DATA DRIVEN COMPUTATION AND LEARNING ALGORITHMS

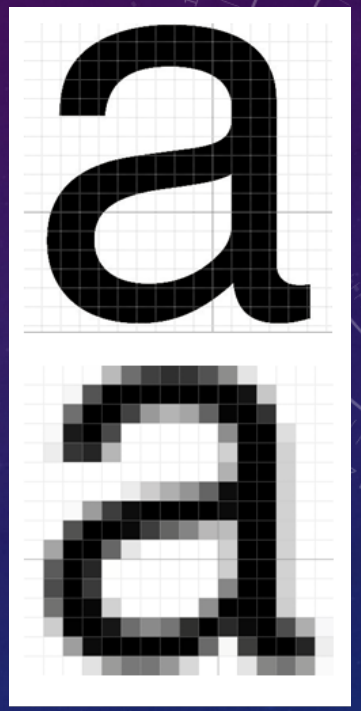
- Abuses, **Bias** and Blessings of Data

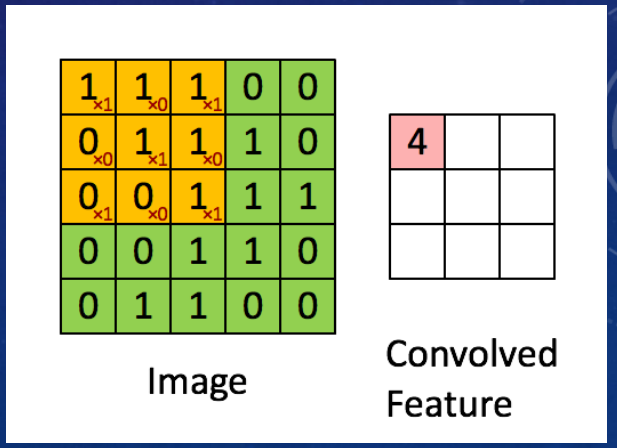
RULE-BASED AI

- Design and describe a rule or set of rules to distinguish these animals into their species

SMALL COMPLEX LEARNING CHALLENGE HAND WRITTEN DIGIT RECOGNITION

- Design template (mask)
- Compare template with various positions in image
  - Convolution





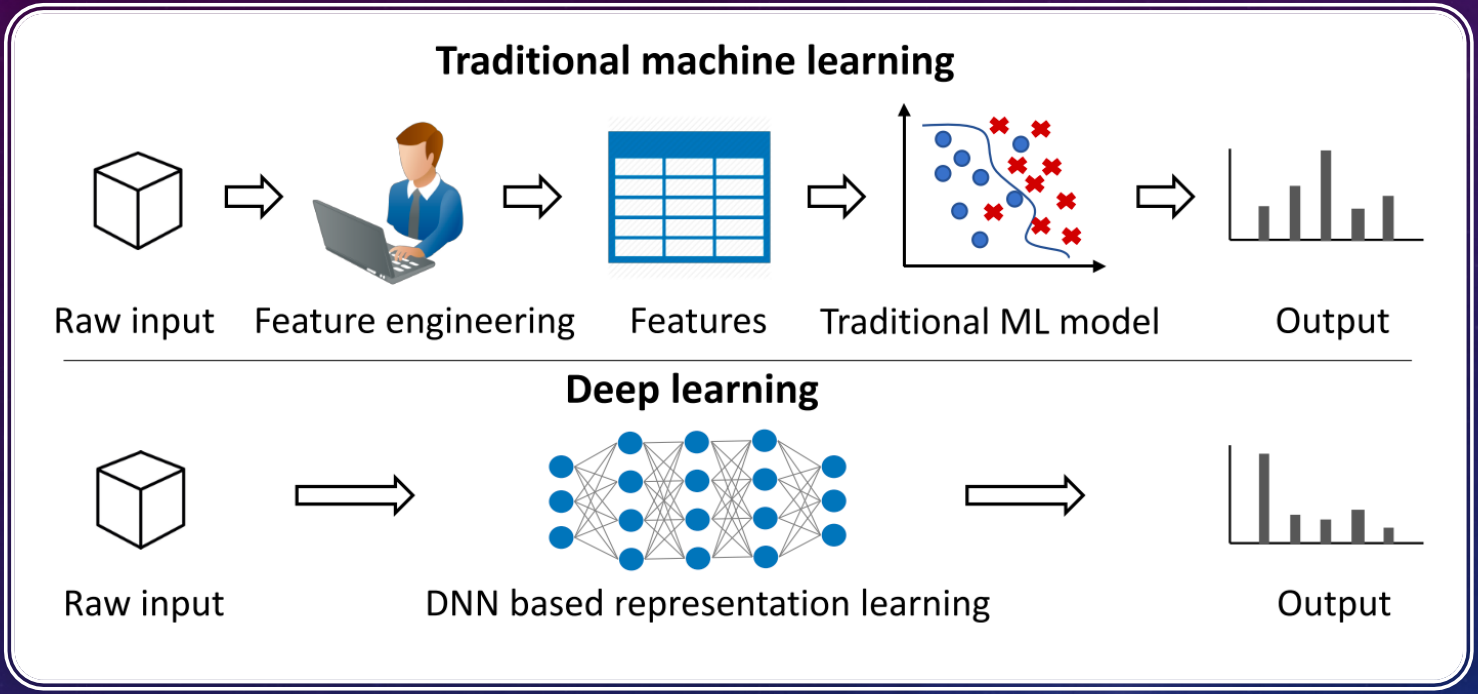
MACHINE LEARNING PIPELINE

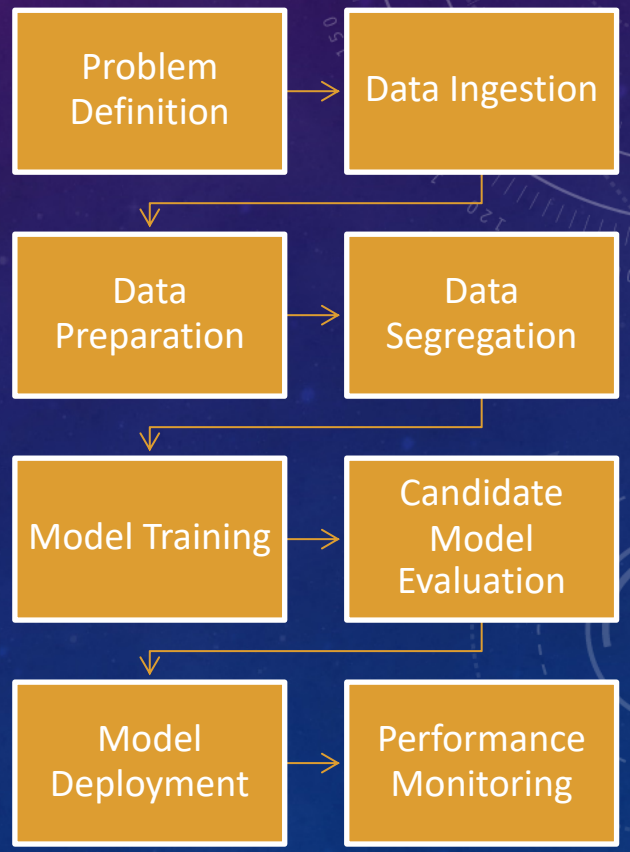
DATA **CONTAINS** PATTERNS

MACHINE LEARNING ALGORITHM **FINDS** PATTERNS

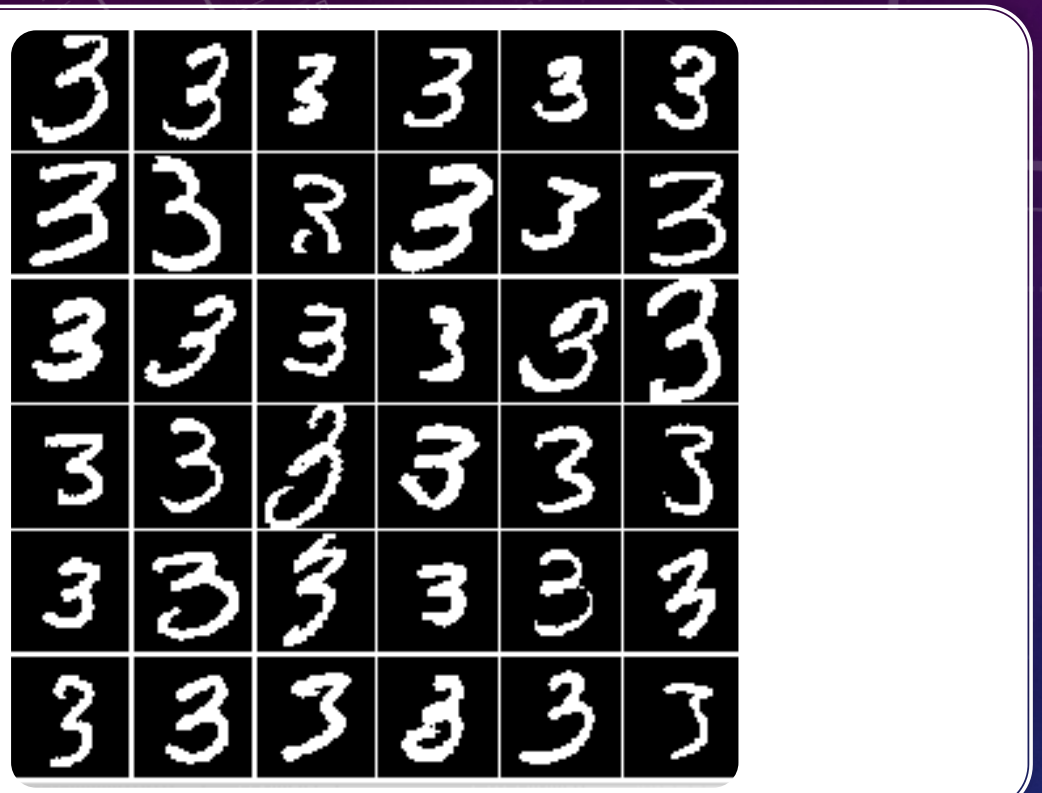
MODEL **RECOGNIZES** PATTERNS

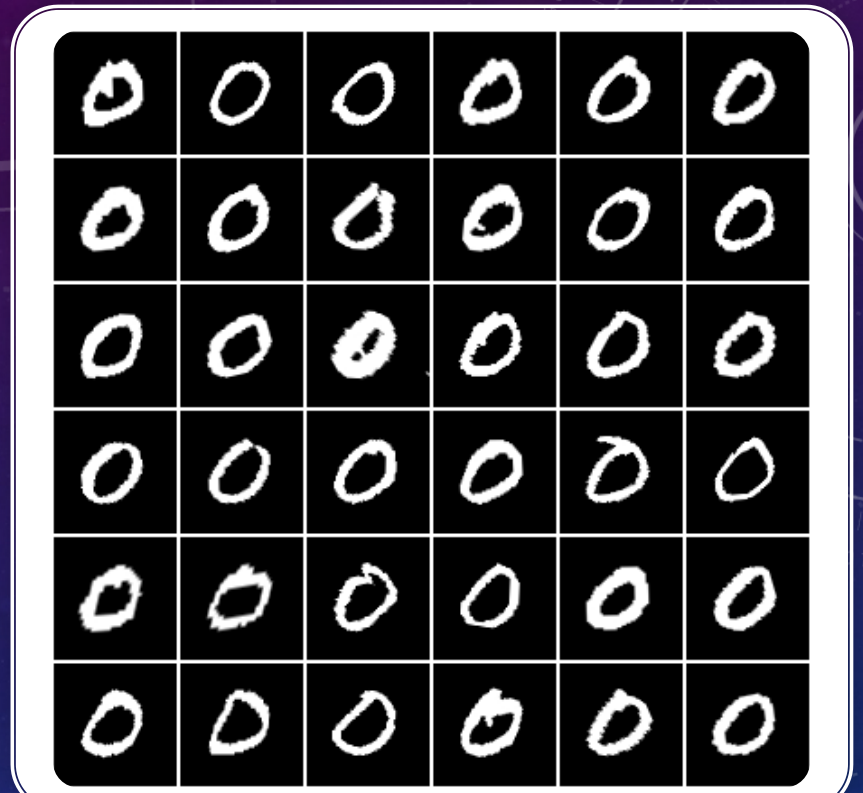
AI / ML PIPELINE





M. Du, N. Liu, and X. Hu, "Techniques for Interpretable Machine Learning," *Commun. ACM*, Jul. 2019. http://arxiv.org/abs/1808.00033





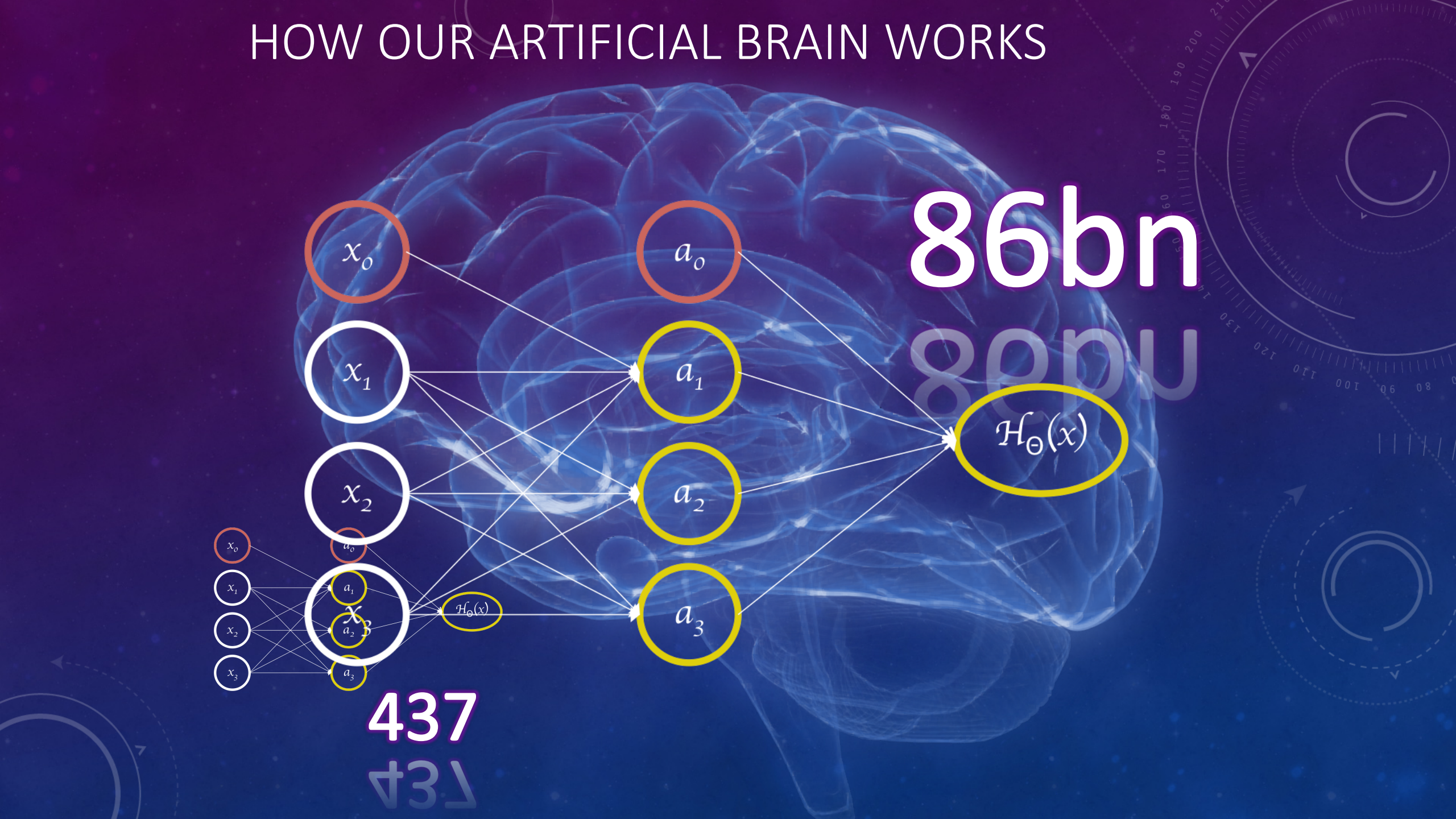
SMALL COMPLEX LEARNING CHALLENGE
HAND WRITTEN DIGIT RECOGNITION

MINI AI BRAIN V'S THE REAL THING

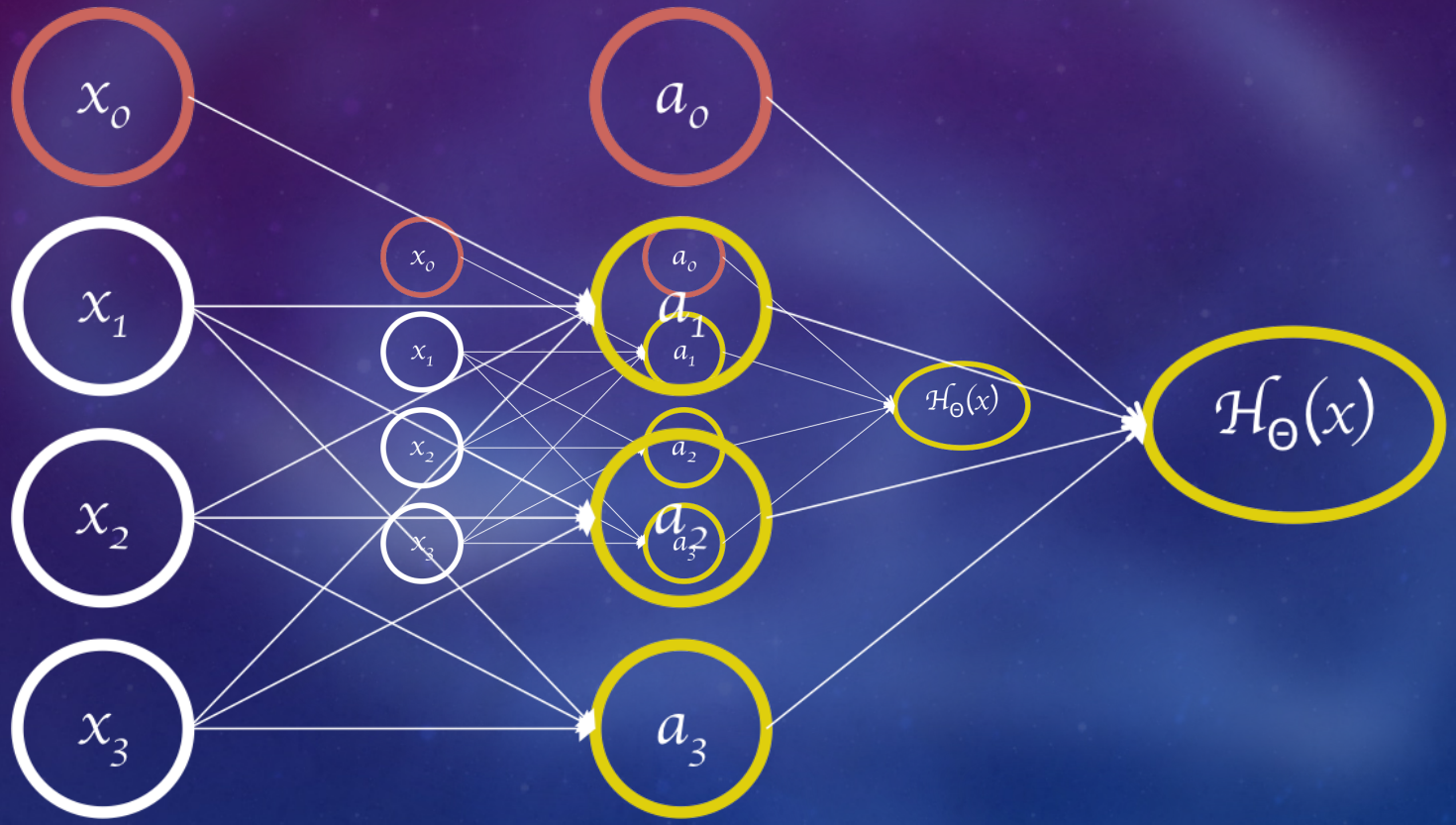
86bn

437

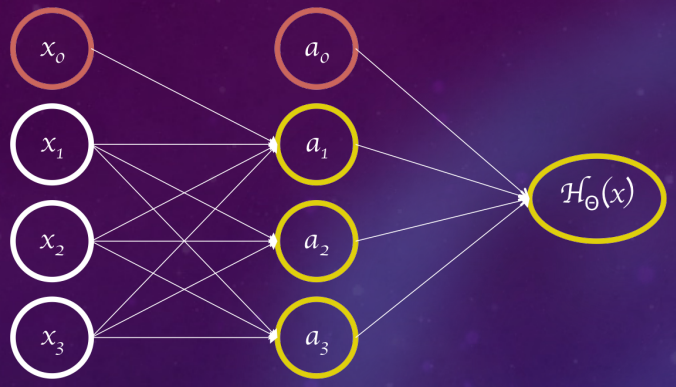
HOW OUR ARTIFICIAL BRAIN WORKS

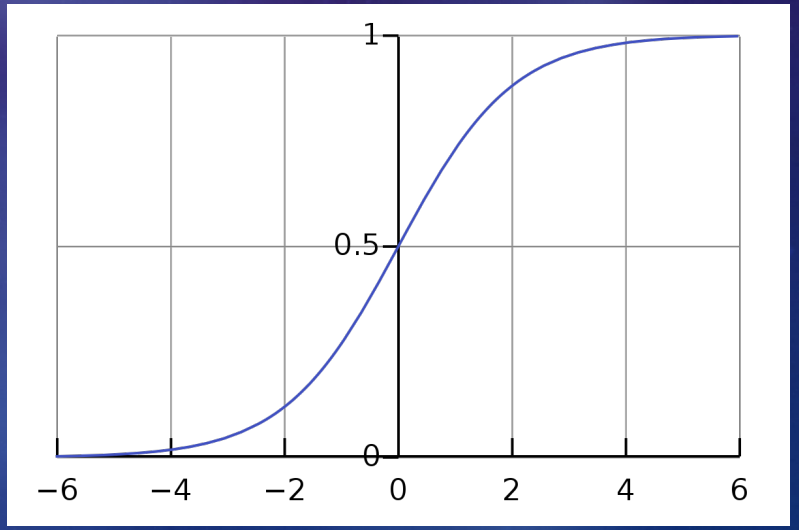


SWITCHING OUR NEURONS ON & OFF



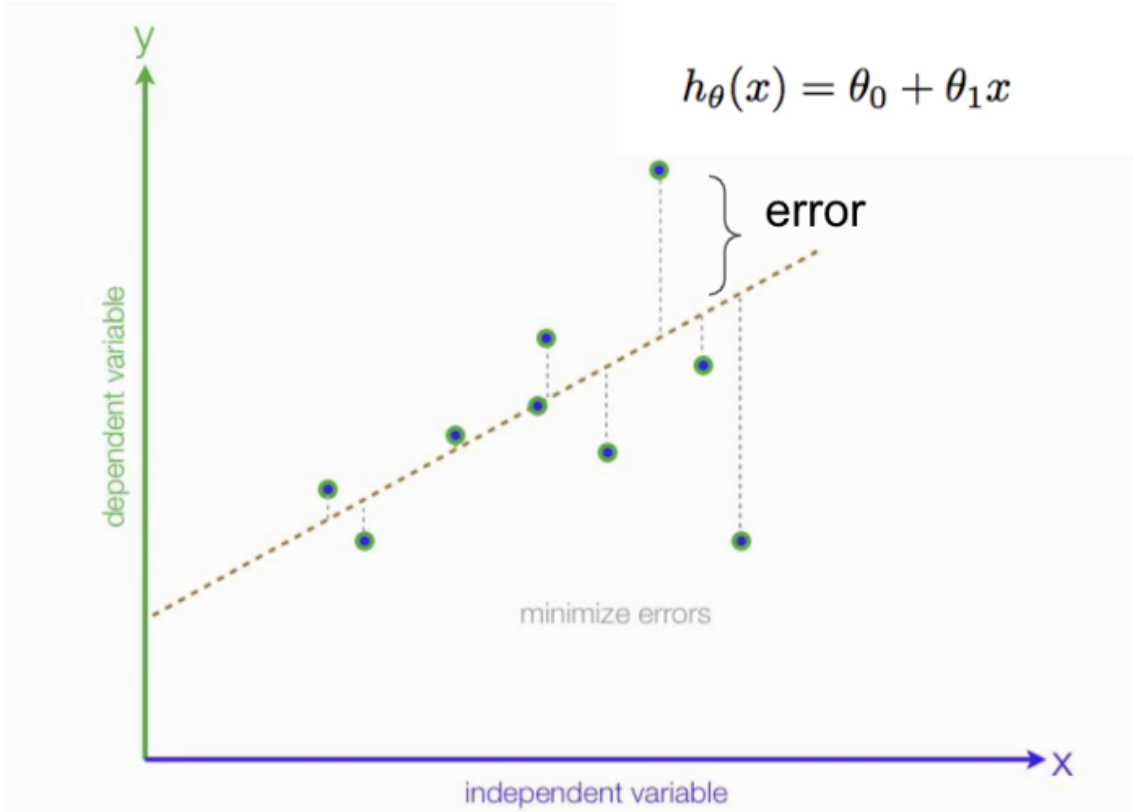
SWITCHING OUR NEURONS ON & OFF





SUPERVISED LEARNING

- 5000 examples 5000 known answers
  - Plug them in
  - Measure a cost function
  - Gradient decent to optimize a convex function
  - Iterate round and round to complete optimization



$$h_\theta(x) = \theta_0 + \theta_1 x$$

Hypothesis:

$$h_\theta(x) = \theta_0 + \theta_1 x$$

Parameters:

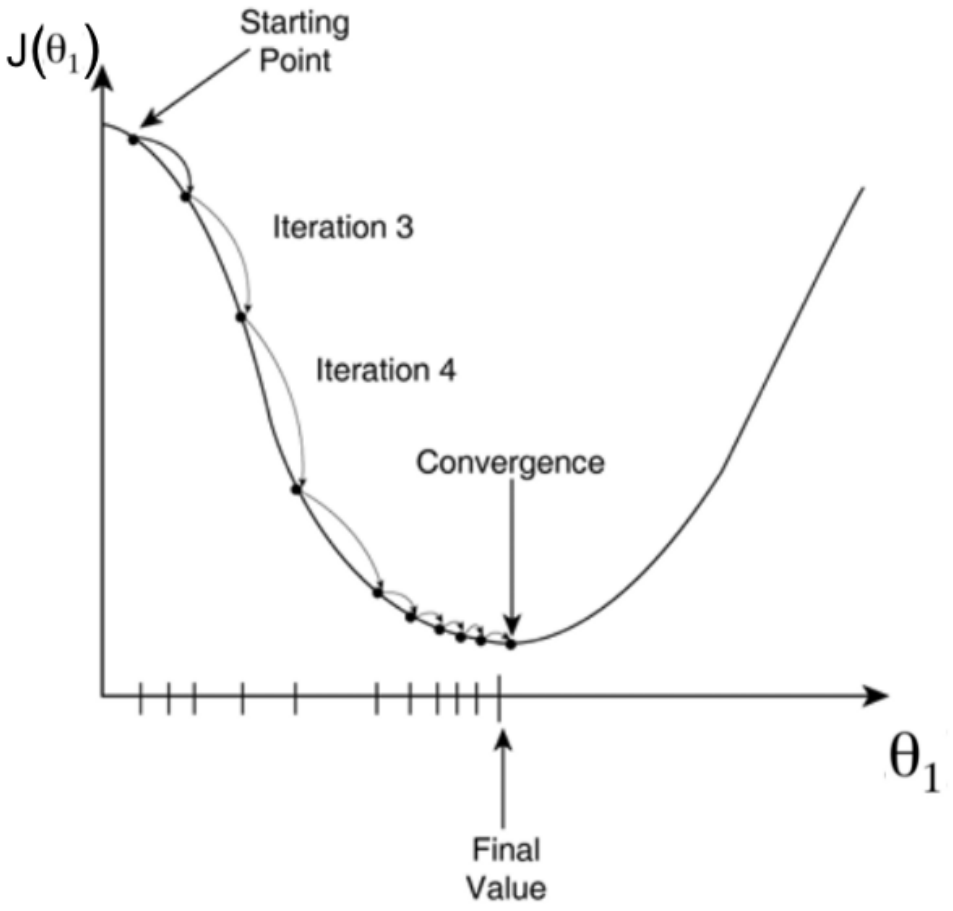
$$\theta_0, \theta_1$$

Cost Function:

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2$$

Goal:

$$\underset{\theta_0, \theta_1}{minimize} J(\theta_0, \theta_1)$$



Cost Function – "One Half Mean Squared Error":

$$J(\theta_0, \theta_1) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2$$

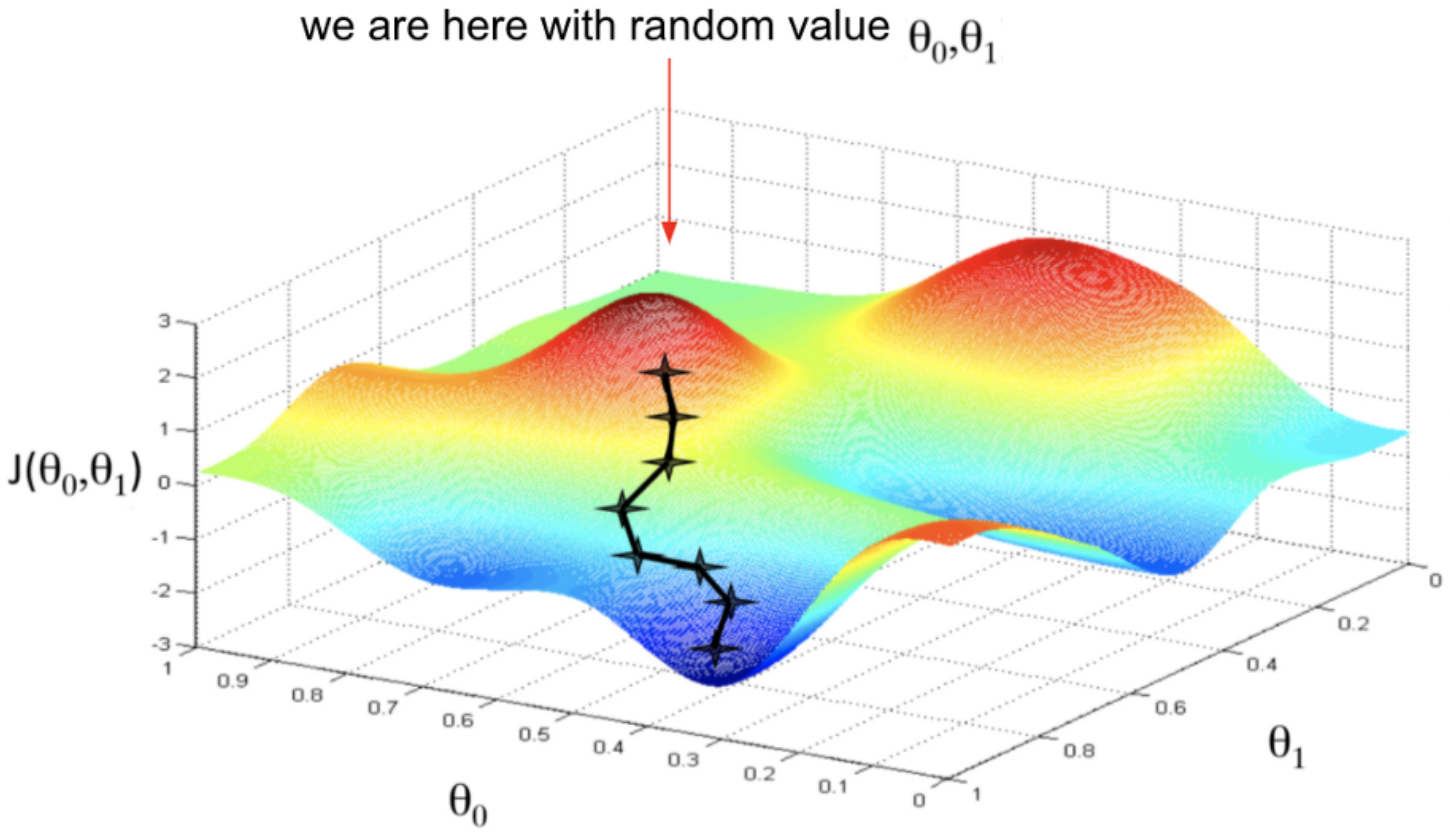
Objective:

$$\min_{\theta_0, \theta_1} J(\theta_0, \theta_1)$$

Derivatives:

$$\frac{\partial}{\partial\theta_0}J(\theta_0, \theta_1) = \frac{1}{m}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)$$

$$\frac{\partial}{\partial\theta_1}J(\theta_0, \theta_1) = \frac{1}{m}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)\cdot x^{(i)}$$



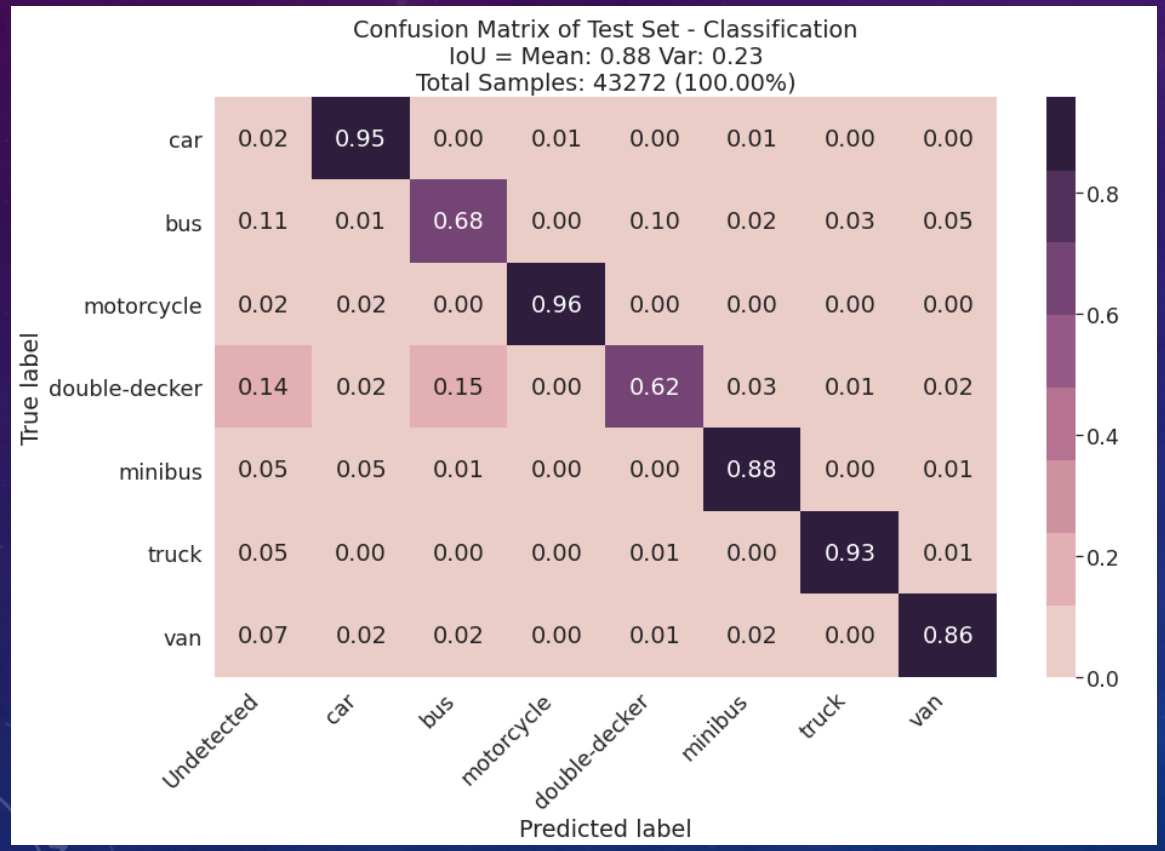
- Start with some θ_0,θ_1
- Keep changing θ_0,θ_1 to reduce $J(\theta_0,\theta_1)$

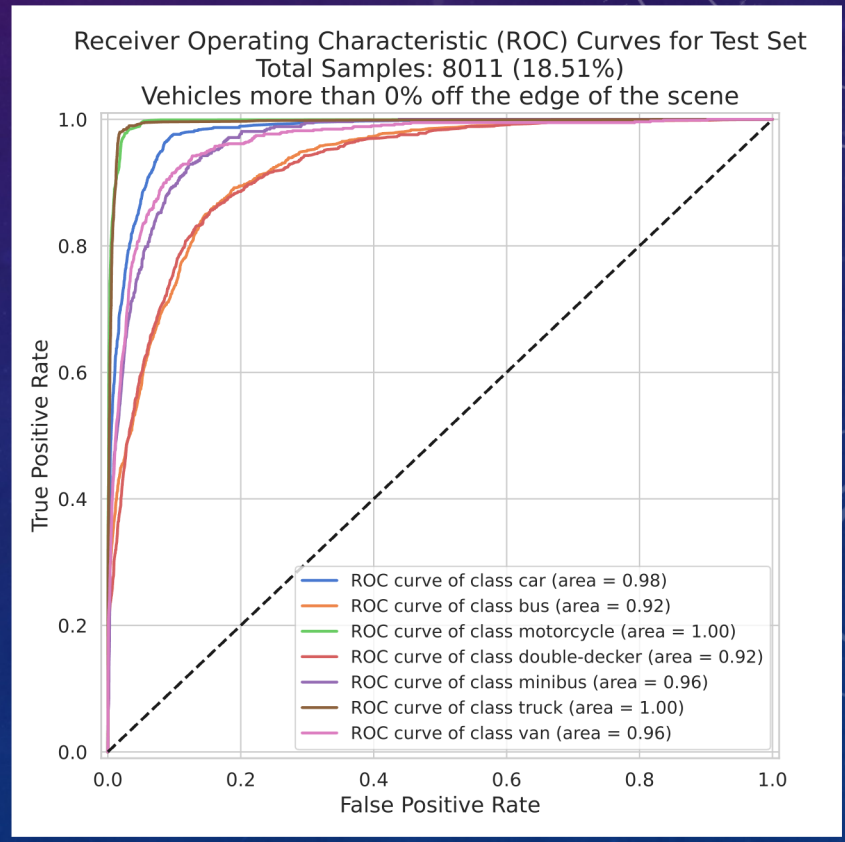
until we hopefully end up at a minimum

MACHINE LEARNING (CLASSIFICATION) PERFORMANCE METRICS

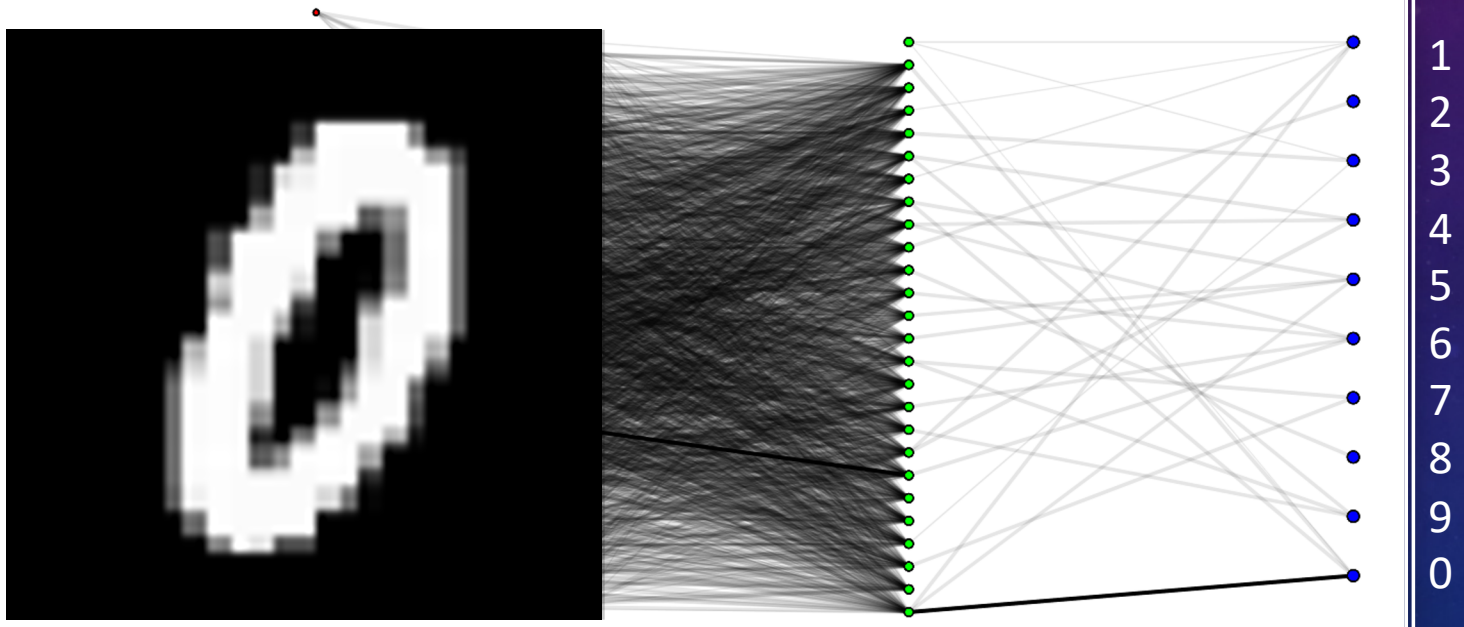
		True condition				
	Total population	Condition positive	Condition negative	Prevalence = $\frac{\Sigma \text{ Condition positive}}{\Sigma \text{ Total population}}$	Accuracy (ACC) = $\frac{\Sigma \text{ True positive} + \Sigma \text{ True negative}}{\Sigma \text{ Total population}}$	
Predicted condition	Predicted condition positive	**True positive**	**False positive**, Type I error	Positive predictive value (PPV), Precision = $\frac{\Sigma \text{ True positive}}{\Sigma \text{ Predicted condition positive}}$	False discovery rate (FDR) = $\frac{\Sigma \text{ False positive}}{\Sigma \text{ Predicted condition positive}}$	
	Predicted condition negative	**False negative**, Type II error	**True negative**	False omission rate (FOR) = $\frac{\Sigma \text{ False negative}}{\Sigma \text{ Predicted condition negative}}$	Negative predictive value (NPV) = $\frac{\Sigma \text{ True negative}}{\Sigma \text{ Predicted condition negative}}$	
		True positive rate (TPR), Recall, Sensitivity, probability of detection, Power = $\frac{\Sigma \text{ True positive}}{\Sigma \text{ Condition positive}}$	False positive rate (FPR), Fall-out, probability of false alarm = $\frac{\Sigma \text{ False positive}}{\Sigma \text{ Condition negative}}$	Positive likelihood ratio (LR+) = $\frac{\text{TPR}}{\text{FPR}}$	Diagnostic odds ratio (DOR) = $\frac{\text{LR+}}{\text{LR−}}$	$\mathbf{F_1}$ **score** = $2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$
		False negative rate (FNR), Miss rate = $\frac{\Sigma \text{ False negative}}{\Sigma \text{ Condition positive}}$	Specificity (SPC), Selectivity, True negative rate (TNR) = $\frac{\Sigma \text{ True negative}}{\Sigma \text{ Condition negative}}$	Negative likelihood ratio (LR−) = $\frac{\text{FNR}}{\text{TNR}}$		

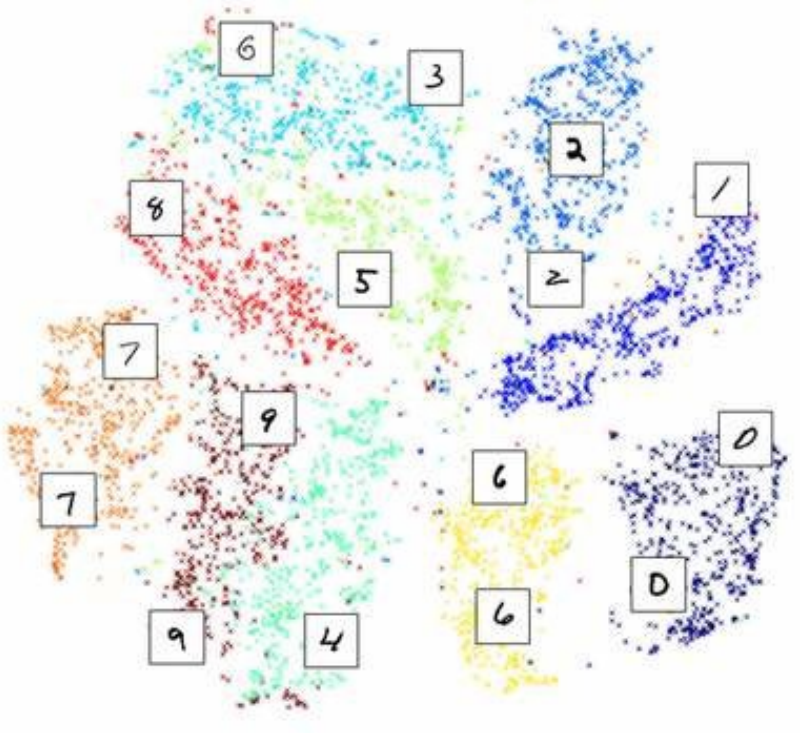
CLASSIC ML METRICS – CONFUSION MATRIX, ROC



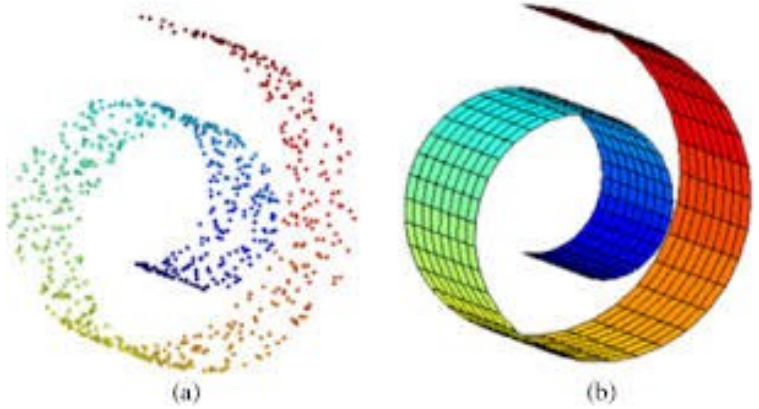


DEEP NEURAL NET VISUALIZATION

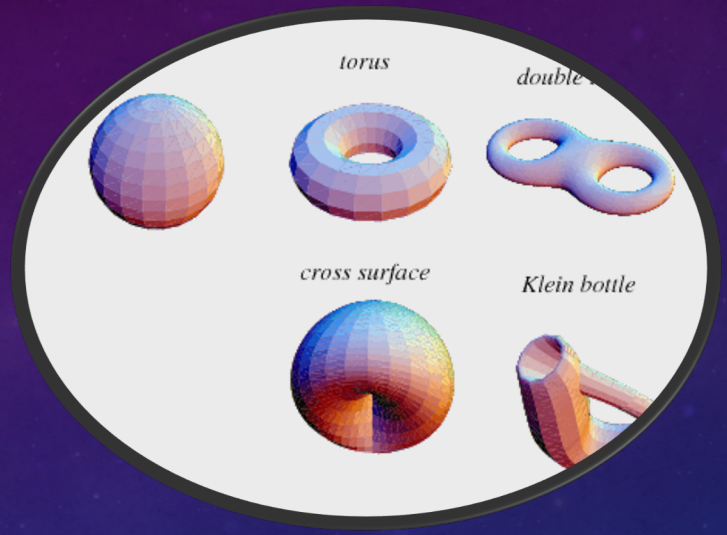




Manifold of handwritten digits as a two-dimensional representation.



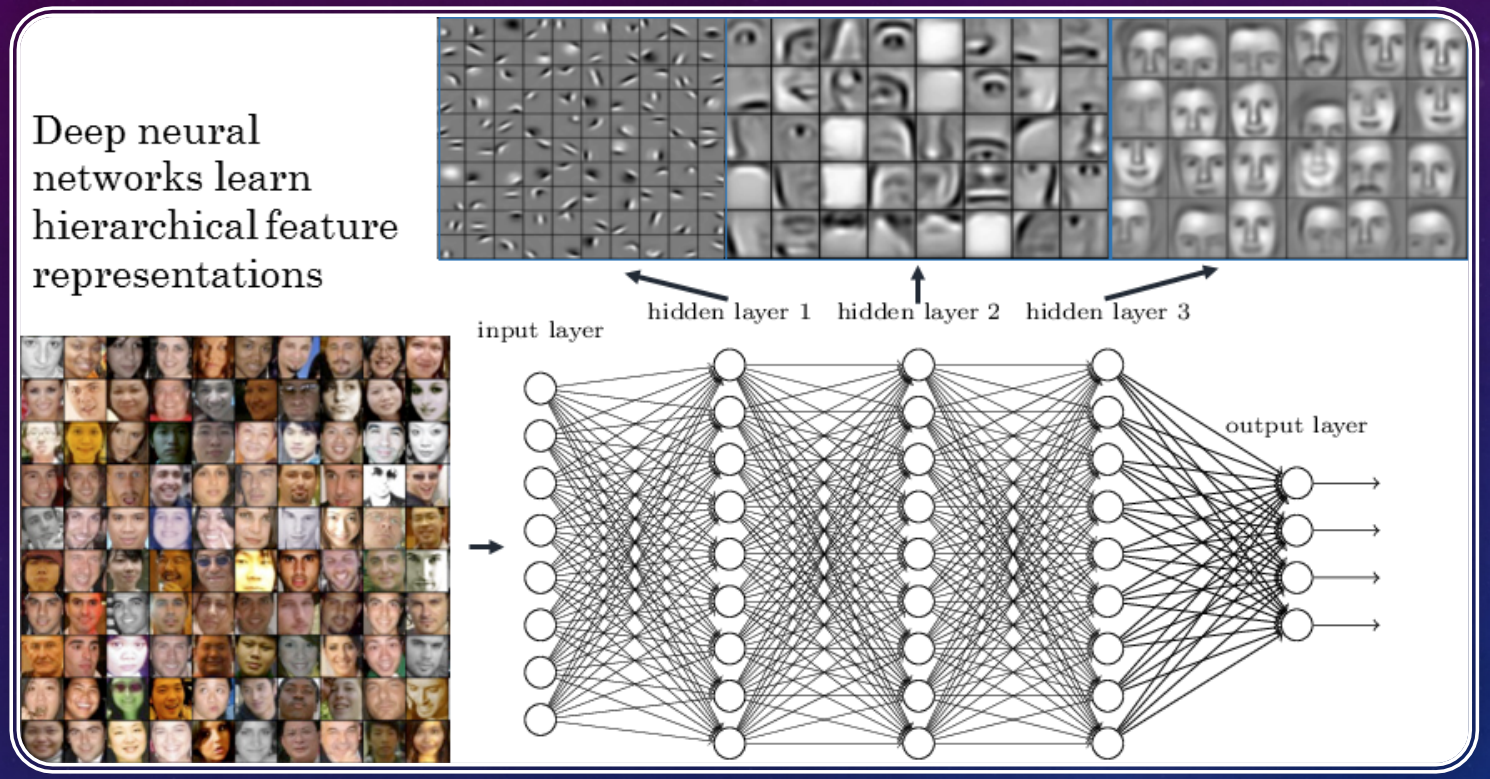
a) Dataset spatial representation. (b) Smooth surface approximation.



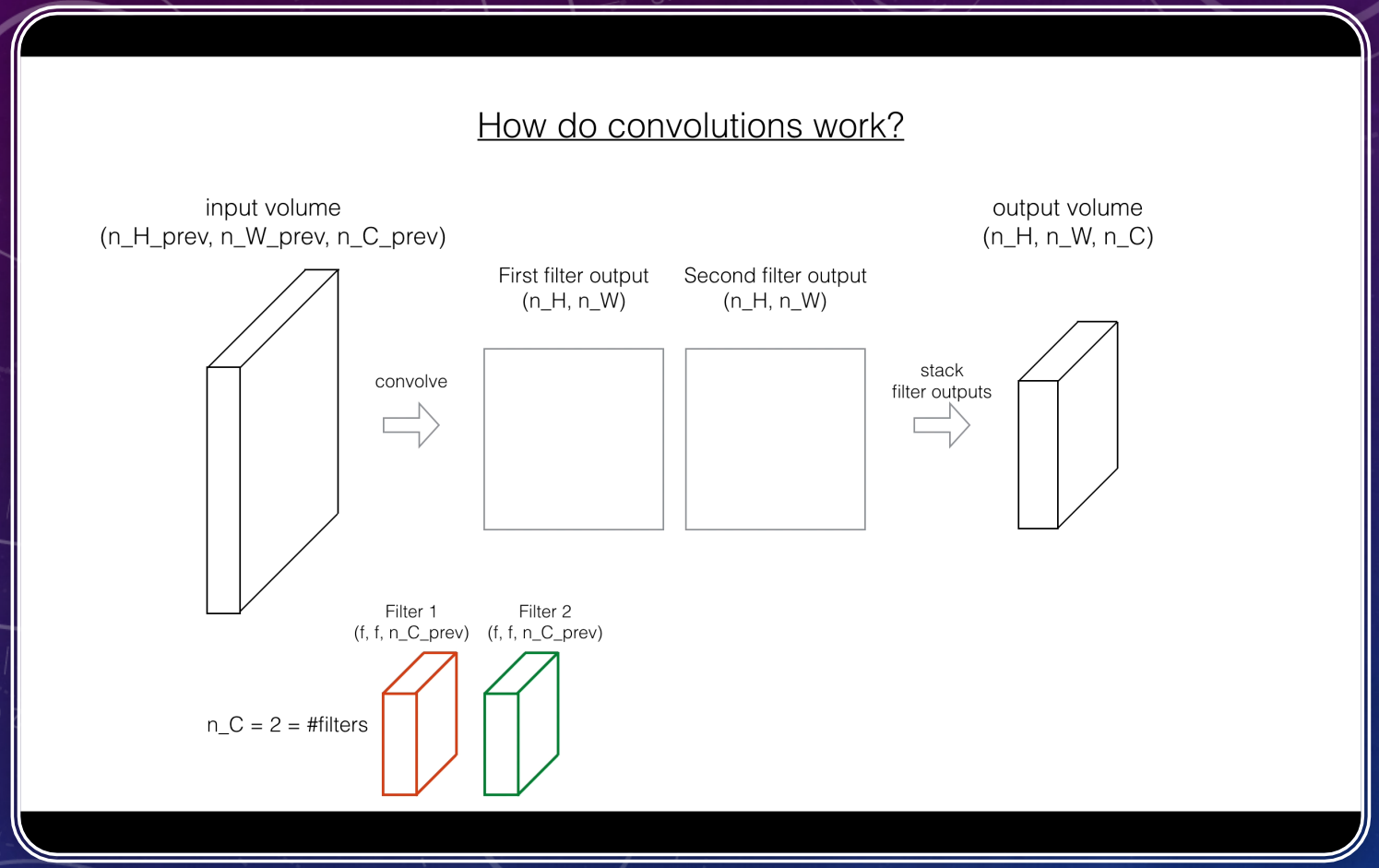
VECTOR SPACE & MANIFOLDS

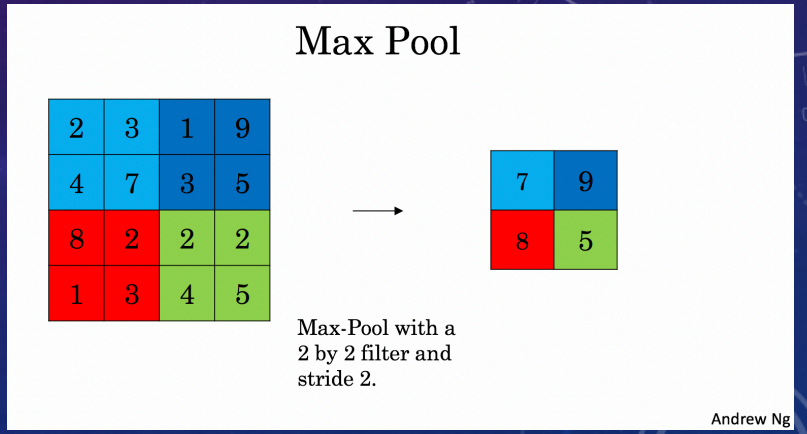
- Smoothness assumption
- Discover

DEEP NEURAL NET VISUALIZATION



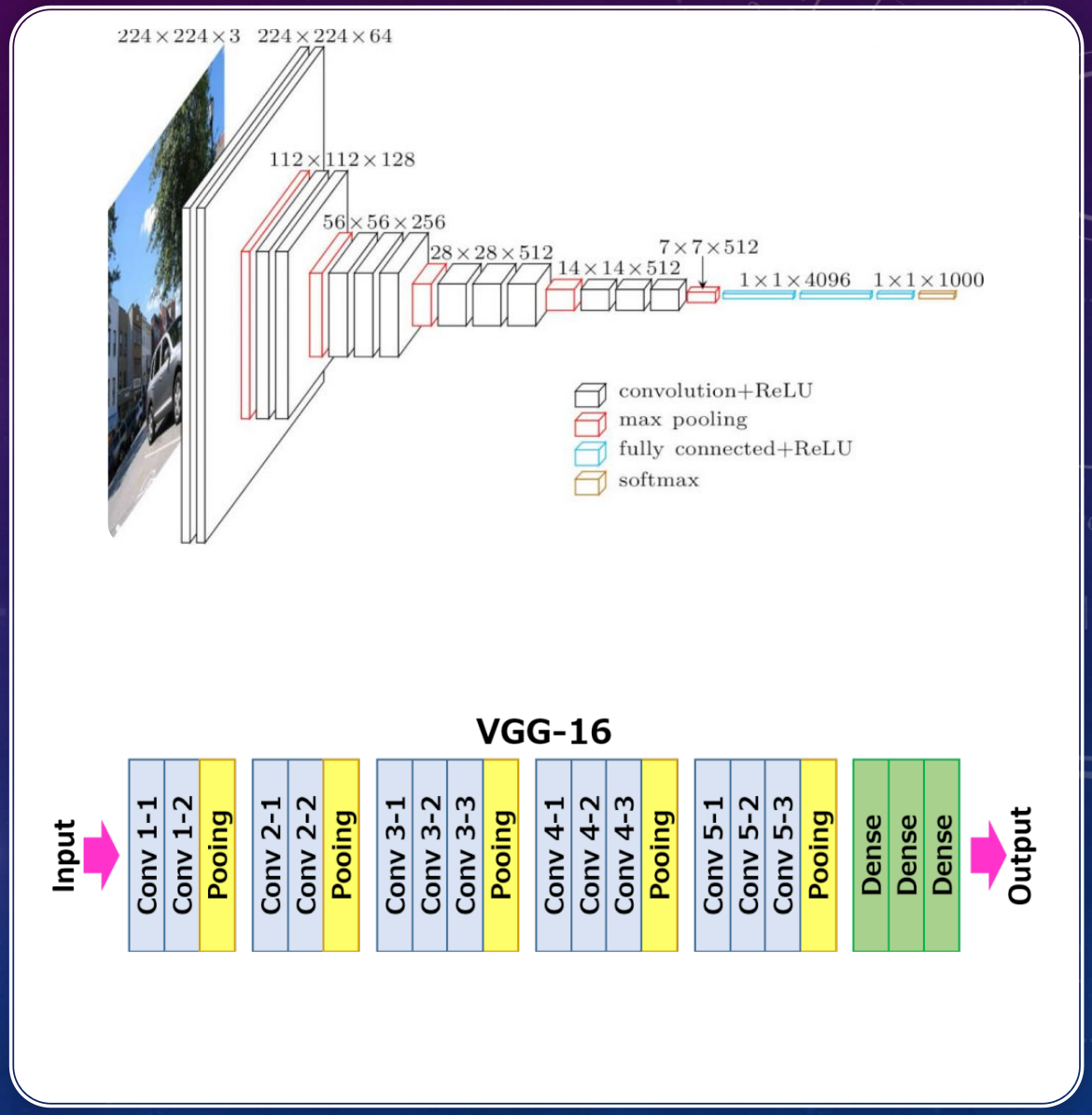
CONVOLUTIONAL NEURAL NETS





COMPLEX DEEP NEURAL NETS

- AlexNet
- Very Deep Convolutional Networks for Large-Scale Image Recognition(VGG-16)
  - most popular pre-trained models for image classification
  - Introduced in 2014 and still one of the best
  - 16-layers (pooling layers don't count - don't ask me why)
    - Convolutional Layers = 13
    - Pooling Layers = 5
    - Dense Layers = 3
  - number of parameters (weights) is **138 Billion**
- Inception net (Google)
  - 42 layers
- ResNet50
  - <guess> layers



GAVE RISE RAISE TO THE IDEA IT WAS A BLACK BOX

Training Data

Input:

Output:

Cat
0.96

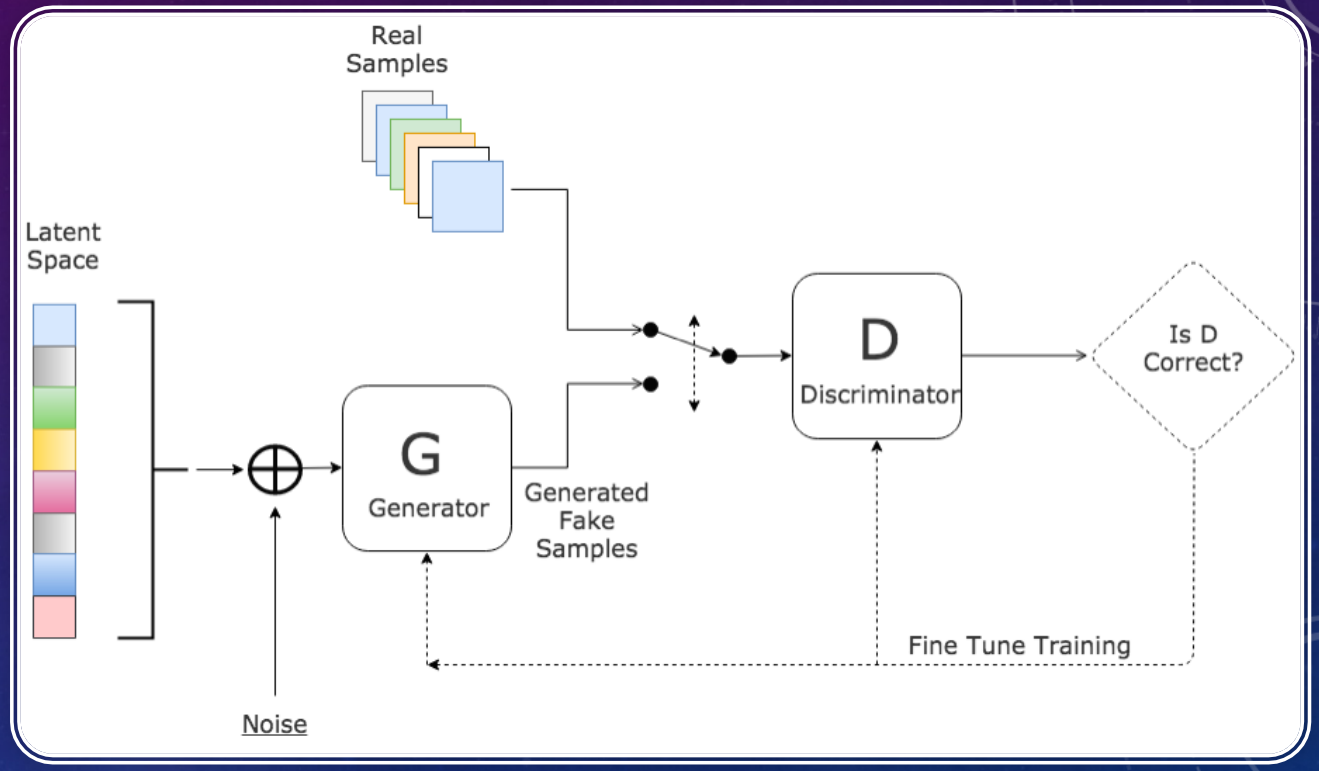
Input Layer Hidden Layer Output Layer

NOTE WORTHY

LEARNING MODELS AND EFFECTS IN FEATURE SPACE

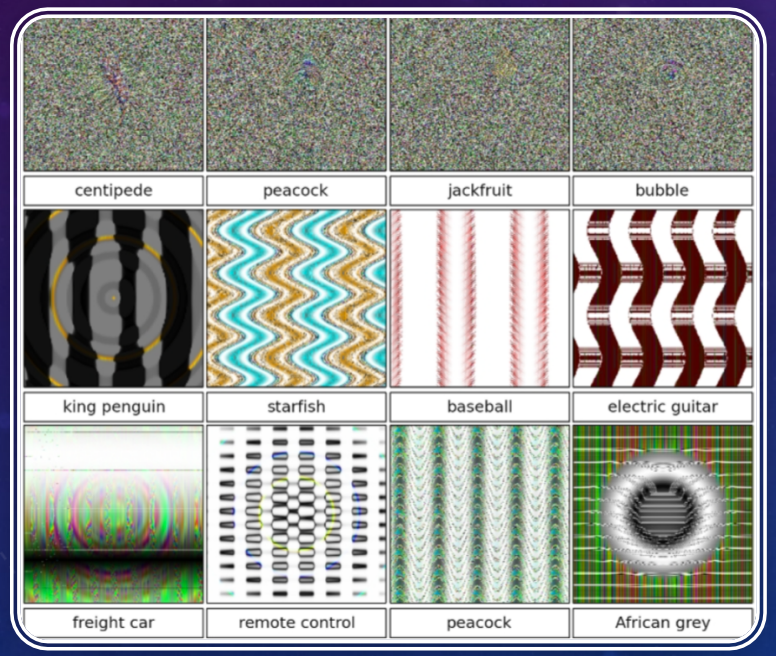
GENERATIVE ADVERSARIAL NETWORKS (GAN)

Image Source: https://www.kdnuggets.com/2017/01/generative-adversarial-networks-hot-topic-machine-learning.html



SECURITY: ADVERSARIAL ATTACK EXAMPLE

Reference A Tutorial on Attacking DNNs using Adversarial Examples
Christian Both, Arun Rawlani, Andi Ryyhan Chibrandy
https://github.com/arunrawlani/AdversarialMachineLearning_COMP551/blob/master/ML_Project_4_Final.pdf



IMPERCEPTIBLE CHANGES & UNRECOGNIZABLE OBJECTS

- C. Szegedy *et al.*, “Intriguing properties of neural networks,” *2nd Int. Conf. Learn. Represent. ICLR 2014 - Conf. Track Proc.*, pp. 1–10, 2014.
- A. Nguyen, J. Yosinski, and J. Clune, “Deep neural networks are easily fooled: High confidence predictions for unrecognizable images,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07-12-June-2015, pp. 427–436, 2015.