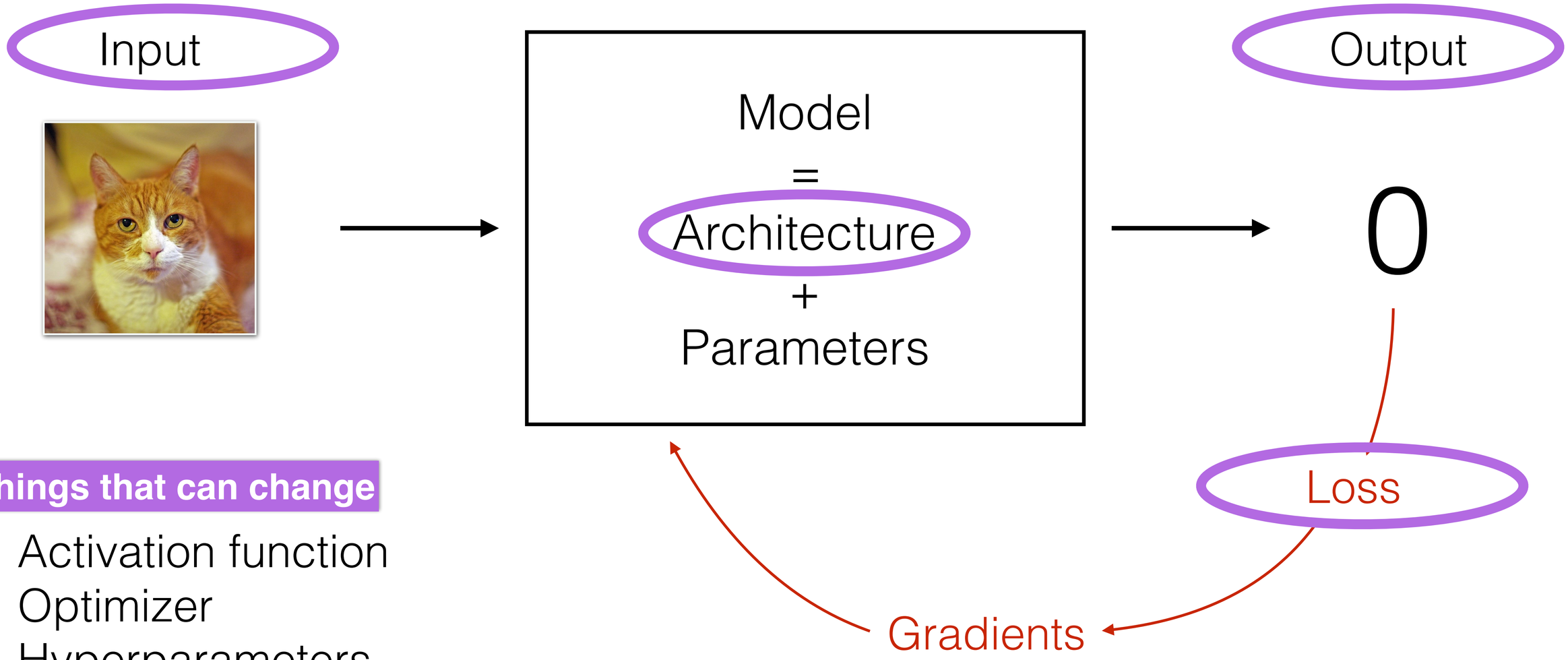
CS230: Lecture 2

Deep Learning Intuition

Kian Katanforoosh

Recap

Learning Process

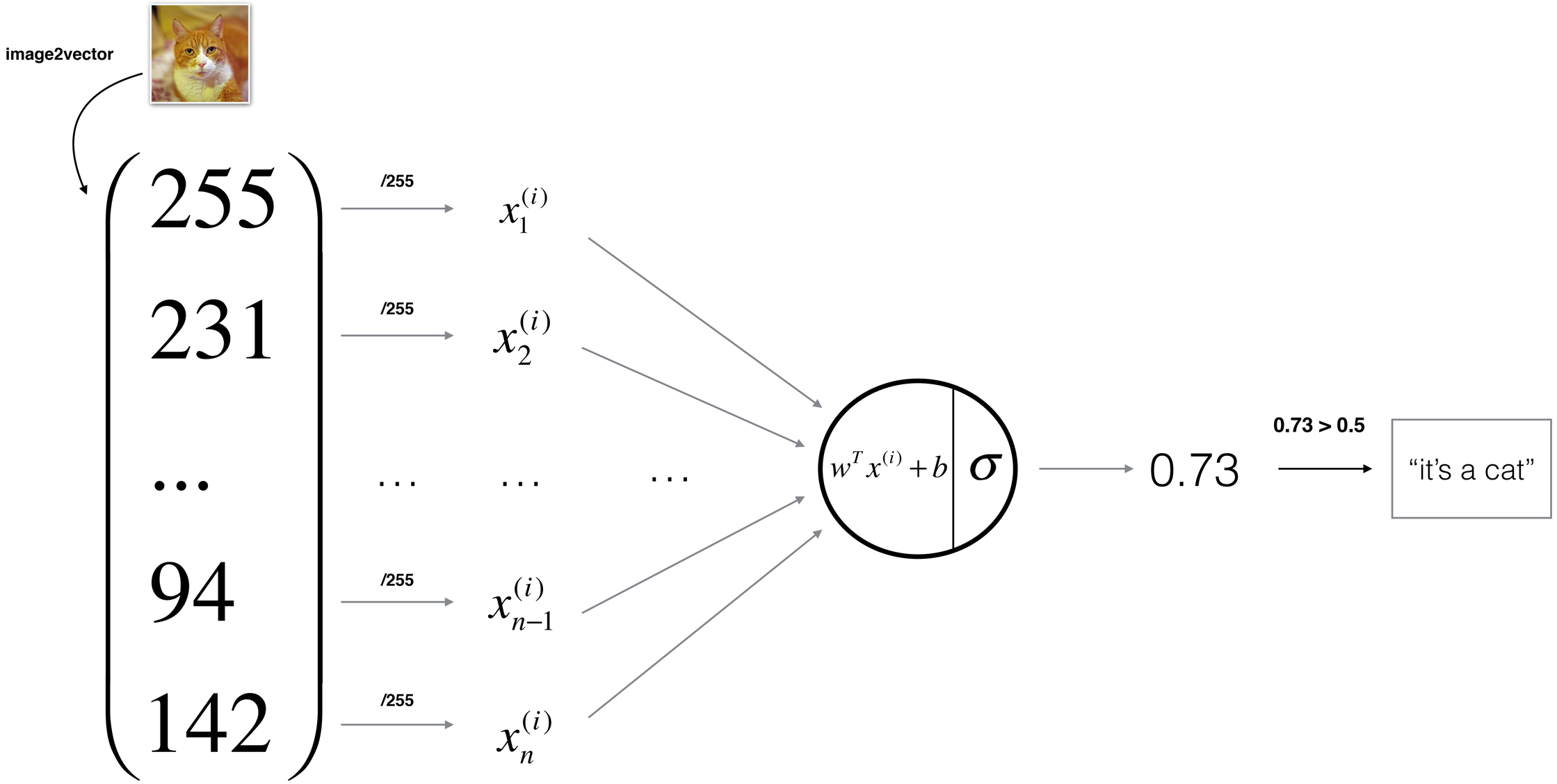


Things that can change

- Activation function
- Optimizer
- Hyperparameters
- …

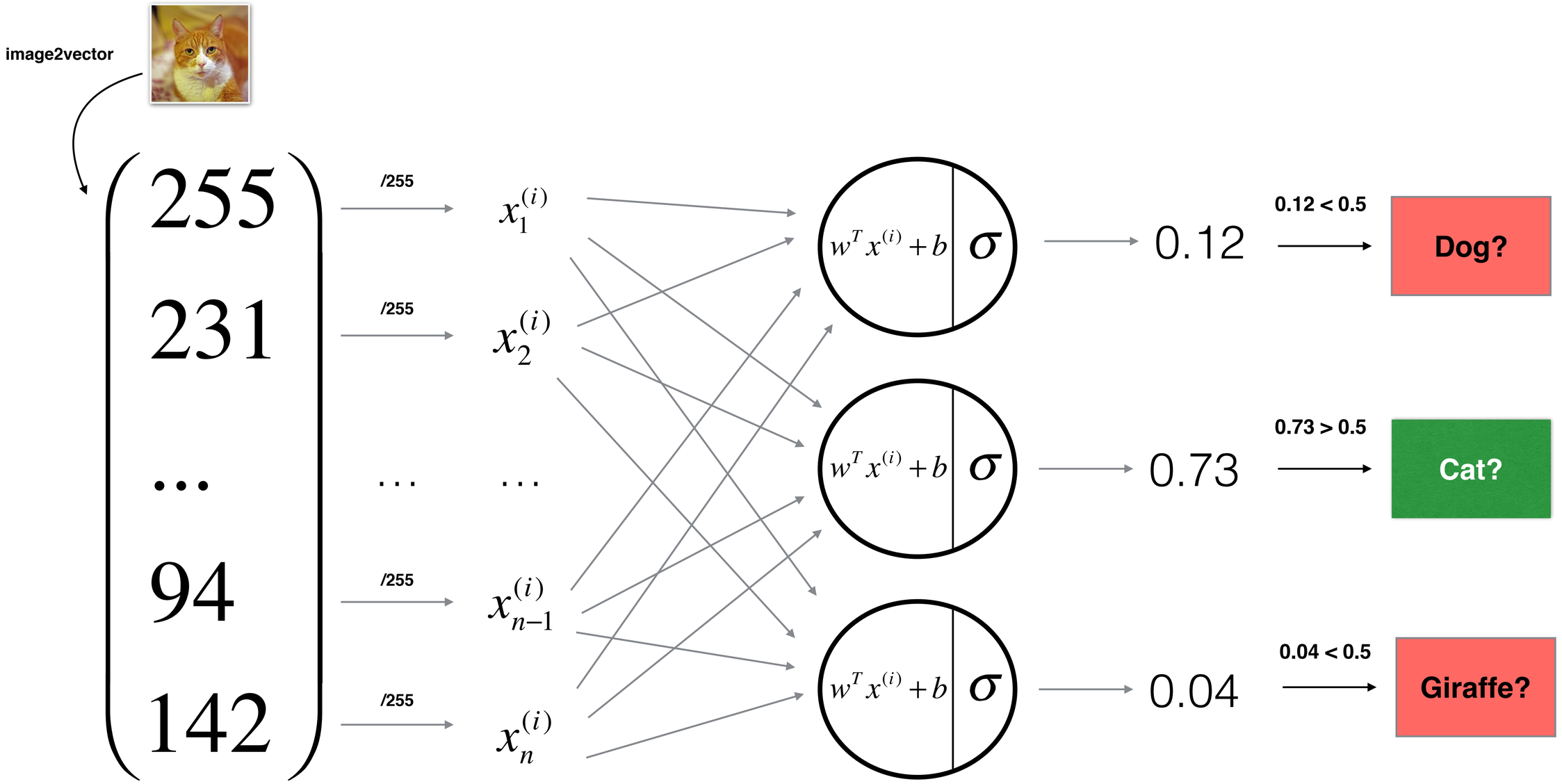
Kian Katanforoosh, Andrew Ng, Younes Bensouda Mourri

Logistic Regression as a Neural Network



Kian Katanforoosh, Andrew Ng, Younes Bensouda Mourri

Multi-class



Kian Katanforoosh, Andrew Ng, Younes Bensouda Mourri

Neural Network (Multi-class)

image2vector

$$\begin{pmatrix} 255 \\ 231 \\ \dots \\ 94 \\ 142 \end{pmatrix}$$

/255 → $x_1^{(i)}$

/255 → $x_2^{(i)}$

$\dots$ $\dots$

/255 → $x_{n-1}^{(i)}$

/255 → $x_n^{(i)}$

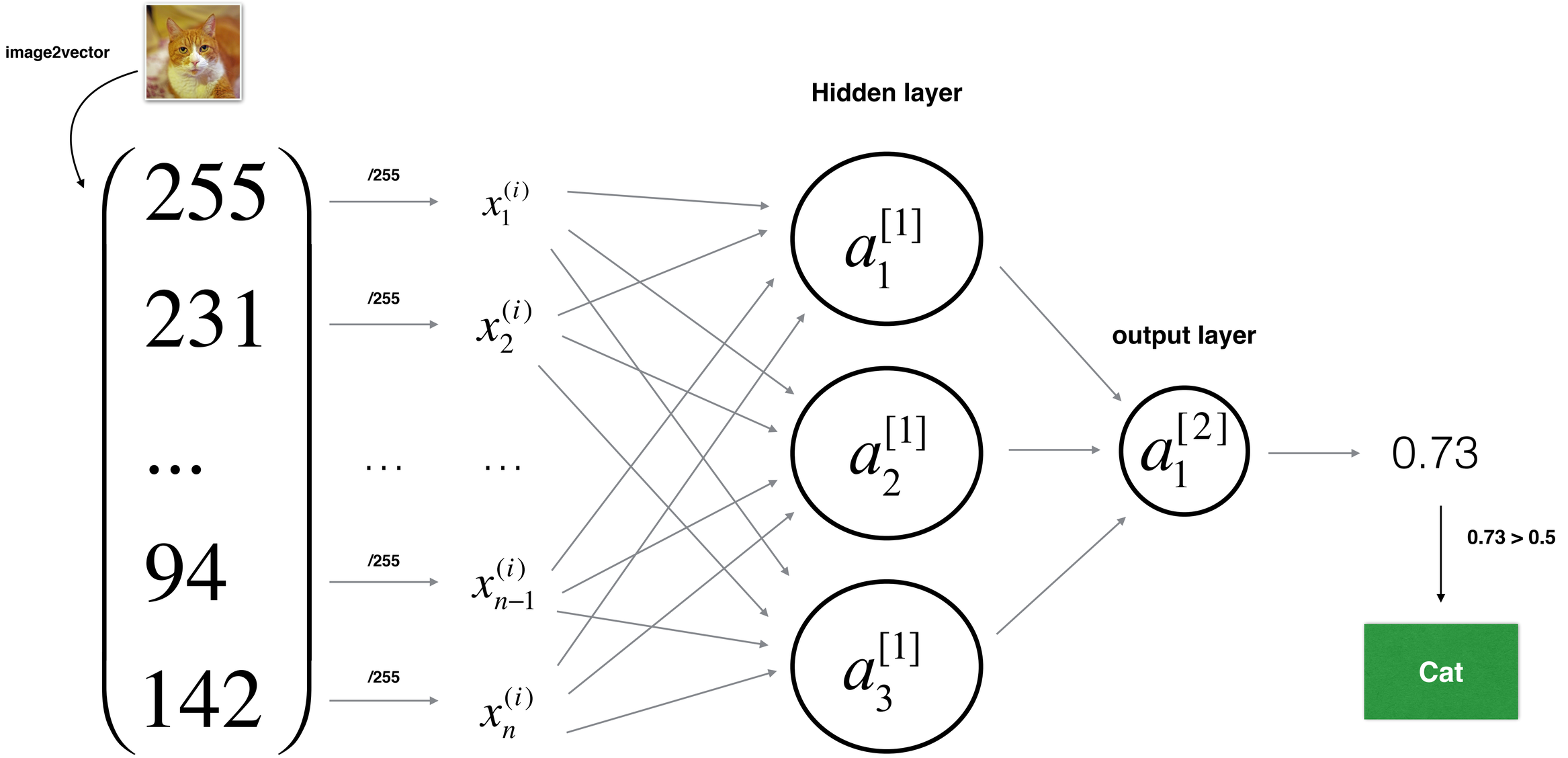
$w^T x^{(i)} + b \mid \sigma$

$w^T x^{(i)} + b \mid \sigma$

$w^T x^{(i)} + b \mid \sigma$

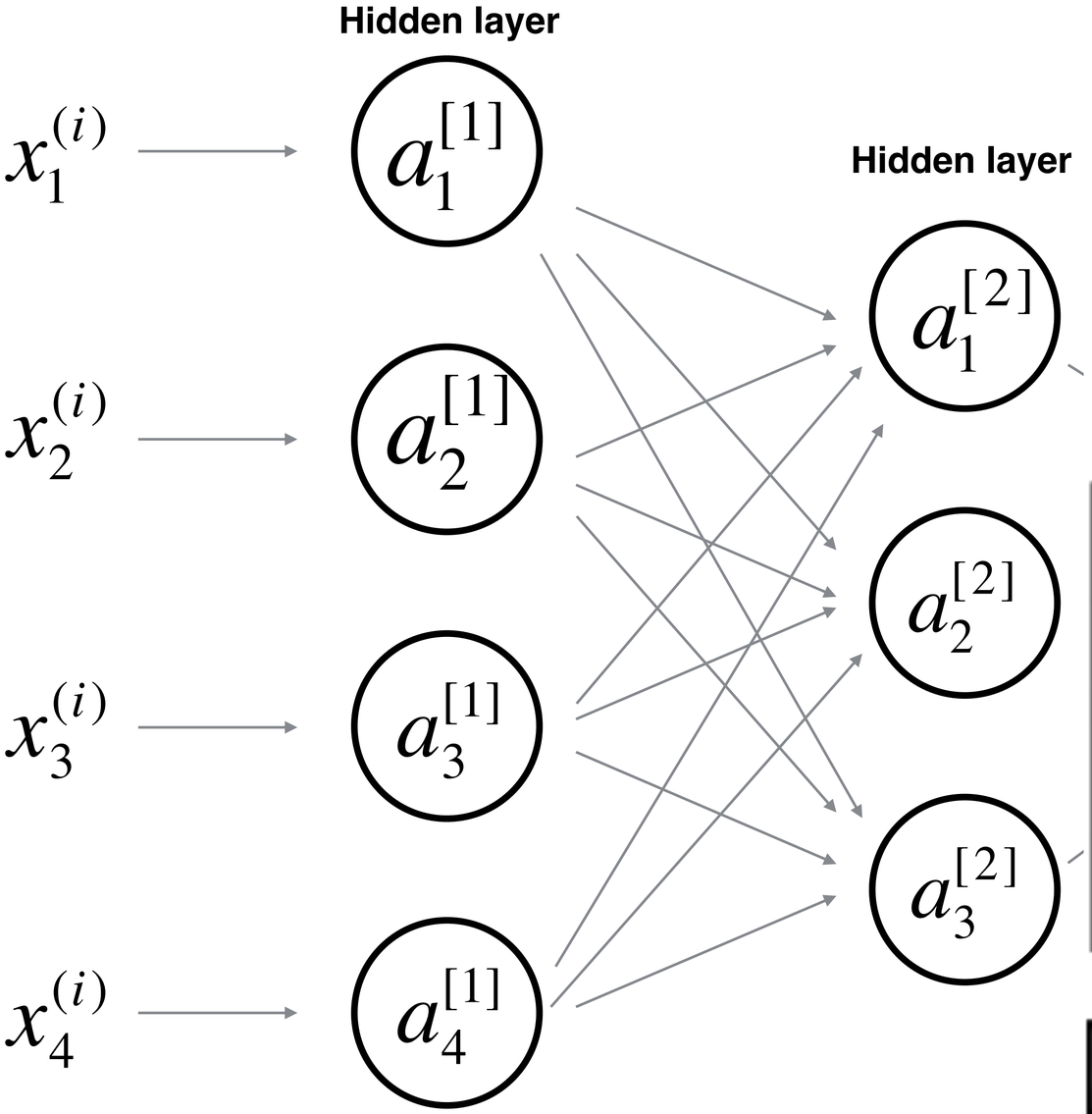
Kian Katanforoosh, Andrew Ng, Younes Bensouda Mourri

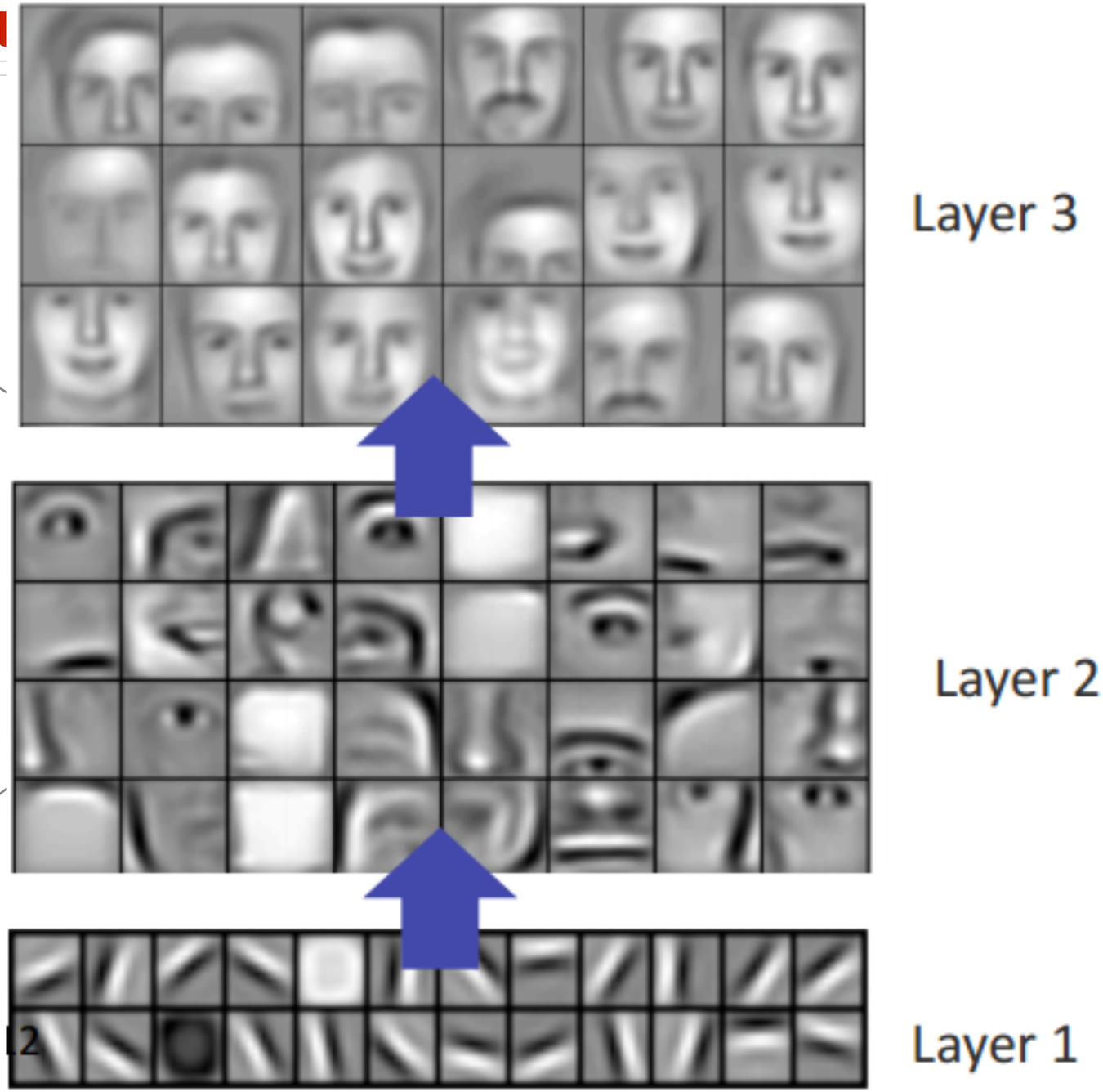
Neural Network (1 hidden layer)



Kian Katanforoosh, Andrew Ng, Younes Bensouda Mourri

Deeper net





Technique called “encoding”

Let’s build intuition on concrete applications

Kian Katanforoosh, Andrew Ng, Younes Bensouda Mourri

Today's outline

We will learn how to:

- Analyze a problem from a deep learning approach
- Choose an **architecture**
- Choose a **loss** and a **training strategy**

I. Day'n'Night classification

II. Face Recognition

III. Art generation

IV. Keyword Spotting

V. Shipping model

Day'n'Night classification (warm-up)

Goal: Given an image, classify as taken "during the day" (0) or "during the night" (1)

1. **Data?** 10,000 images Split? Bias?
2. **Input?**  Resolution? (64, 64, 3)
3. **Output?** y = 0 or y = 1 Last Activation? sigmoid
4. **Architecture ?** Shallow network should do the job pretty well
5. **Loss?** $L = -[y \log(\hat{y}) + (1-y)\log(1-\hat{y})]$

Easy

Kian Katanforoosh, Andrew Ng, Younes Bensouda Mourri

Face Verification

Goal: A school wants to use Face Verification for validating student IDs in facilities (dinning halls, gym, pool …)

1. Data?

Picture of every student labelled with their name

Bertrand

2. Input?

Resolution?

(412, 412, 3)

3. Output?

y = 1 (it's you)

or

y = 0 (it's not you)

Face Verification

Goal: A school wants to use Face Verification for validating student IDs in facilities (dinning halls, gym, pool …)

4. What architecture?

Simple solution:

database image

compute distance pixel per pixel

if less than threshold then y=1

input image

Issues:

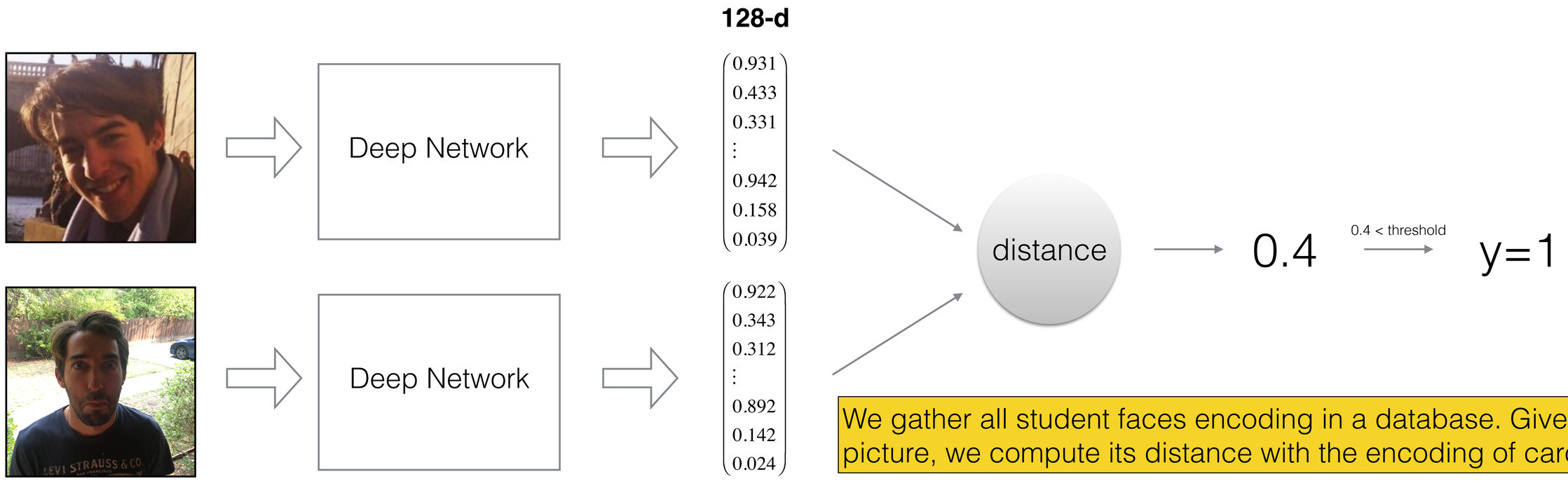
- Background lighting differences
- A person can wear make-up, grow a beard…
- ID photo can be outdated

Face Verification

Goal: A school wants to use Face Verification for validating student IDs in facilities (dinning halls, gym, pool …)

4. What architecture?

Our solution: encode information about a picture in a vector



We gather all student faces encoding in a database. Given a new picture, we compute its distance with the encoding of card holder

Kian Katanforoosh, Andrew Ng, Younes Bensouda Mourri

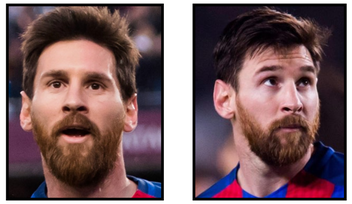
Face Recognition

Goal: A school wants to use Face Verification for validating student IDs in facilities (dinning hall, gym, pool …)

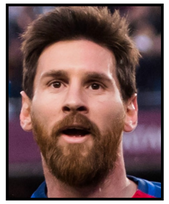
4. Loss? Training?

We need more data so that our model understands how to encode:
Use public face datasets

What we really want:

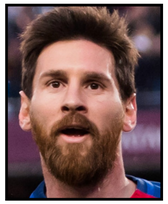


similar encoding

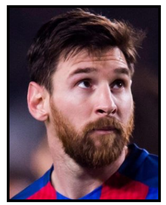



different encoding

So let's generate triplets:

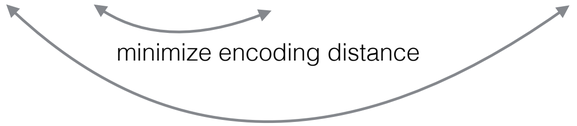


anchor



positive

negative



Kian Katanforoosh, Andrew Ng, Younes Bensouda Mourri

Recap: Learning Process

Input

anchor positive negative

Model
=
Architecture
+
Parameters

Output

$$\begin{pmatrix} 0.13 \\ 0.42 \\ .. \\ 0.10 \\ 0.31 \\ 0.73 \\ .. \\ 0.43 \\ 0.33 \end{pmatrix} \quad \begin{pmatrix} 0.01 \\ 0.54 \\ .. \\ 0.45 \\ 0.01 \\ 0.89 \\ .. \\ 0.12 \\ 0.01 \end{pmatrix} \quad \begin{pmatrix} 0.95 \\ 0.45 \\ .. \\ 0.20 \\ 0.41 \\ 0.89 \\ .. \\ 0.31 \\ 0.34 \end{pmatrix}$$

Enc(A) Enc(P) Enc(N)

Loss

$$L = \left\|Enc(A) - Enc(P)\right\|_2^2 - \left\|Enc(A) - Enc(N)\right\|_2^2 + \alpha$$

Gradients

Kian Katanforoosh, Andrew Ng, Younes Bensouda Mourri

Face Recognition

Goal: A school wants to use Face Identification for recognize students in facilities (dinning hall, gym, pool …)

K-Nearest Neighbors

Goal: You want to use Face Clustering to group pictures of the same people on your smartphone

K-Means Algorithm

Maybe we need to detect the faces first?

Kian Katanforoosh, Andrew Ng, Younes Bensouda Mourri

Art generation (Neural Style Transfer)

Goal: Given a picture, make it look beautiful

1. Data?

Let's say we have any data

2. Input?

content image

style image

3. Output?

generated image

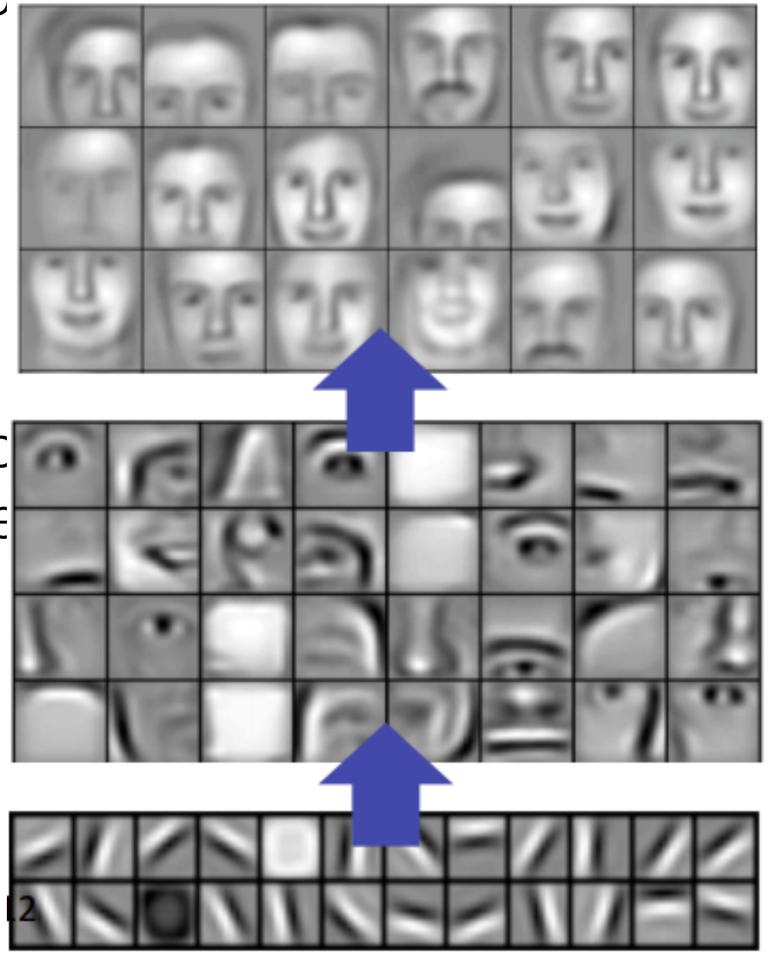
Leon A. Gatys, Alexander S. Ecker, Matthias Bethge: A Neural Algorithm of Artistic Style, 2015

Kian Katanforoosh, Andrew Ng, Younes Bensouda Mourri

Art generation (Neural Style Transfer)

4. Architecture?

We want a model that **understands images** very well
We load an **existing model trained on ImageNet** for example



ation

When this image forward p
about its content & its style

$Content_C$
$Style_S$

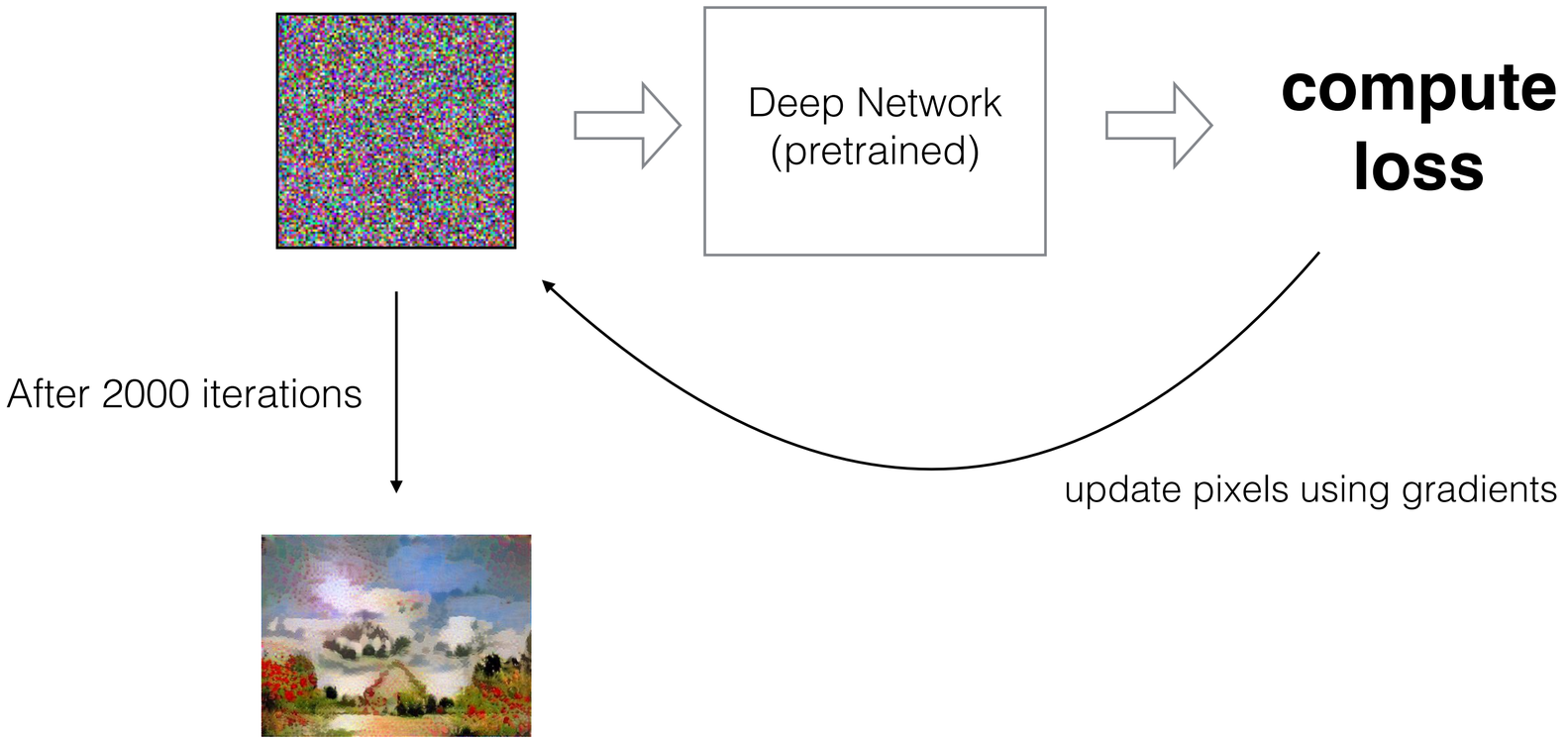
5. Loss?

Art generation (Neural Style Transfer)

Correct Approach

$$L = \left\| Content_C - Content_G \right\|_2^2 + \left\| Style_S - Style_G \right\|_2^2$$

We are not learning parameters by minimizing L. We are learning an image!



Leon A. Gatys, Alexander S. Ecker, Matthias Bethge: A Neural Algorithm of Artistic Style, 2015

Speech recognition: Keyword Spotting

Goal: Given an audio speech, detect the word “lion”.

1. **Input?**

2. **Output?**

   y = 0 (there is “lion”) or y = 1 (there isn’t “lion”)

   y = (0,0,0,0,0,0,0,0,0,0,….,0,0,0,0,0,0,1,0,0, ….,0,0,0,0,0,0,0,0,0,0)

3. **Data?**

   Many audio recordings (“words”)

Speech recognition: Keyword Spotting

Goal: Given an audio speech, detect the word “lion”.

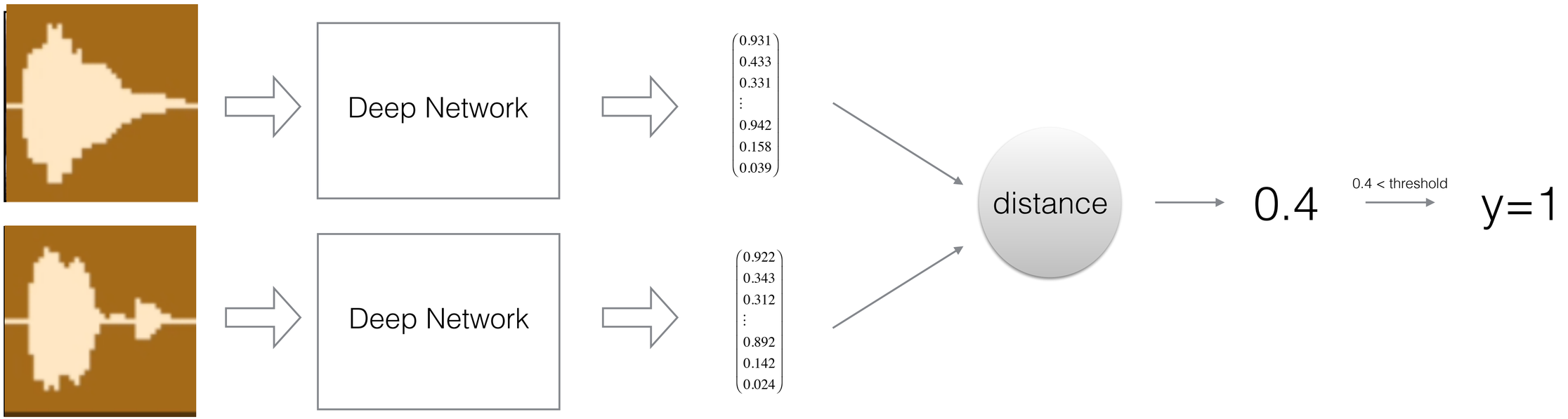
$$L = \left\| Enc(A) - Enc(P) \right\|_2^2 - \left\| Enc(A) - Enc(N) \right\|_2^2 + \alpha$$

4. What architecture?

y = (0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,….,0,0,0,0,0,0,1,0,0, ….,0,0,0,0,0,0,0,0,0,0,0)

0.12 0.01 0.27 … … … … … … … 0.21 **0.92** 0.43 … … … … … …

Threshold: 0.6

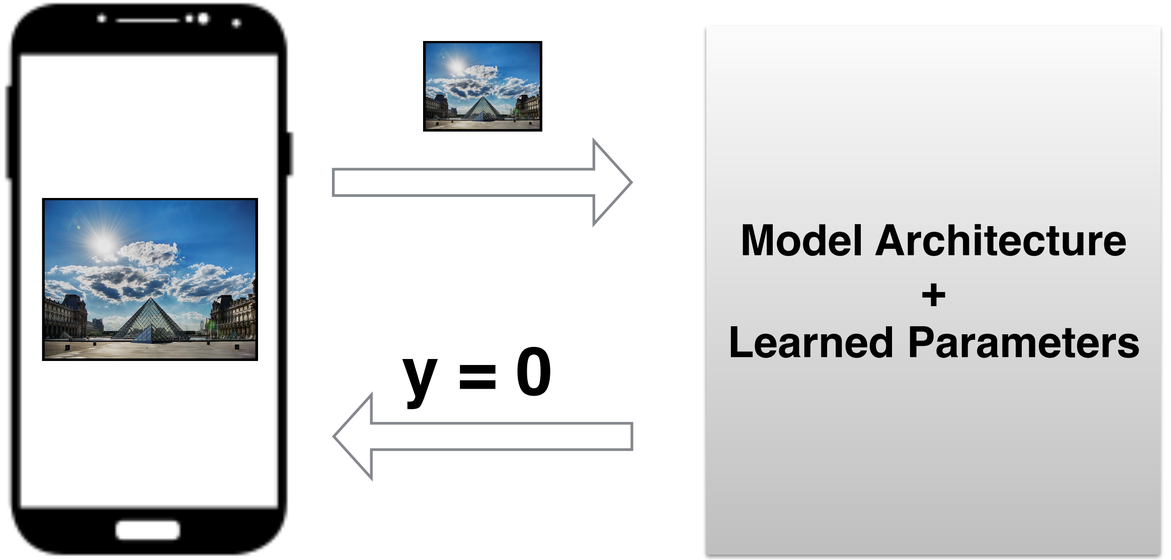


Kian Katanforoosh, Andrew Ng, Younes Bensouda Mourri

App implementation

Server-based or on-device?

Server-based



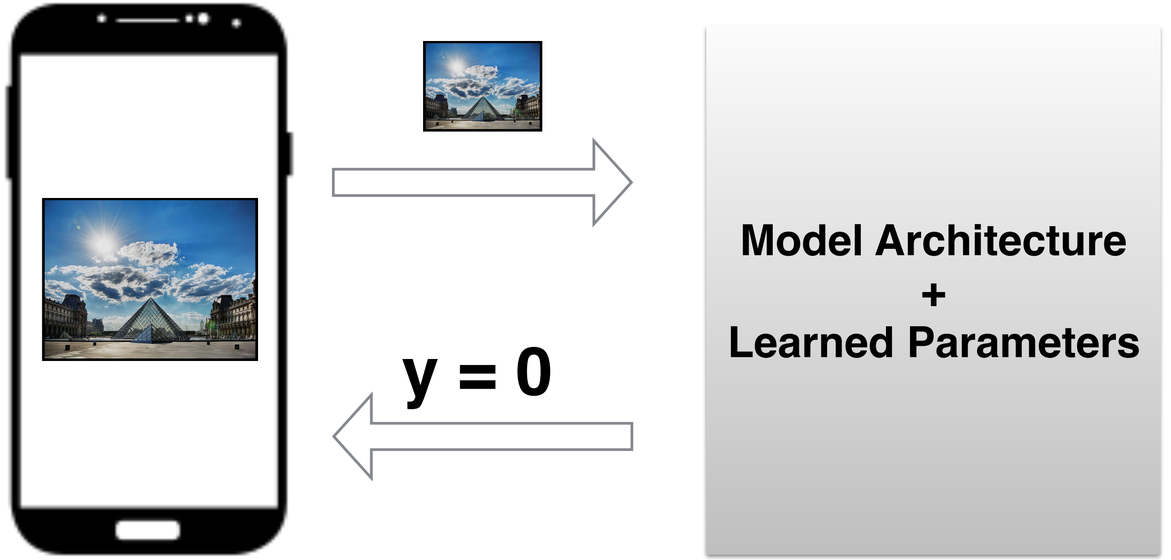
On-device

Kian Katanforoosh, Andrew Ng, Younes Bensouda Mourri

Server-based or on-device?

Server-based

- App is light-weight
- App is easy to update



On-device

- Faster predictions
- Works offline

Kian Katanforoosh, Andrew Ng, Younes Bensouda Mourri

Duties for next week

For Tuesday 04/17, 9am:

C1M3

- **Quiz: Shallow Neural Networks**
- **Programming Assignment: Planar data classification with one-hidden layer**

C1M4

- **Quiz: Deep Neural Networks**
- **Programming Assignment: Building a deep neural network - Step by Step**
- **Programming Assignment: Deep Neural Network Application**

Project

- **For this Friday (04/13): find teammate and submit the Team-members form with your project category**
- **Fill-in AWS Form to get GPU credits**

This Friday (04/13):

- **(optional) Project section: How to get started with your projects?**