWFHWa-w [21:00]

Example 1: Training LeNet on Mnist

- Mnist database is dataset for handwritten digits.
- The dataset consists of pair, “handwritten digit image” and “label”.
  - handwritten digit image: This is gray scale image with size 28 x 28 pixel.
  - label : This is actual digit number this handwritten digit image represents. It is either 0 to 9.



Sample

Le Net

- Yann LeCun and his collaborators developed a really good recognizer for handwritten digits by using back-propagation in a feed-forward net with:
  - Many hidden layers
  - Many maps of replicated units in each layer.
  - Pooling of the outputs of nearby replicated units.
  - A wide net that can cope with several characters at once even if they overlap.
  - A clever way of training a complete system, not just a recognizer.
- This net was used for reading ~10% of the checks in North America.
- Look the impressive demos of LENET at http://yann.lecun.com