

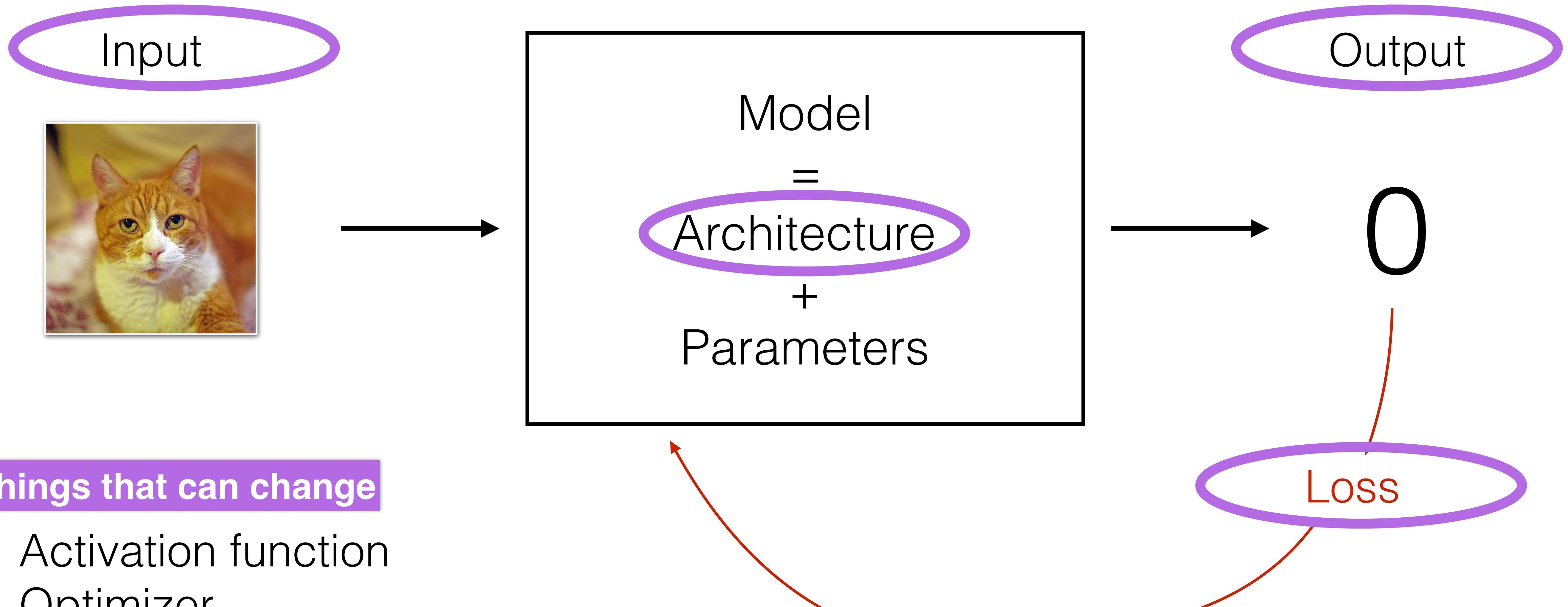
CS230: Lecture 2

Deep Learning Intuition

Kian Katanforoosh

Recap

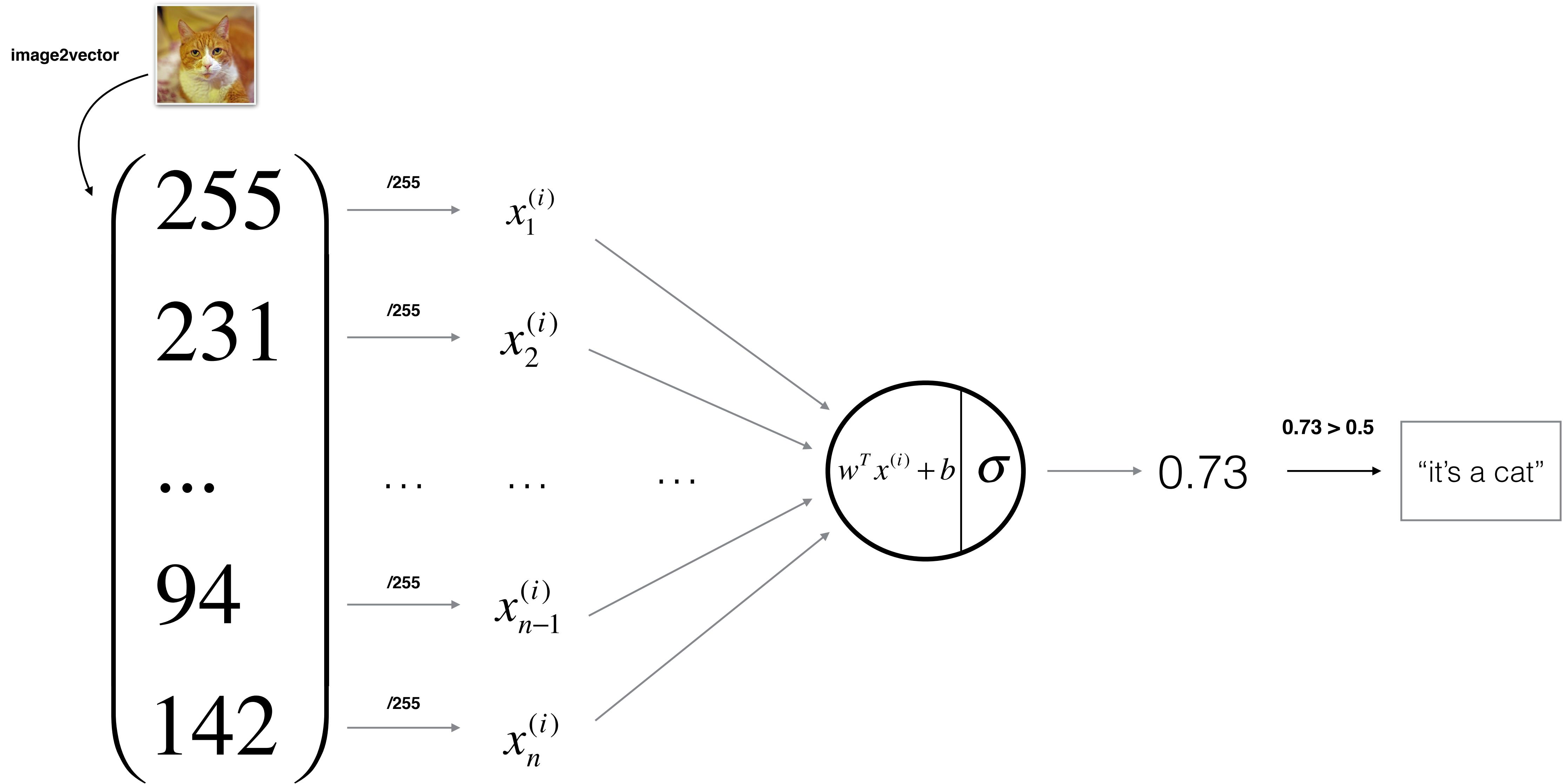
Learning Process



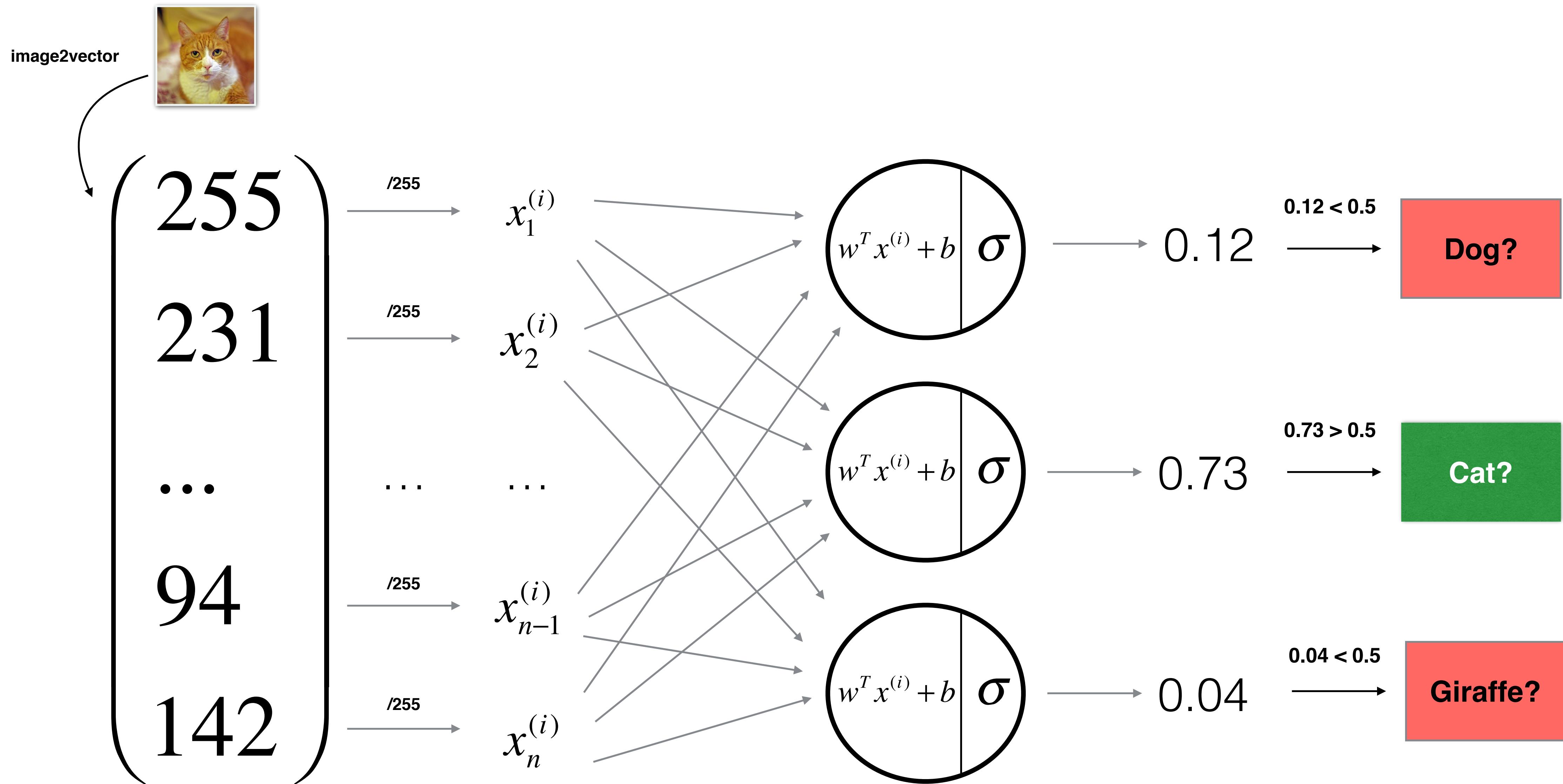
Things that can change

- Activation function
- Optimizer
- Hyperparameters
- ...

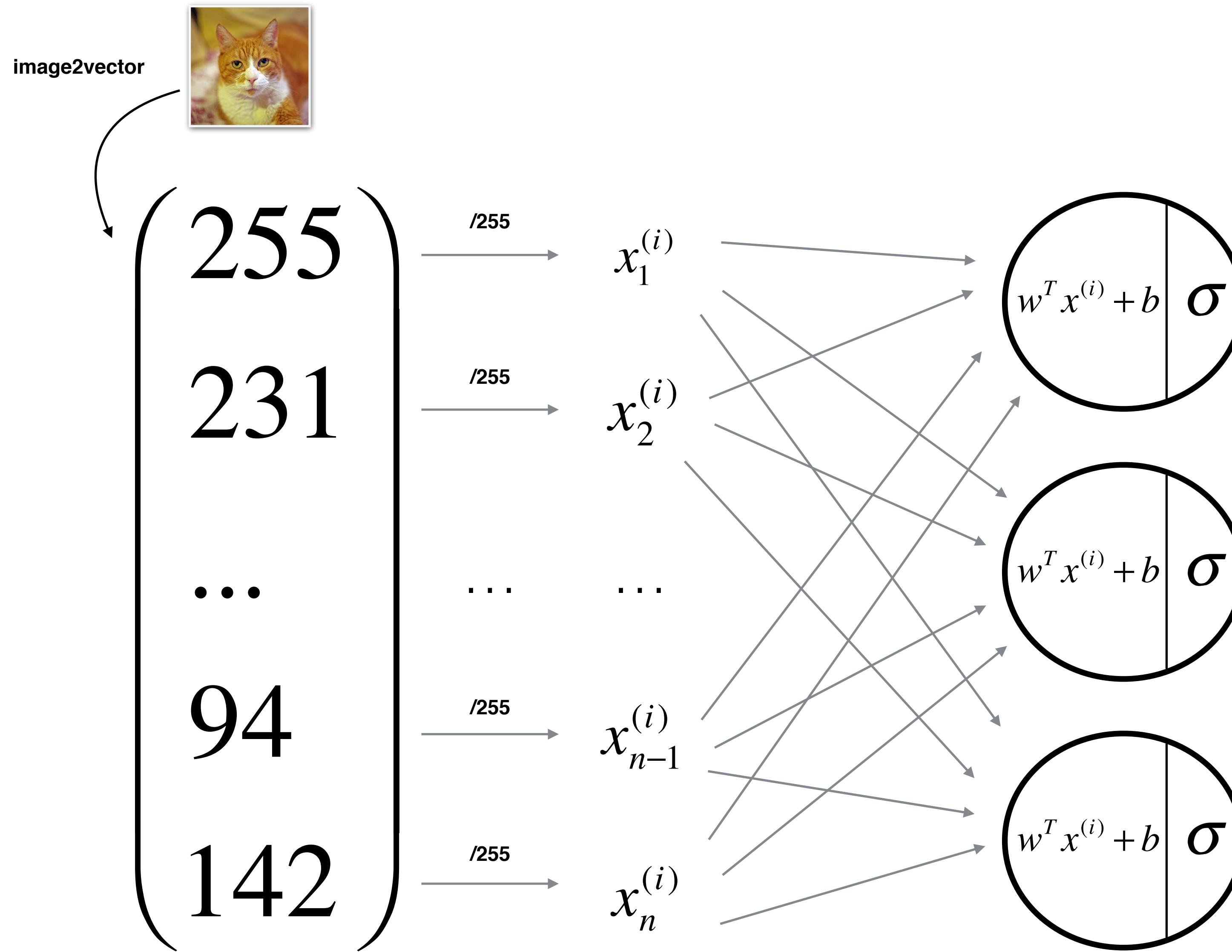
Logistic Regression as a Neural Network



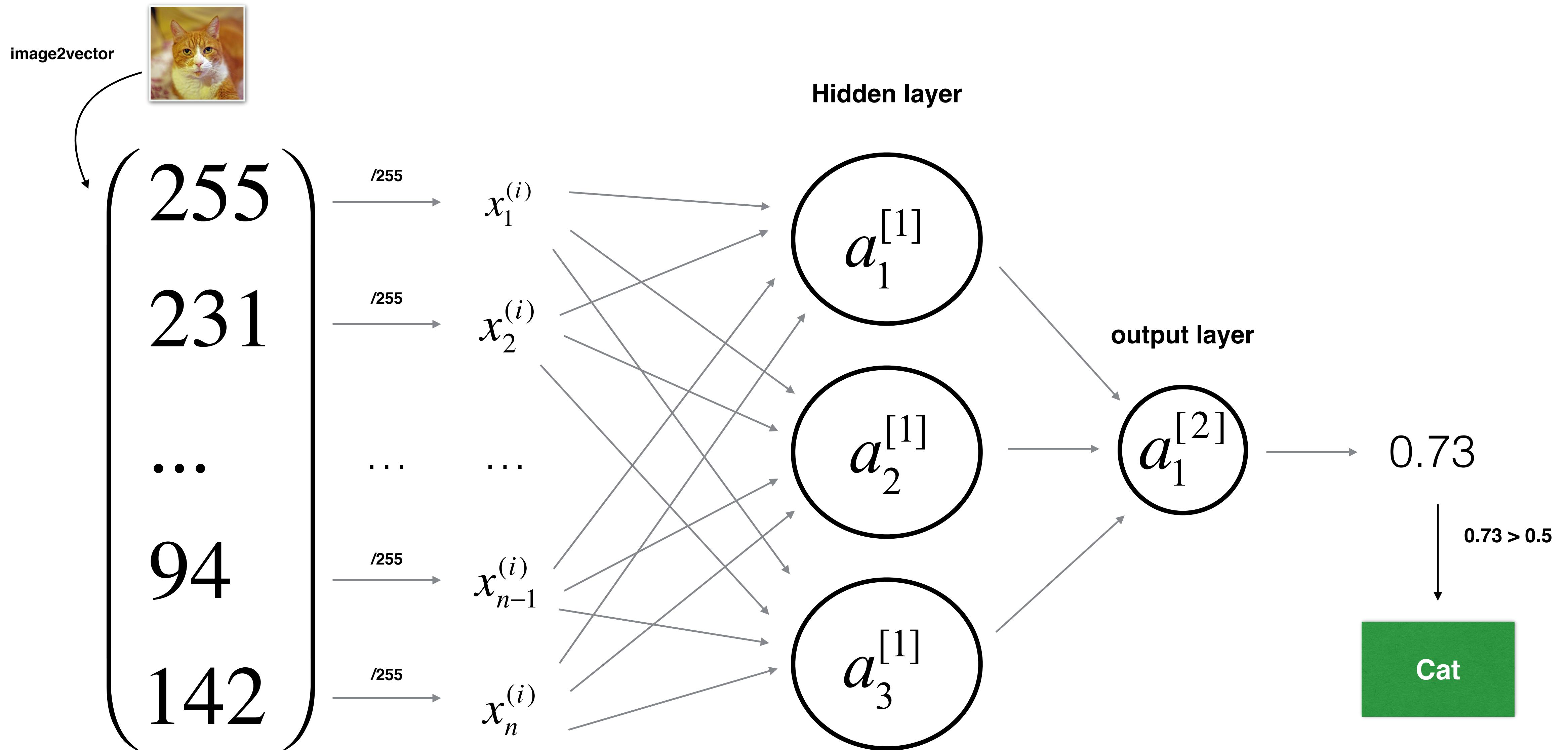
Multi-class



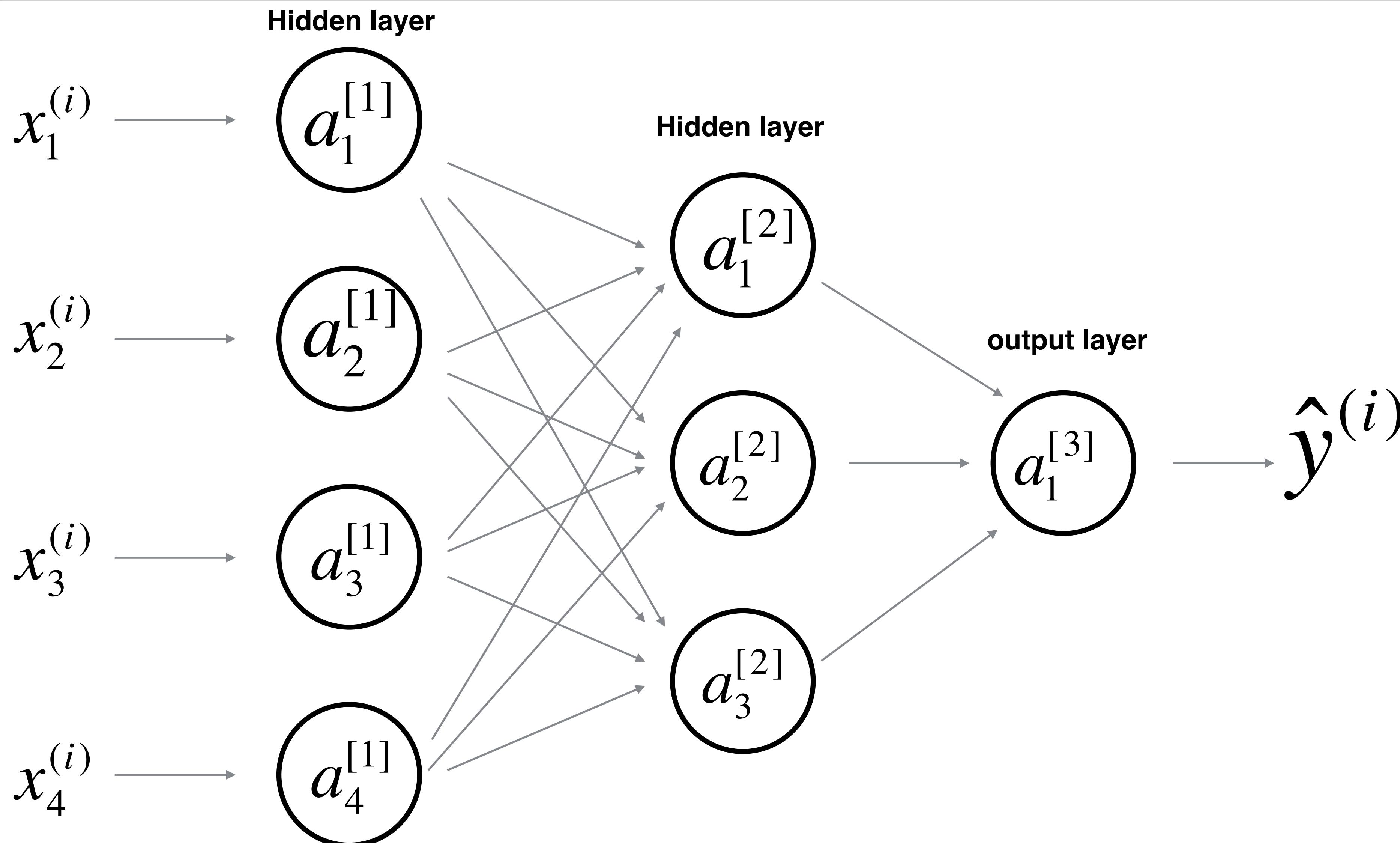
Neural Network (Multi-class)



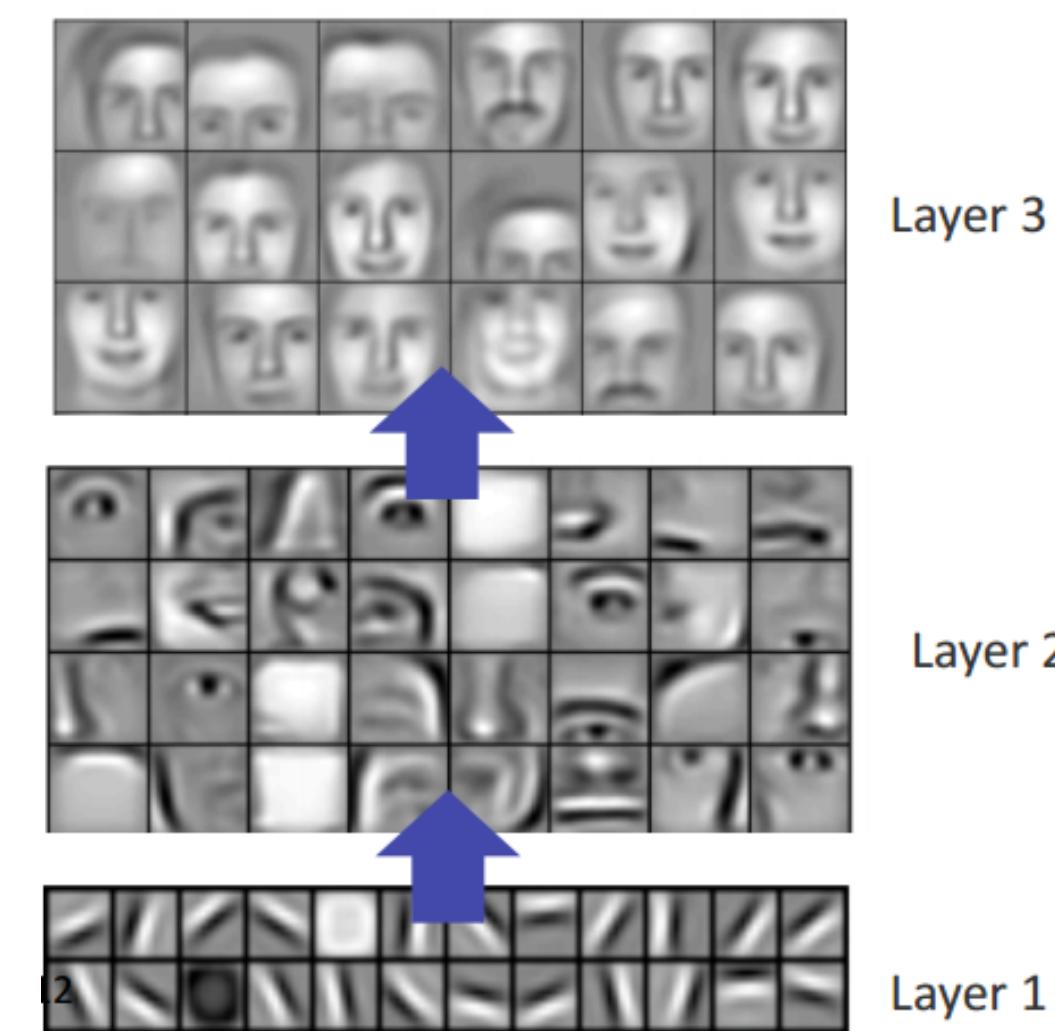
Neural Network (1 hidden layer)



Deeper network: Encoding



Technique called “encoding”



Let's build intuition on concrete applications

Today's outline

We will learn tips and tricks 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 verification and recognition
 - III. Neural style transfer (Art generation)
 - IV. Trigger-word detection
 - V. Shipping model

Day'n'Night classification

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 warm up

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

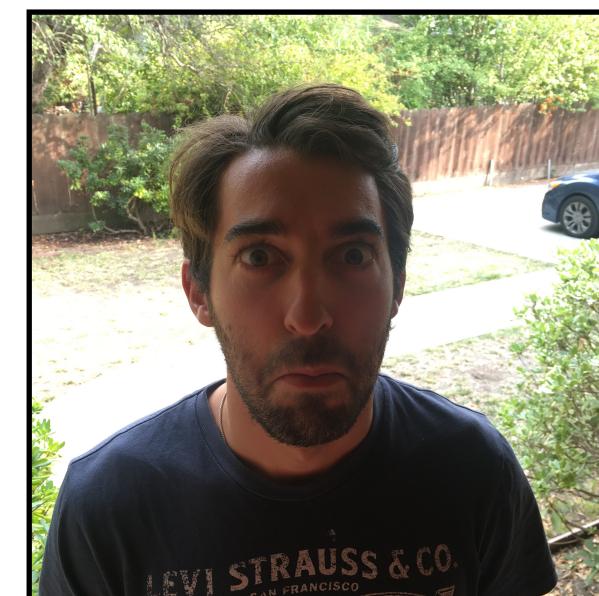
Goal: A school wants to use Face Verification for validating student IDs in facilities (dinning halls, gym, pool ...)

4. What architecture?

Simple solution:



compute distance
pixel per pixel
if less than threshold
then $y=1$



database image

input image

Issues:

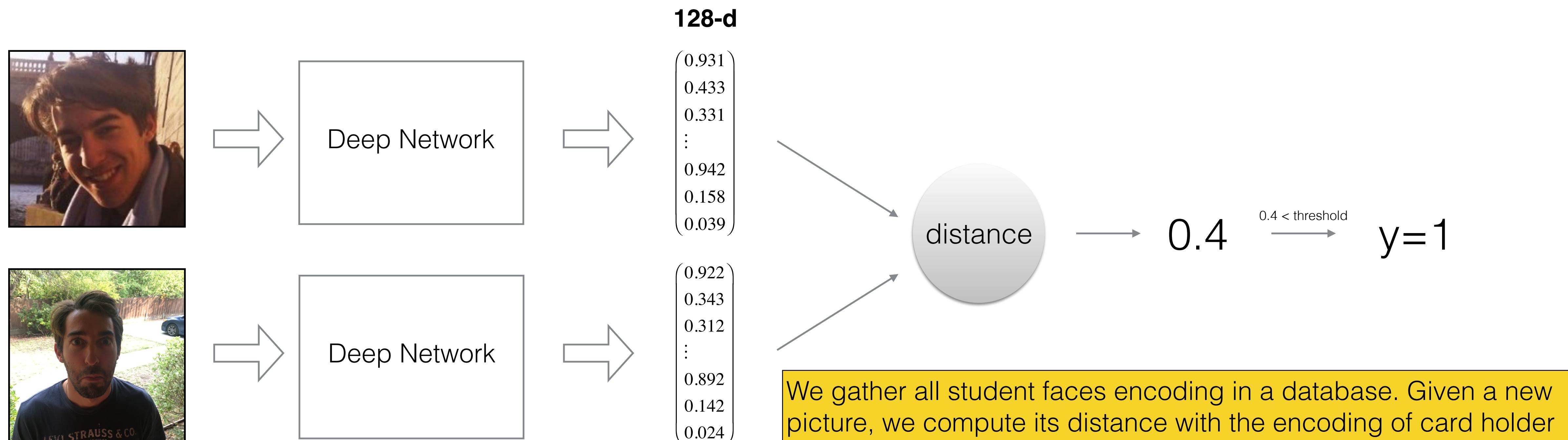
- Background lighting differences
- A person can wear make-up, grow a beard...
- ID photo can be outdated

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



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

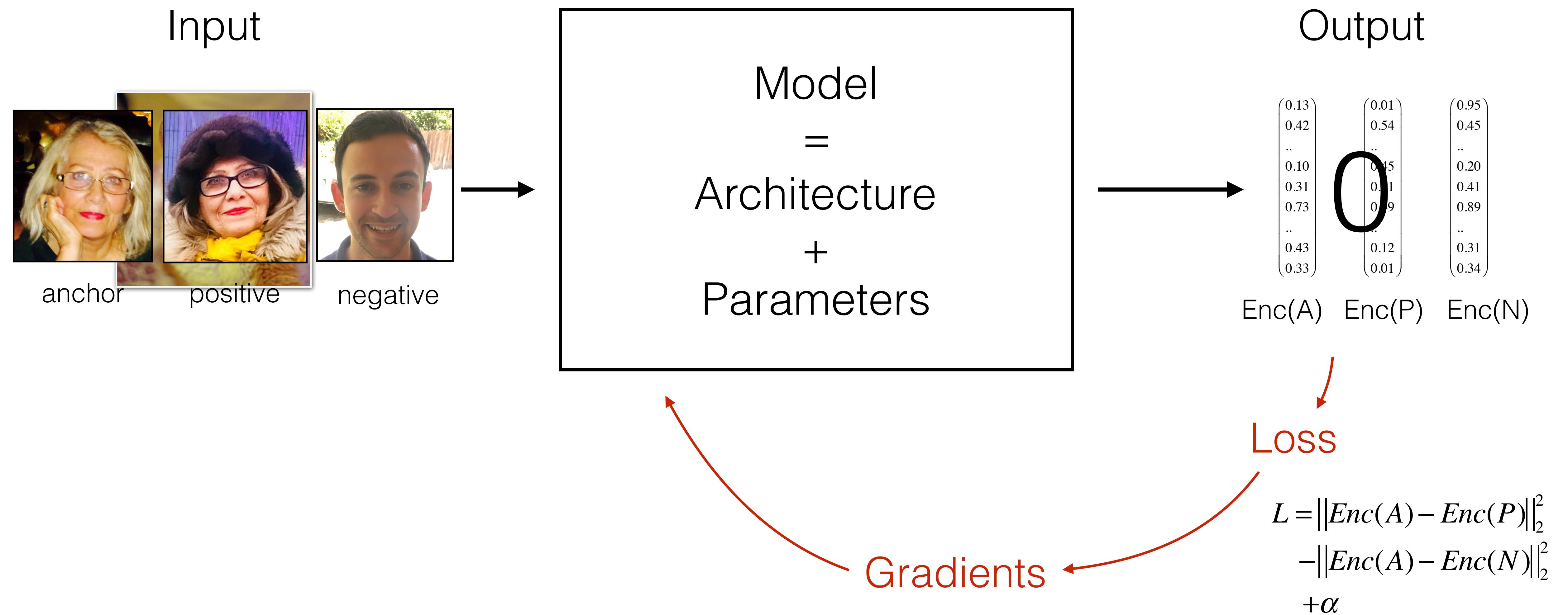


negative

minimize encoding distance

maximize encoding distance

Recap: Learning Process



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?

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

3. Output?



style
image

generated
image

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



When this image forward propagates, we can get information about its content & its style by inspecting the layers.

Content_C
Style_S

5. Loss?





A graphic element consisting of a central globe icon with a blue and white color scheme, set against a dark background. The globe is partially visible within a dark rectangular frame. The frame is flanked by two smaller, semi-transparent rectangular boxes, each containing a 2x2 grid of colored squares (light blue, medium blue, dark blue). Below this graphic, the text 'LiveSlides web content' is displayed, followed by a play button icon and instructions for viewing the content.



LiveSlides web content

To view

Download the add-in.

liveslides.com/download

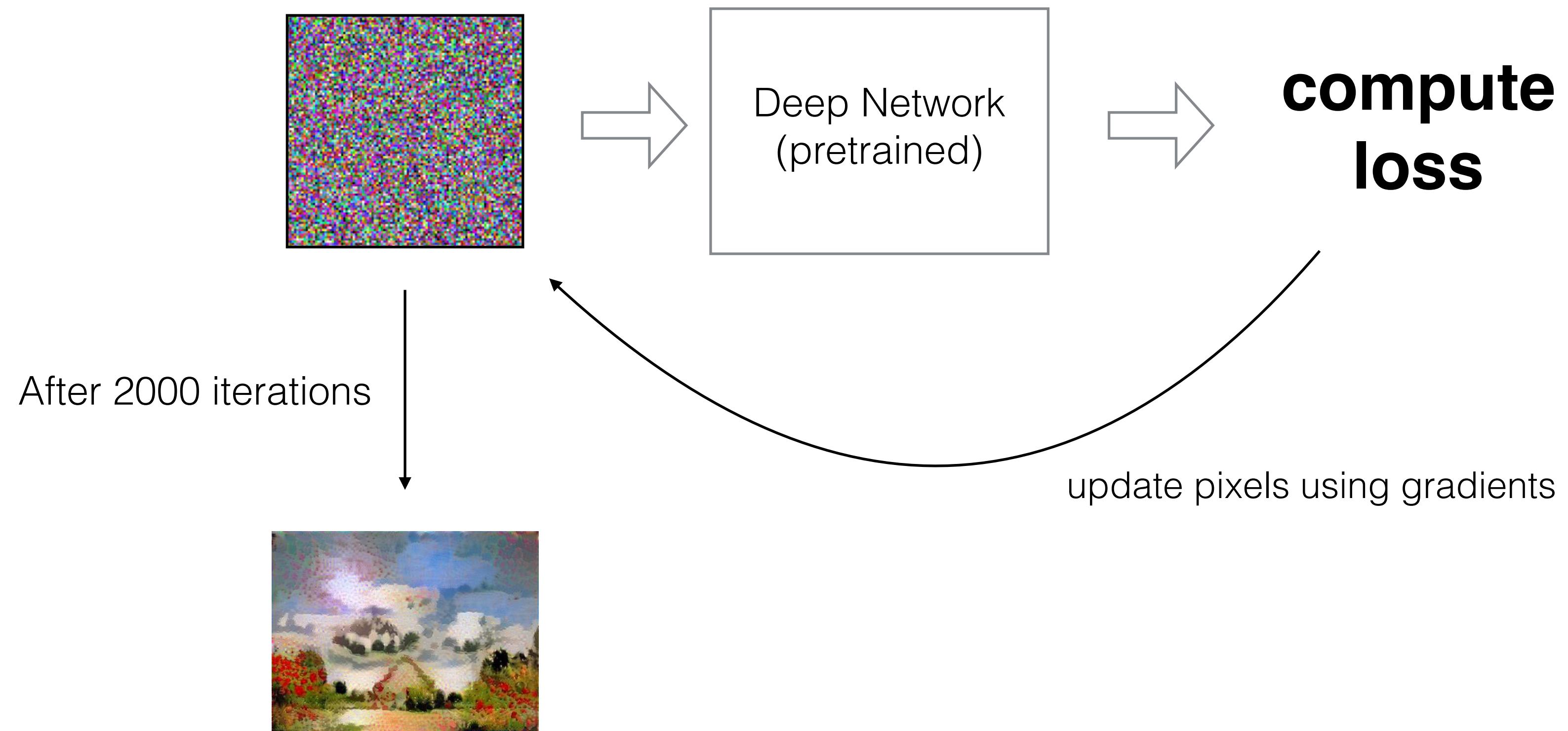
Start the presentation.

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!



Trigger word detection

Goal: Given a 10sec audio speech, detect the word “activate”.

1. Data?

A bunch of 10s audio clips

Distribution?

2. Input?

$x = \text{A 10sec audio clip}$

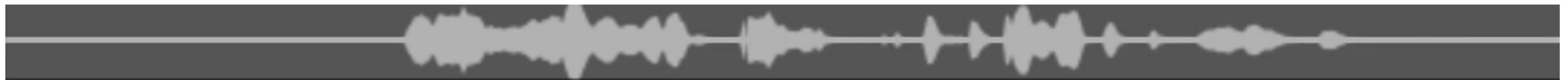


Resolution? (sample rate)

3. Output?

$y = 0 \text{ or } y = 1$

Let's have an experiment!



$$y = 1$$



$$y = 0$$



$$y = 1$$



$$y = 0000000000001000$$

Trigger word detection

Goal: Given a 10sec audio speech, detect the word “activate”.

1. Data?

A bunch of 10s audio clips

Distribution?

2. Input?

$x = \text{A 10sec audio clip}$



Resolution? (sample rate)

3. Output?

$y = 0 \text{ or } y = 1$

$y = 00..0000\mathbf{1}00000..000$

$y = 00..0000\mathbf{1}..1000..000$

Last Activation?
sigmoid
(sequential)

4. Architecture ?

Sounds like it should be a RNN

5. Loss?

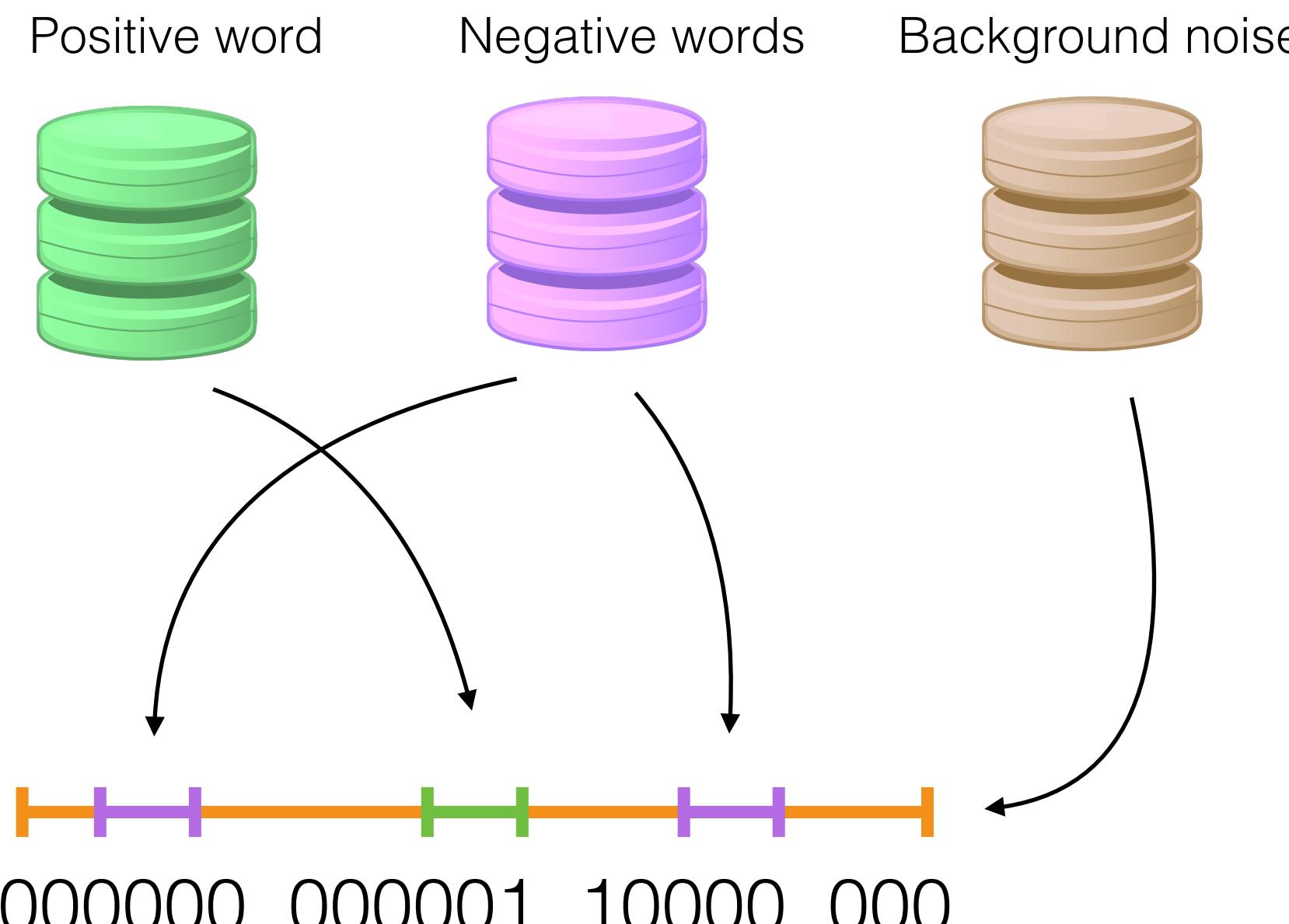
$$L = -(y \log(\hat{y}) + (1 - y) \log(1 - \hat{y}))$$

(sequential)

Trigger word detection

What is critical to the success of this project?

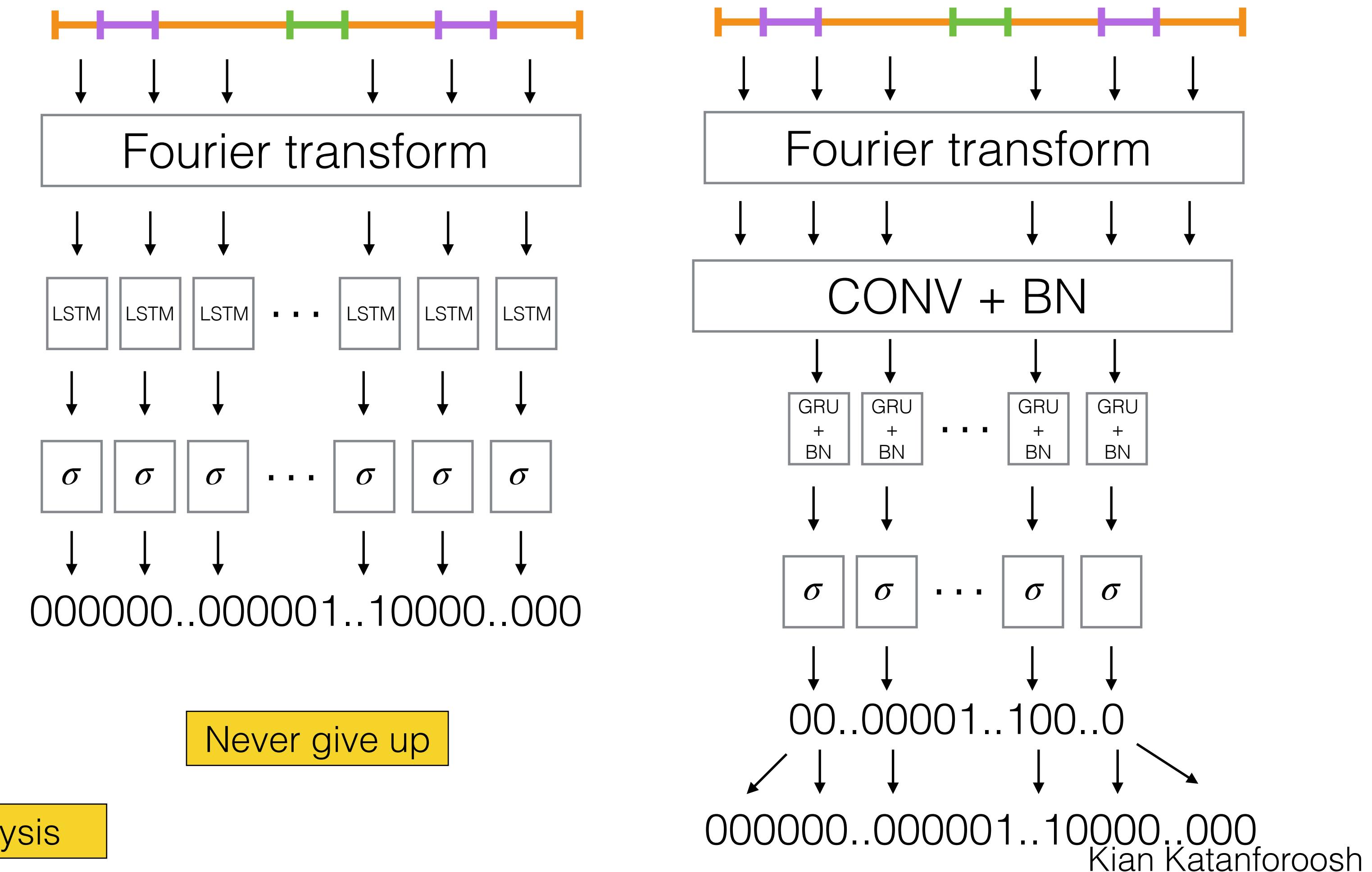
1. Strategic data collection/ labelling process



Automated labelling

+ Error analysis

2. Architecture search & Hyperparameter tuning



Kian Katanforoosh

Another way of solving the TWD problem?

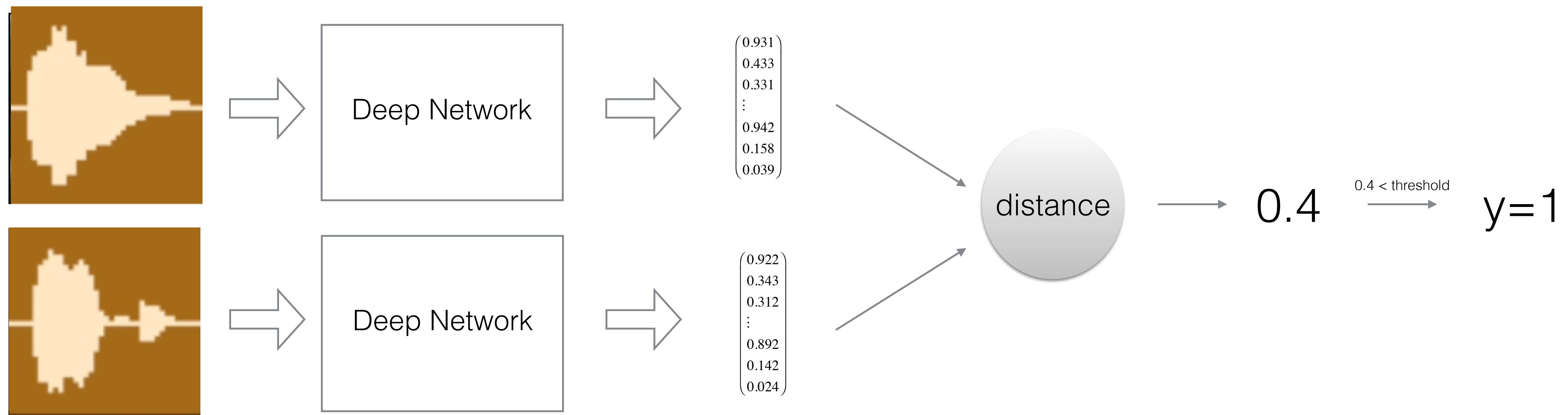
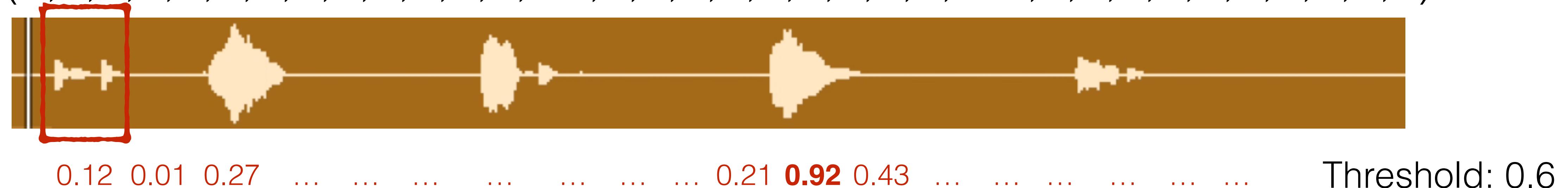
Trigger word detection (other method)

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?

$$y = (0, 0, 0, 0, 0, 0, 0, 0, 0, 0, \dots, 0, 0, 0, 0, 0, 0, 1, 0, 0, \dots, 0, 0, 0, 0, 0, 0, 0, 0)$$



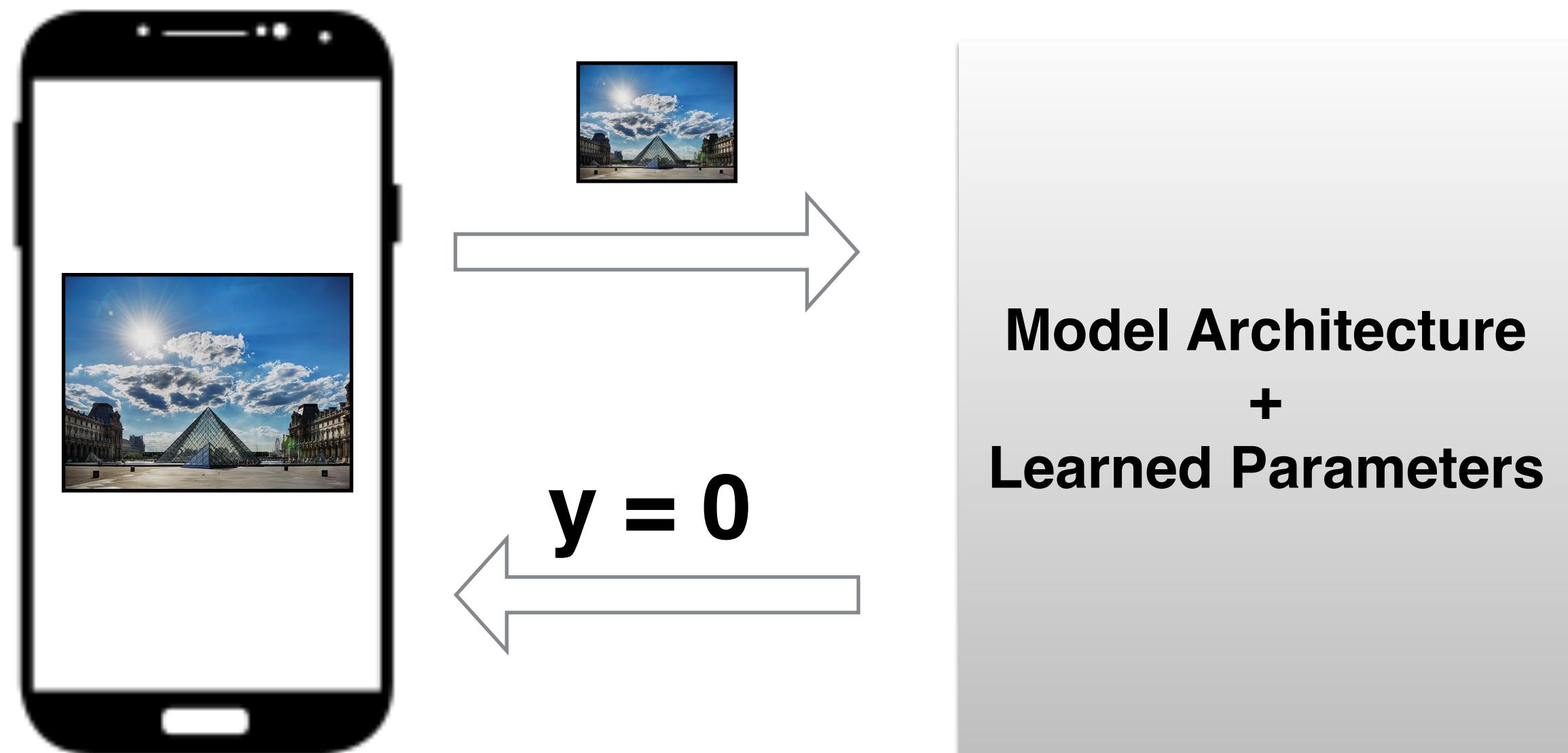
Featured in the Magazine “the Most Beautiful Loss functions 2015”

$$\begin{aligned}
& \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] \\
& + \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right] \\
& + \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} \left(C_i - \hat{C}_i \right)^2 \\
& + \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{noobj}} \left(C_i - \hat{C}_i \right)^2 \\
& + \sum_{i=0}^{S^2} \mathbb{1}_i^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2
\end{aligned}$$

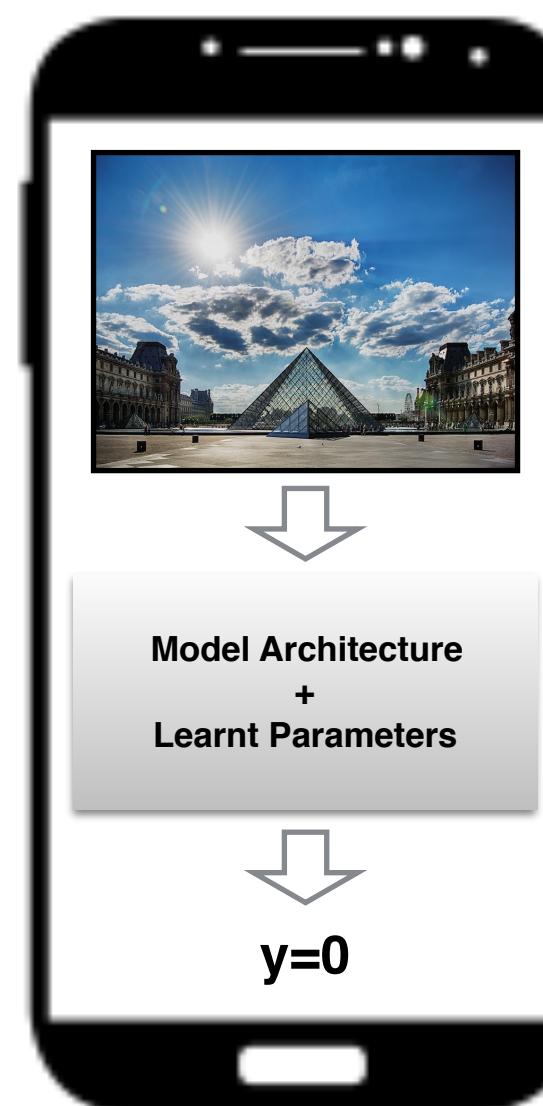
App implementation

Server-based or on-device?

Server-based



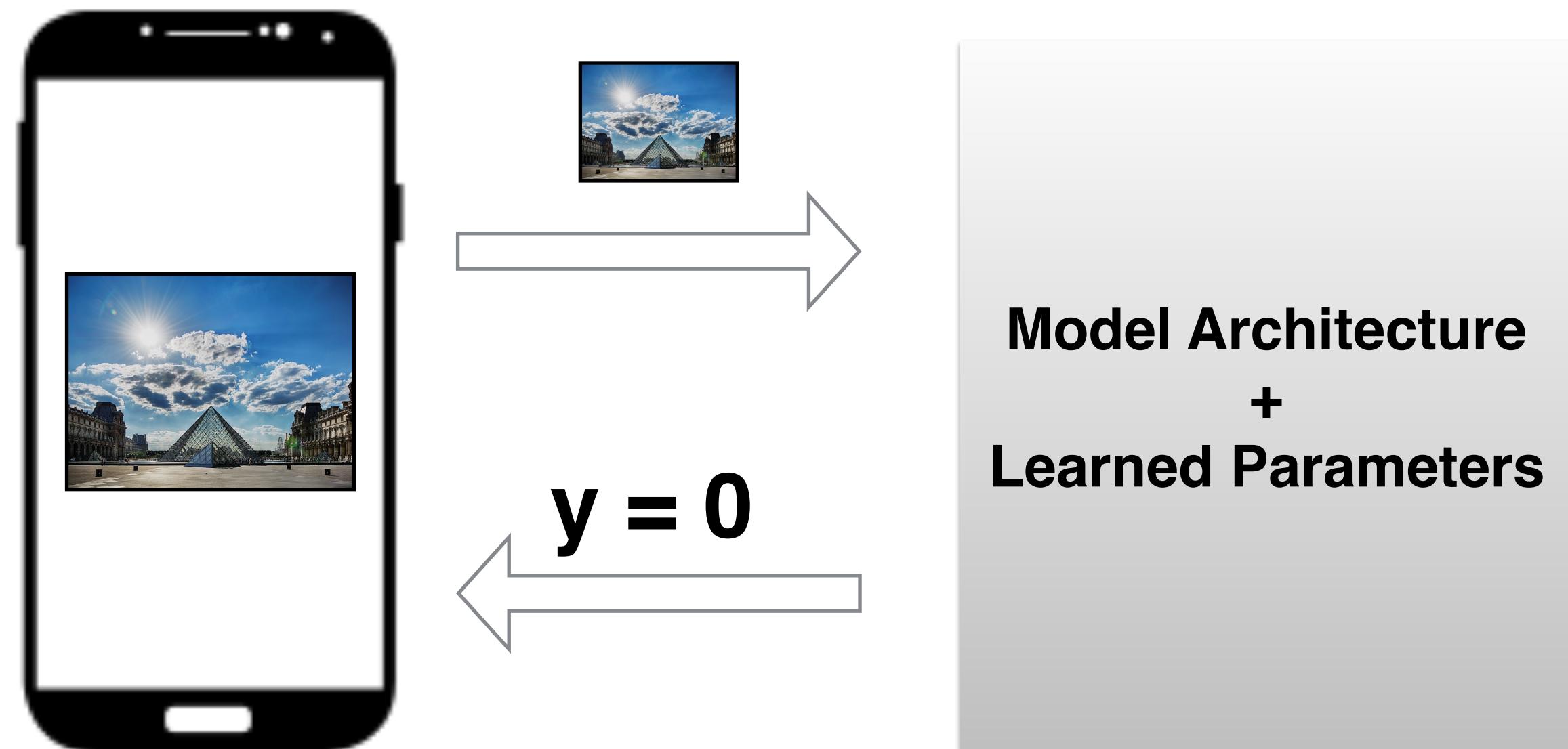
On-device



Server-based or on-device?

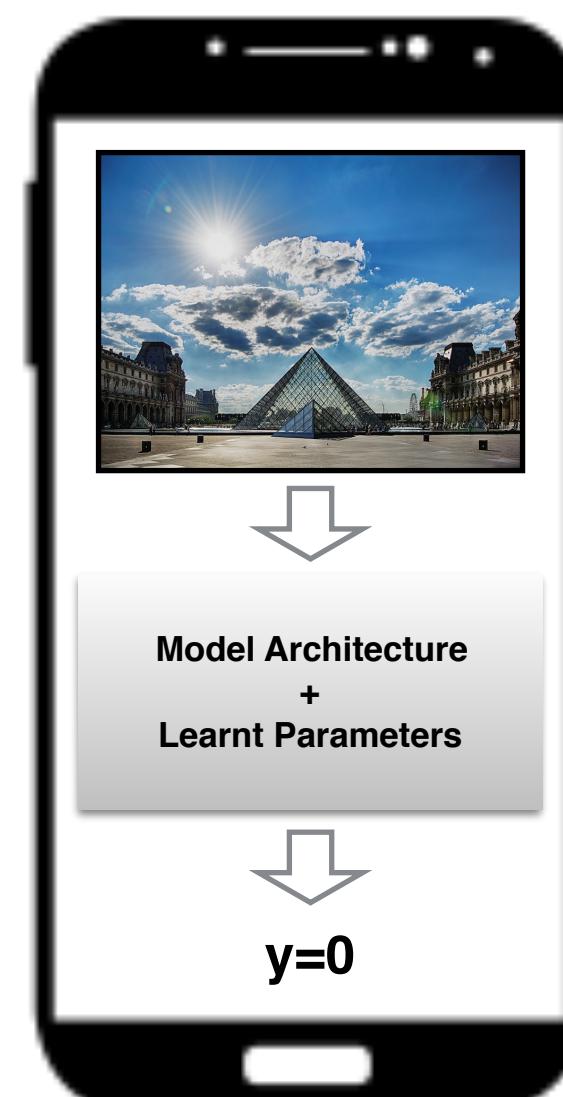
Server-based

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



On-device

- + Faster predictions
- + Works offline



Duties for next week

For Wednesday 10/09, 10am:

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

Others:

- TA project mentorship (mandatory this week)
- Friday TA section (10/05): focus on git and neural style transfer.
- Fill-in AWS Form to get GPU credits for your projects