

August 23 (Tue)

First Day of the Class

1. Organizational meeting

"Green Sheet"

Repo: github.com/kuaili10/opencv/deep-learning-2023/Email: hua.li@sjtu.edu(656) 490-1116 Cellphone for
Text message Only.

Office Hours: M.W. On Zoom

(See Syllabus for the
Zoom link).

2 Software Tools:

Anaconda — Install it by the end
of this week;

TensorFlow, TF 2.0

OpenCV.

3. Prerequisites: CMPS259 & CMPS251

Homework: To upload a copy of
your un-official transcript to
show the required courses satisfied.

On CANVAS.

4. Textbook: Deep Learning with Python.

Keras (API) for TF

Robot Vision Book By Horn (Horn, Hancock)

Book, (and Reference for OpenCV Algorithms. From the Architecture diagram:

Grad Theoretical Foundations)

5. Projects { Mandatory Assigned Project

Team Project
(Mandatory)(4-Person) Team. Presentation By the
End of the Semester.

August 25 (Wed)

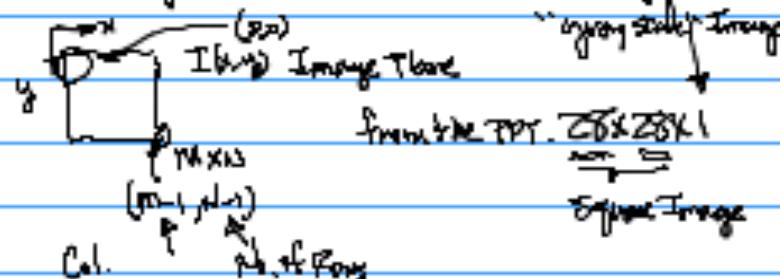
Note: 1. The lecture notes will be posted on
to the github.2. Zoom Recording will be posted on the
the github.Homework: By A week from today. 1. Anaconda
Installation; 2. OpenCV Installation. Submission
on CANVAS. $JPG/PNG/JPEG \rightarrow PDF \rightarrow ZIP$
 ZIP .

Example (github: 2022Fu-13)

Minist Architecture for Handwritten Digit

Recognition.

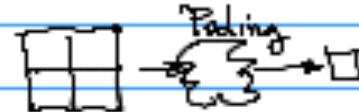
Single channel

(gray scale image \rightarrow 1 channel \rightarrow bit \rightarrow [0, 255])

First Layer of the Minst Architecture

1) Channel/Plane of the 1st convolutional layer

C1

Next, pooling \rightarrow Reduction \rightarrow Resolution \rightarrow 256 bins,

M1

P

F

L

I

O

R

S

T

U

V

W

X

Y

Z

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z

A

B

C

D

E

F

G

H

I

J

K

L

COMPENG

Aug. 25, 22

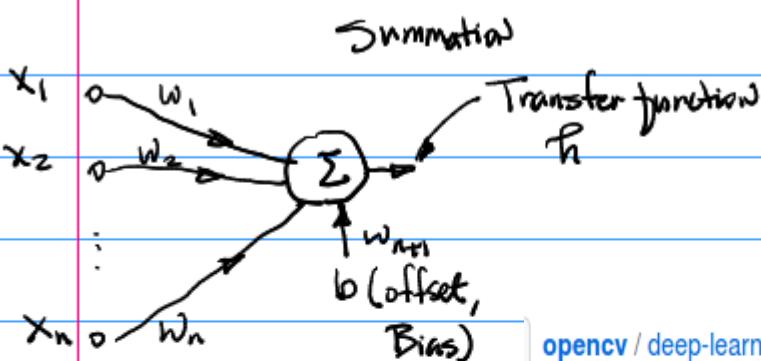
August 30, Tue

2/

Ref. 1

<https://github.com/hualili/opencv/blob/master/deep-learning-2022s/2022F-103b-NN-Intro-Python-v5-2022-8-25.pdf>

[opencv / deep-learning-2022s / 2022S-103a-notation-neuro-loss-function-2022-2-8.pdf](https://github.com/hualili/opencv/blob/2022S-103a-notation-neuro-loss-function-2022-2-8.pdf)



Input / Excitation in Vector Form: $\mathbf{x} = (x_1, x_2, \dots, x_n) \dots \mathbf{v}$

Weights, links each excitation to the Neuron Ref 2. Code

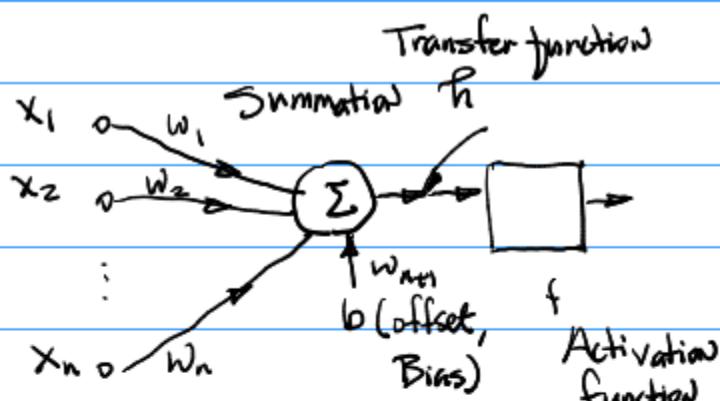
$$\mathbf{W} = (w_1, w_2, \dots, w_n) \dots (2)$$

<https://github.com/hualili/opencv/blob/master/deep-learning-2022s/2022S-110b-%232019S-31-6mnist-numerals-ch02.py>

$$x_1 w_1 + x_2 w_2 + \dots + x_n w_n + b w_{n+1} = f \quad \text{Example:}$$

$$\sum_{i=1}^n x_i w_i + b w_{n+1} = f(x_i w_i) \text{ or Simply } f(\mathbf{x} \mathbf{w}) \dots (3)$$

$$f(\mathbf{x} \mathbf{w}; b), f(\mathbf{x}, \mathbf{w})$$



$$f(f(\mathbf{x} \mathbf{w})) = f\left(\sum_{i=1}^n x_i w_i + b w_{n+1}\right) \dots (4)$$

$$f(\mathbf{x} \mathbf{w}) = \sum_{i=1}^n x_i w_i + b w_{n+1}$$

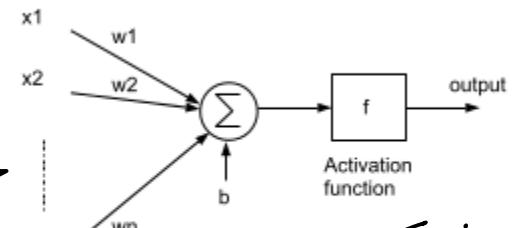


Fig. 1

(x_1, x_2, \dots, x_n) Feature Vector with Dimension N.

$$h = \sum_{i=1}^N w_i x_i = W \cdot X + b \quad (11)$$

Transfer function $f(\cdot)$.

$$w_{n+1} b = b'$$

Examples of Different Activation functions

include RELU. A piecewise linear.

Note: Be Able to Build A Single Neuron per a technical specification, Such as

ReLU, Activation $f(\cdot)$, Draw a Block

$$y = f(\sum_{i=1}^N w_i x_i = W \cdot X + b). \quad (17)$$

Activation function. Its output is the Response of the Neuron.

Aug. 30.

Consider the output of the Neuron

 y from Eqn(17).

Output of a Single Neuron.

For Multiple Neuron Output, see Fig.2

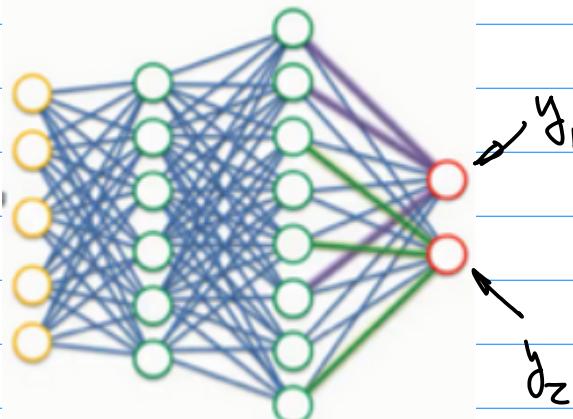


Fig.2

 y_{di} ... (1)Subscript: $i=1, 2$

No. of Output at the Output Layer.

 $y_i, i=1, 2, \dots, M$.

In practical Application,

 y_{ij}^* ... (2)

Superscript

 $j=1, 2, \dots, P$ No. of Experiments

Performed, Training Performed.

Look at the Concept & Definition of Loss function.

Mathematically To Compare a Neural Network Output (Single Neuron Output)

function f . function g Comparison of the Similarity or difference between f and g . $f - g$ f/g

Difference Between Two Functions.

Take this Approach to define Loss function,

 $y - \hat{y}$... (3)

Ground Truth.

Output (Prediction) from the Neuron

Outputs

-3 connected Network :activation

fc_4 Fully-Connected

Neural Network

(with dropout)

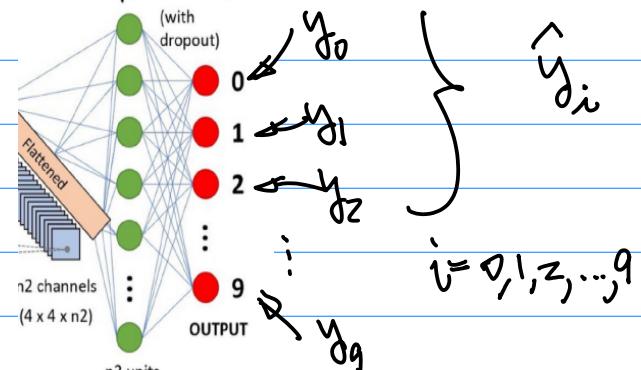
Flattened

n2 channels (4 x 4 x n2)

n3 units

OUTPUT

Fig.3.



August 30.

$$y_i - \hat{y}_i \quad \dots (4-a)$$

To measure All the content for Each

Training Experiment

$$\sum_{i=0}^q (y_i - \hat{y}_i) \quad \dots (4-b)$$

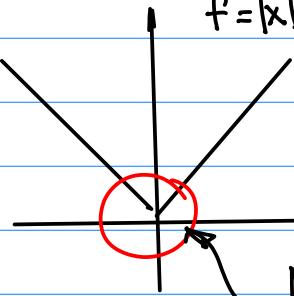
Expand this to Experiment/Training up to "P" Times

$$\sum_{j=1}^P \left[\sum_{i=0}^q (y_i - \hat{y}_i) \right] \quad \dots (4-c)$$

Note: Eqn (4-c) may lead to positive & Negative Terms Cancellation.

Fix: Absolute Value? \rightarrow Squared Instead,

$$f = |x|$$



Not "well-behaved"!

$$\frac{\partial L}{\partial w_{i,k}} = \frac{\partial}{\partial w_{i,k}} \frac{1}{2} \sum_{j=1}^P \sum_{i=1}^M (\tilde{y}_i^j - y_i^j)^2$$

(24)

Sept 1st (Th).

Ref:

[2022S-103a-notation-neuro-loss-function-2022-2-8.pdf](#)

Example: Background on "Learning" of ANNs.

[2022F-103b-NN-Intro-Python-v5-2022-8-25.pdf](#)

$$L = \sum_{j=1}^P \left[\sum_{i=0}^q (\tilde{y}_i^j - \hat{y}_i^j)^2 \right] \quad \dots (4-d)$$

J , or Φ

$$L_{total} = \frac{1}{2} \sum_{j=1}^P (\tilde{y}^j - y^j)^2$$

Ground Truth

(23)

For a Single Neuron @ the Output Layer

Training Based "Steepest Gradient Descent"

Example: Given A

function $f(x) = x^2$, Find its Derivative

$$\frac{df}{dx} = \frac{d}{dx} x^2$$

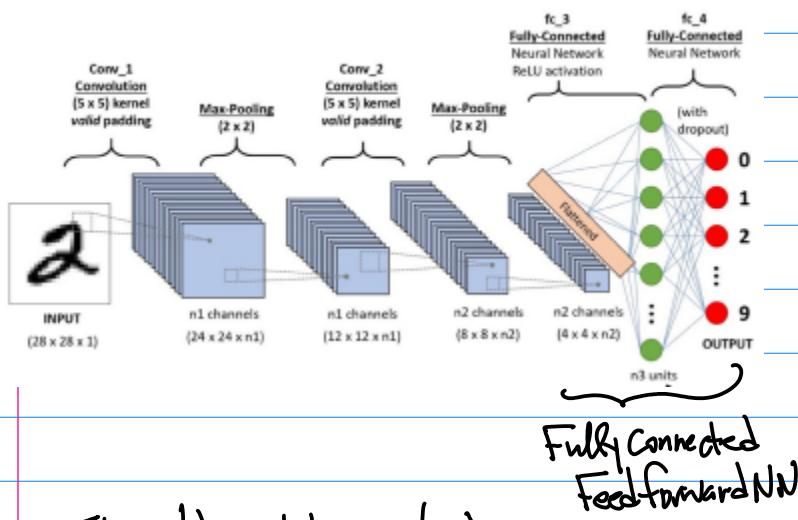
To get rid of Coefficient from the derivative,

$$\text{Let's define } f(x) \triangleq \frac{1}{2} x^2$$

$$\frac{d}{dx} f(x) = \frac{1}{2} \cdot 2 \cdot x = x$$

$$\frac{\partial L}{\partial w_{i,k}} = \frac{\partial}{\partial w_{i,k}} \frac{1}{2} \sum_{j=1}^P \sum_{i=1}^M (\tilde{y}_i^j - y_i^j)^2$$

(24)



To Train AN, we take eqn (23), e.g.
Loss function (error function),

$$L(\cdot) \rightarrow L(W_i)$$

Independent Variables

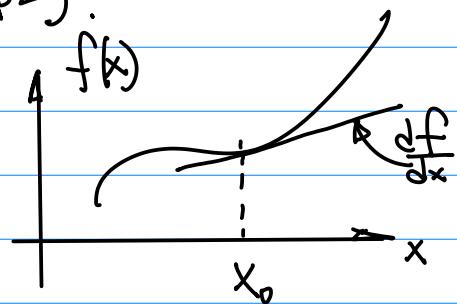
Minimize the Loss $L(\cdot)$, which is the process of Training, which leads to Learning for the NN.

Since the Loss function in Eqn (23) is formulated with a ground truth, this defines a Supervised Learning.

Math. Formula: Prediction of Function's Behavior, we like to know given the current function value (Loss function) how this function is going to change at next moment, increase? stay the same?

Or Decrease?

Derivative of A given function is a good indicator to give us the description of the function behavior
Next Step Ahead (Very small tiny step).



$$\frac{df}{dx} = \lim_{\Delta x \rightarrow 0} \frac{f(x+\Delta x) - f(x)}{\Delta x} \quad \dots (1)$$

f Intuition

$f(x+\Delta x) - f(x) \rightarrow > 0$ (derivative) if the derivative is greater than 0
 $\Delta x \approx 1$ unit

then $f(x+\Delta x) > f(x)$, the next Step function $f(x)$ is increased;

if the derivative $\frac{df}{dx} < 0$, then

$$f(x+\Delta x) < f(x)$$

if the derivative $\frac{df}{dx} = 0$, then
 $f(x+\Delta x) = f(x)$

Consider Two Dimensional Case as an Example for n-dimensional Case.

Sept. 1st.

6

$$f(x_1, x_2) \left\{ \begin{array}{l} \frac{\partial f(x_1, x_2)}{\partial x_1} \\ \frac{\partial f(x_1, x_2)}{\partial x_2} \end{array} \right. \rightarrow \text{Single Neuron}$$

Multi-dimensional Derivative

In Case of Training
They are "Weights"
 w_1, w_2 .

gradient

the Steepest gradient

the Steepest Descent
gradientThe core technique to
train NN. SGD

Ref: from the github

$$\frac{\partial f}{\partial x_i} \rightarrow \frac{\partial f}{\partial w_i} \text{ In the Context of Training.}$$

Conclusion: Use Partial Derivatives

With respect to the weight w_i as an indicator to measure if the Loss function is got reduced or not.

$$\frac{\partial L}{\partial w_{i,k}} = \frac{\partial}{\partial w_{i,k}} \frac{1}{2} \sum_{j=1}^P \sum_{i=1}^M (\tilde{y}_i^j - y_i^j)^2$$

Example: Consider A Technique which allows training to be more effective, e.g., to minimize training and prediction error (Loss) function

Steps for Development of this technique :

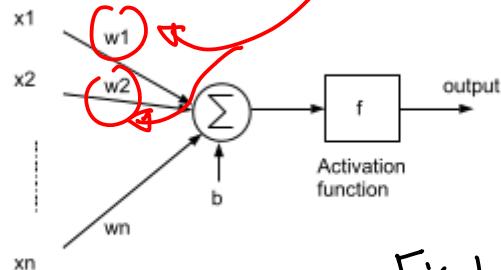


Fig.1

/ 2022S-105c-#20-2021S-4gradient-descent-v2-final-2021-2-8.pdf

Homework (Opt), Due A week from Today.
Sept. 8th, Thursday. ON CANVAS.

1^o Installation of TensorFlow, Version 2.1 or higher.

2^o Screen Capture to show the installation is successful

Note: All Different Development Tools/Environments including google colab, jupyter Note Book etc. Are OK, However for the Deployment purpose, projects homework Submission must be in Python Stand-Alone form.

Sept. 6 (Th)

Homework: Due 1 week from Today Sept 13.

1. OpenCV Installation, Python.

2. Use Smart phone to Capture
5~10 Seconds Video Clips.
.avi, .mp4 (.mpeg4).

3. Sample Code, github.
See CANVAS for the Detailed
Links & Requirements.

4. Submission to CANVAS.

- (1) Python Code;
- (2) Original & Processed image
Side by Side with your Name +
SID.
- (3) Create One pdf file to
Cover the Source Code, And
Screen Captured Images.

5. Naming Convention

HW-CV-First-LastName-Cmpe258-SID.zip

Example: Gradient Definition

Ref:

<https://github.com/hualili/opencv/blob/master/deep-learning-2022s/2022S-105c-%2320-2021S-4gradient-descent-v2-final-2021-2-8.pdf>

<https://github.com/hualili/opencv/blob/master/deep-learning-2022s/2022S-104d-%232-pdisplay-2019-1-30.py>

Higher Dimensional Function

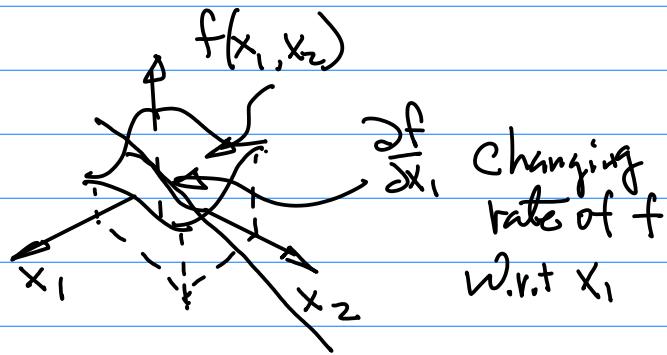
$$f(x_1, x_2, \dots, x_n) \quad \dots (z)$$

Weights, w_1, w_2, \dots, w_N

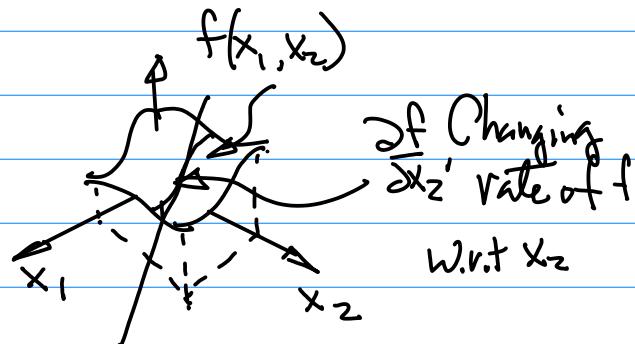
Partial Derivatives:

$$\frac{\partial f}{\partial x_1} \text{ w.r.t } x_1, \frac{\partial f}{\partial x_2} \text{ for } x_2, \dots$$

$$\frac{\partial f}{\partial x_n} \text{ w.r.t } x_n.$$



Changing rate of f
w.r.t x_1



Changing rate of f
w.r.t x_2

Loss function

Derivative, e.g. Given $f(x)$, then

$$\frac{df}{dx} = \lim_{\Delta x \rightarrow 0} \frac{f(x+\Delta x) - f(x)}{\Delta x} \dots (1)$$

$$\frac{\partial L}{\partial w_{i,k}} = \frac{\partial}{\partial w_{i,k}} \frac{1}{2} \sum_{j=1}^P \sum_{i=1}^M (\tilde{y}_i^j - y_i^j)^2 \quad (24)$$

Sept. 6, 22

Consider the minimization of function f (Loss Function) w.r.t. All possible weights.

Therefore, put all the partial derivatives together to form a vector, e.g., gradient.

$$\nabla f = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_i} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix} \dots (za)$$

for $n=2$,

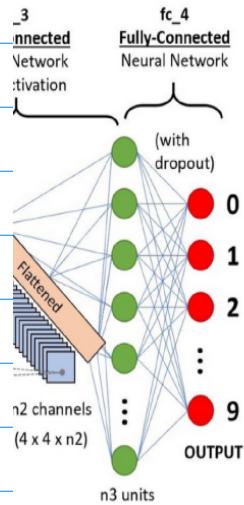
$$\nabla f(x_1, x_2) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{pmatrix} \dots (zb)$$

for $n=3$

$$\nabla f(x_1, x_2, x_3) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \frac{\partial f}{\partial x_3} \end{pmatrix} \dots (zc)$$

c. On the Right hand side of Eqn(5):
 (x_1^k, x_2^k) Dimension $n=2$, (x_1, x_2)
 Time Index "K", Superscript

Output of the NN with its weights at Step K (Time) is



$$x_1^k, x_2^k$$

e. On the left (x_1^{k+1}, x_2^{k+2}) , at the step $k+1$, to Reduce the Loss function, so update the new step by following

$$-\nabla f = -\eta \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{pmatrix} \dots (z)$$

IV. GRADIENT STEEPEST DESCENT FOR MINIMIZATION

Conclusion:

$$(x_1^{k+1}, x_2^{k+1}) = (x_1^k, x_2^k) + [-\eta(\nabla f)]$$

a. Loss function
 f

(5) b. $n=2$

e.g.

$$f(x_1, x_2, \dots, x_n)$$

$$\Rightarrow f(x_1, x_2)$$

c. Gradient

$$\nabla f(x_1, x_2), \text{ or}$$

$$\nabla f$$

Background:

Given a function $f(x)$, How do you Approximate this function By using Basic Building Blocks (B^3)?

$$f(x) = \text{Constant} + \text{A Linear Term} + \text{A Quadratic Term} + \text{A Cubic Term} + \dots \dots (4)$$

CmpE258

Sept 6, 22

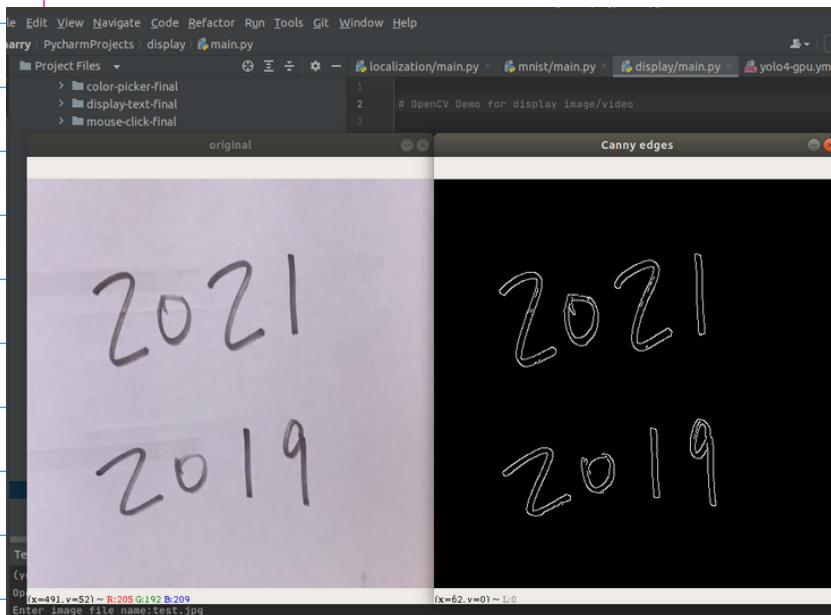
9.

Taylor Expansion:

$$f(x) = f(x_0) + \frac{df}{dx} \cdot (x - x_0) + \frac{d^2f}{dx^2} (x - x_0)^2 + \dots + R_n(x)$$

$$f(x_1, x_2) = f(x_1, x_2) \Big|_{\substack{\text{Constant} \\ x_1=x_{10} \\ x_2=x_{20}}} + \frac{\partial f}{\partial x_1}(x_1 - x_{10}) + \frac{\partial f}{\partial x_2}(x_2 - x_{20})$$

Note: The screen capture for your homework reference.



Sept. 8 (Thu)

Note: 1^o Check the CANVAS for Both Homeworks.

Example: From 1D Case in Eqn(4), we can expand the Taylor Expansion to higher dimension n. to Capture multiple excitations, multiple weights $w_i, i=1, 2, \dots, n$.

Consider n=2

$$+ \frac{\partial^2 f}{\partial x_1^2}(x_1 - x_{10})^2 + \frac{\partial^2 f}{\partial x_2^2}(x_2 - x_{20})^2 + \dots + R_n(x_1, x_2)$$

... (5)

Higher Order Terms

The goal is to verify the formula for updating the weights of A given NN. (see Handout 1, Eqn(5)).

Step 1: Taylor Expansion \rightarrow Step 2

Simplify the Taylor Expansion By just using upto the Linear terms \rightarrow Step 3: Re-arrange the Taylor Expansion in the form of Training formula (In Eqn(5), in Handout 1) \rightarrow Step 4: Analyze the re-arranged formula, to Reach the Conclusion, e.g., Using gradient descent, we can Reduce the Loss

function through each step of the training.

From the Handout, we have as Step 1 & 2:

$$f(x_1, x_2) \approx f(a, b) + \frac{\partial f}{\partial x_1}(x_1 - a) + \frac{\partial f}{\partial x_2}(x_2 - b) \quad (6)$$

$$\underbrace{f(x_1, x_2) - f(a, b)}_{\text{Comparison of A "loss" function}} \approx f_{x_1}(x_1 - a) + f_{x_2}(x_2 - b)$$

Comparison of A "loss" function, write

$$\Delta x_1 = x_1 - a$$

$$\Delta x_2 = x_2 - b$$

And $\nabla f = \begin{pmatrix} f_{x_1} \\ f_{x_2} \end{pmatrix}$

Hence, we have

$$f(x_1, x_2) - f(a, b) = (\Delta x_1, \Delta x_2) \cdot \nabla f$$

$$\text{Let } \Delta x_1 = -f_{x_1}, \Delta x_2 = -f_{x_2}$$

Therefore,

$$f(x_1, x_2) - f(a, b) = (-f_{x_1}, -f_{x_2}) \cdot \begin{pmatrix} f_{x_1} \\ f_{x_2} \end{pmatrix}$$

$$= -\left(f_{x_1}^2 + f_{x_2}^2\right) \leq 0$$

Hence, $f(x_1, x_2) - f(a, b) \leq 0$

OR, $f(x_1, x_2) \leq f(a, b)$

Loss function updated at the next step by Eqn(5) in Handout 1.

Note: The requirement for this discussion on Notations, And Formulation, especially, the Eqn(5) in Handout 1 is required

To Be Able to use these tools to analyze the problem, And to perform Verification (eng. design).

Example: [2022F-103b-NN-Intro-Python-v5-2022-8-25.pdf](#)

To Be Continued.

OpenCV Homework. Sample code

[opencv/2022S-104d-#2-pdisplay-2](#)

```

9     #include<sys.argv[1]>
10    window_name = 'Display Image'
11
12    imageName = sys.argv[1] #get file name from command line
13
14    src = cv2.imread(imageName, cv2.IMREAD_COLOR)
15
16    if src is None:
17        print ('Error opening image!')
18        print ('Usage: pdisplay.py image_name\n')
19
20    ind = 0
21
22    while True:
23        cv2.imshow(window_name, src)
24
25        c = cv2.waitKey(500)
26        if c == 27: #ESC

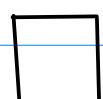
```

Loss function at the current step.

(0,0) \rightarrow $(m-1, N-1)$ \rightarrow $m \times N$

Note: OpenCV Resize Function.

Resize



```

12 import numpy as np
13 #import argparse
14 import cv2
15
16 img = input('Enter image file name:')
17
18 image = cv2.imread(img, cv2.IMREAD_COLOR)
19
20 if image is None:
21     print('Error opening image!')
22     print('Usage: pdisplay.py image_name\n')
23
24 image = cv2.resize(image, (512, 512))
25
26 gray = cv2.cvtColor(image, cv2.COLOR_BGR2GRAY) a.
27 edges = cv2.Canny(gray, 100, 200) b.

```

- a. Conversion to Gray-Scale Image;
- b. Canny Edge Detection.

c. Preprocessing of the dataset

Name	x_1	Weight (minus 135)	Height (minus 66)	Gender
Alice	-2	-1	1	
Bob	25	6	0	
Charlie	17	4	0	
Diana	-15	-6	1	

Sept.13 (Tue)

Homework: (opt) Due to A week from today. Capture the screen to the T.F. Installation is successful.

Example: [2022F-103b-NN-Intro-Python-v5-2022-8-25.pdf](#)

a. Feature Vectors

Example: Collecting data for training

Name	Weight (lb)	Height (in)	Gender
Alice	133	65	F
Bob	160	72	M
Charlie	152	70	M
Diana	120	60	F

Sig	Mij	Mpq
V1	133	65
V2	160	72
V3	152	70
V4	120	60

Sign
Stop
Right
Right
Stop

$$V = (v_1, v_2)$$

$$\Xi = (x_1, x_2)$$

$$\Xi_i = (x_{i1}, x_{i2})$$

b. Ground

Truth y_i corresponds to Ξ_i

Simple feedforward NN.

Step1. Match the Dimension of the

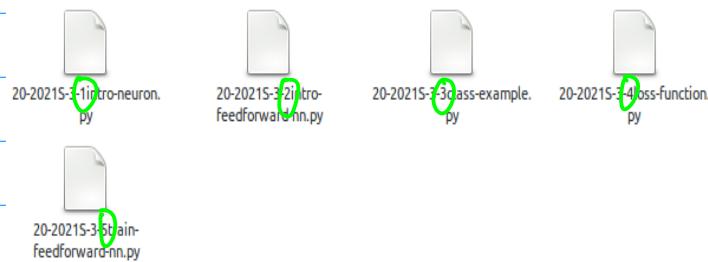
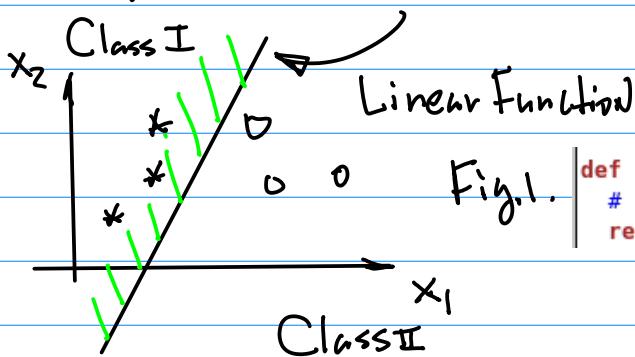
Feature Vector Ξ_i to the Input Neurons

$$\frac{\partial L}{\partial w_{i,k}} = \frac{\partial}{\partial w_{i,k}} \frac{1}{2} \sum_{j=1}^P \sum_{i=1}^M (\tilde{y}_i^j - y_i^j)^2$$

(24)

Also, Determining the Output Layer: Now, Let's use Python Code to Inspect the Nature of this Prediction, Hence, Choose One Single Neuron Output

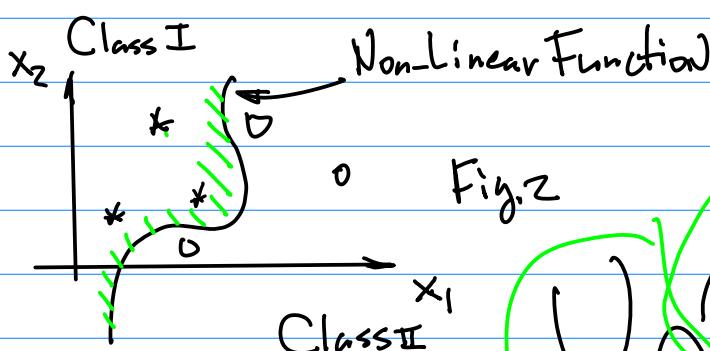
Hidden Layer(s) → Affects the Behavior of Decision-making function.



Define An Activation function.

$$y = f(h(w_i x_i; b)) = f(\mathbf{I} \mathbf{w})$$

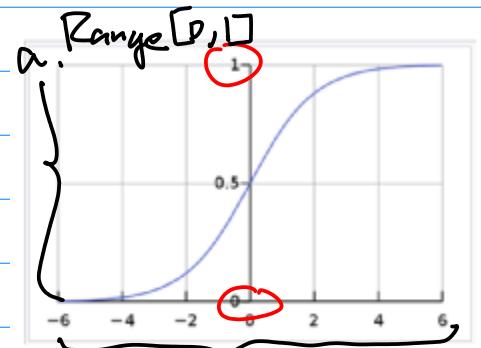
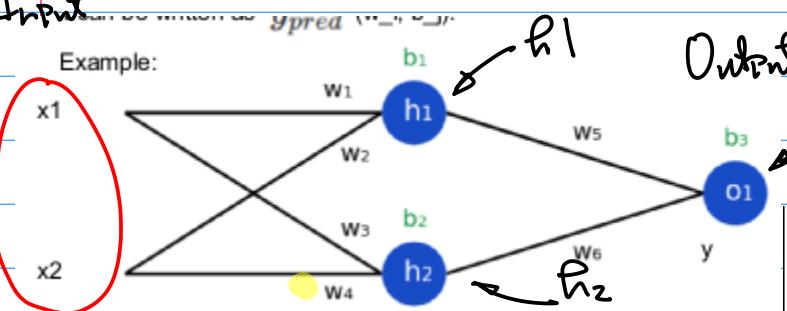
```
def sigmoid(x):
    # Sigmoid activation function: f(x) = 1 / (1 + e^(-x))
    return 1 / (1 + np.exp(-x))
```



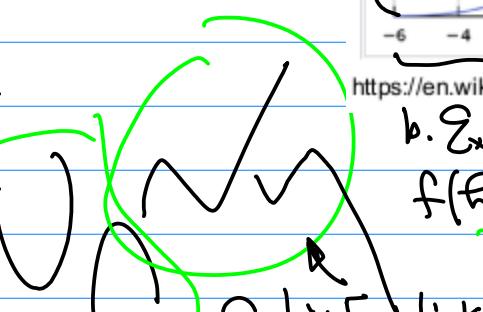
No. of Hidden Layers →

For the Simplicity in this example, Let's choose just 1 Hidden Layer.

\mathbf{I} Input



b. Excitation Range.
 $f(h(\mathbf{I} \mathbf{w}))$



$$S(x) = \frac{1}{1 + e^{-x}} = \frac{e^x}{e^x + 1}$$

Note: A Sigmoid Function. (1)
A sigmoid function is a

The Derivatives of the Activation Function. Done By hand Calculation of the Derivative.

```
def deriv_sigmoid(x):
    # Derivative of sigmoid: f'(x) = f(x) * (1 - f(x))
    fx = sigmoid(x)
    return fx * (1 - fx)
```

$$\frac{d}{dx} \text{Sigmoid}(x) = \frac{d}{dx} \frac{1}{1+e^{-x}}$$

Define the Loss function

$$\frac{\partial L}{\partial w_{i,k}} = \frac{\partial}{\partial w_{i,k}} \left(\frac{1}{2} \sum_{j=1}^P \sum_{i=1}^M (\tilde{y}_i^j - y_i^j)^2 \right)$$

```
def mse_loss(y_true, y_pred):
    # y_true and y_pred are numpy arrays of the same length
    return ((y_true - y_pred) ** 2).mean()
```

Mean Square Error
Ground Truth NN output
 \tilde{y}_i^j

Note: The Derivative Code in
 this Example Allows Training
 Implement By Gradient Descent.

```
# Biases
self.b1 = np.random.normal()
self.b2 = np.random.normal()
self.b3 = np.random.normal()

def feedforward(self, x):
    # x is a numpy array with 2 elements.
    h1 = sigmoid(self.w1 * x[0] + self.w2 * x[1] + self.b1)
    h2 = sigmoid(self.w3 * x[0] + self.w4 * x[1] + self.b2)
    o1 = sigmoid(self.w5 * h1 + self.w6 * h2 + self.b3)
    return o1
```

(24)

Define the F.F. Neural Network.

Output:

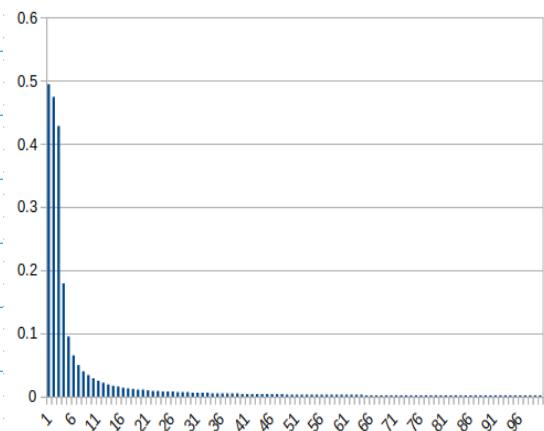
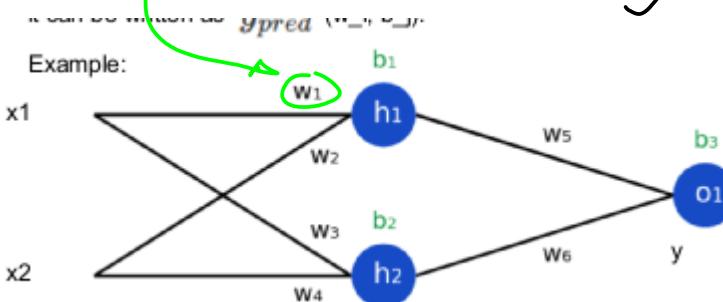
$$y = f(wx) \quad |_{f = \frac{1}{1+e^{-x}}}$$

where x is from the
 Hidden Neurons, h_1 & h_2

https://github.com/hualili/opencv/blob/master/deep-learning-2022s/2022S-103c-%23nn_sample_2022.py

The weight initialization.

```
def __init__(self):
    # Weights
    self.w1 = np.random.normal()
    self.w2 = np.random.normal()
    self.w3 = np.random.normal()
    self.w4 = np.random.normal()
    self.w5 = np.random.normal()
    self.w6 = np.random.normal()
```



Sept 15 (Thu)

Today's Topics:

- 1° Preliminary Sample Code (Python)
- 2° Preprocessing Technique (Computer Vision / OpenCV) for Deep Convolutional NN. Handwritten Digits Recognition

Ref: From Python Sample Code on the github

loss is being Computed \rightarrow

98

 $d_L_d_{ypred} = -2 * (y_{true} - y_{pred})$

```
74     def train(self, data, all_y_trues):
157     #-----
158     # Define dataset and all_y_trues
159     #-----
160     data = np.array([
161         [1, 2.5],      # person A
162         [1, 3],       # person A
163         [2.1, 3.4],   # Person A
164         [2.1, 1],     # person B
165         [3.3, 1],     # person B
166         [3, 2.3],     # person B
167     # HL 2020-9-7 Part E
168     # for the testing of adaptive learning
```

a. Training will take the input from here

then, Compute partial derivatives to form gradient

then, update the weight(s) Based on gradient descent

b. Initialization. w

80 learn_rate = 0.1

81 epochs = 1000 # number of times to loop through

The Number of Trainings is defined with this as a upper Bound if the loss function during the training drops below the pre-set threshold then the training Process will terminate

The Data Point (e.g. $\mathbf{x}(x_1, x_2)$) is substituted into the code to evaluate Transfer function $f(w, x)$, then evaluate the Activation function $f(f(w, x)) = \frac{1}{1 + e^{-x}}$

$$x = f(w, x)$$

```
85     # --- DO A FEEDFORWARD (WE'LL NEED THESE VALUES LATER)
86     sum_h1 = self.w1 * x[0] + self.w2 * x[1] + self.b1
87     h1 = sigmoid(sum_h1)
```

The process propagates through each neuron & each layer till reaches the Output

```
92     sum_o1 = self.w3 * h1 + self.w6 * h2 + self.b3
93     o1 = sigmoid(sum_o1)
94     y_pred = o1
```

96 # --- Calculate partial derivatives.
97 # --- Naming: d_L_d_w1 represents "partial L / partial w1"
98 d_L_d_ypred = -2 * (y_true - y_pred)

Compute total Loss function

```
136
137     self.w1 := learn_rate * d_L_d_ypred * d_ypred_d_h1 * d_h1_d_w1
138     self.w2 := learn_rate * d_L_d_ypred * d_ypred_d_h1 * d_h1_d_w2
139     self.h1 := learn_rate * d_L_d_ypred * d_ypred_d_h1 * d_h1_d_h1
```

--- Calculate total loss at the end of each epoch
if epoch % 10 == 0:
 y_preds = np.apply_along_axis(self.feedforward, 1, data)
 loss = mse_loss(all_y_trues, y_preds)
 print("Epoch %d loss: %.3f" % (epoch, loss))

$$\frac{\partial L}{\partial w_{i,k}} = \frac{\partial}{\partial w_{i,k}} \left(\frac{1}{2} \sum_{j=1}^P \sum_{i=1}^M (\tilde{y}_i^j - y_i^j)^2 \right) \quad (24)$$

Note: the ground truth is given in the code

```
178     all_y_trues = np.array([
179         1, # person A
180         1, # person A
181         1, # person A
182         0, # person B
183         0, # person B
184         0, # person B
185     ])
```

Annotation

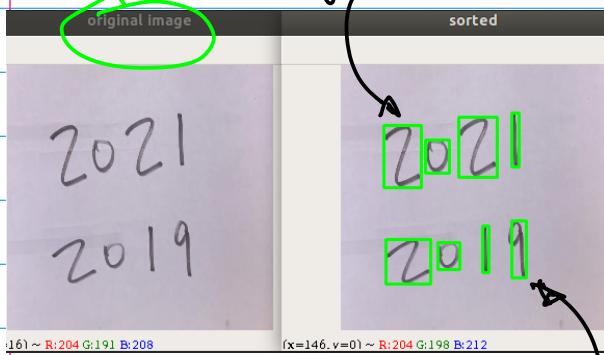
CMPT258
Sept. 15, 22

15

Example: Pre-processing Techniques.
for Handwritten Digits Recognition.

Objective: To Be Able to Localize
the ROIs. (Region of Interests).

a. ROI: Bounding Boxes



b. Localization: localized on
Each hand written digits
Ref: 1. pp. 18 (Starting from)

2022S-101-cmpe258-2022-03-15-Note.pdf

2. 2022S-108-Intro-Binary.pdf PPT.
with math. Formulation.

Background on Color Images.

a. Video Clip.
(A Sequence of Images)
From Ref. 1.

b. A Single frame of
an image $I(x,y,t) \rightarrow I(x,y)$

c. ... (1) at time t

Each Frame of A color image
 $I(x,y)$ consists of 3 channels.

Or. Primitive Color planes, Red,
Green, Blue

In OpenCV, (Color) Image Planes is organized
as B, G, R.

$$I_r(x,y) \in [0,255], I_g(x,y) \in [0,255], I_b(x,y) \in [0,255]$$

8 bit
 $2^8 = 256$
... (z)

Color Image \rightarrow Gray Scale Image.
Example in Fig. 1.
Fig. 1. $I(x,y)$ $I_{gray}(x,y)$

$$I_{gray}(x,y) = \frac{1}{3} (I_r(x,y) + I_g(x,y) + I_b(x,y))$$

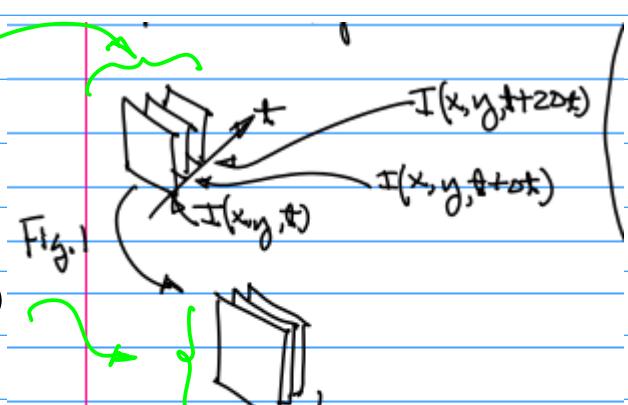
... (3)

↓

Binary Image.

$$B(x,y) = \begin{cases} 255 & \text{if } I_{gray}(x,y) \geq T \\ 0 & \text{o/w} \end{cases}$$

... (4)



3 Image Planes (r, g, b)
Pixel depth, Color Depth: 8 bits per
Each Color plane
 $r \in [0,255], g \in [0,255], b \in [0,255]$

CMPE258
Sept. 15/22

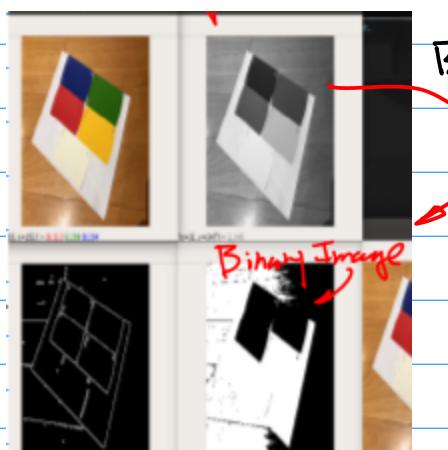
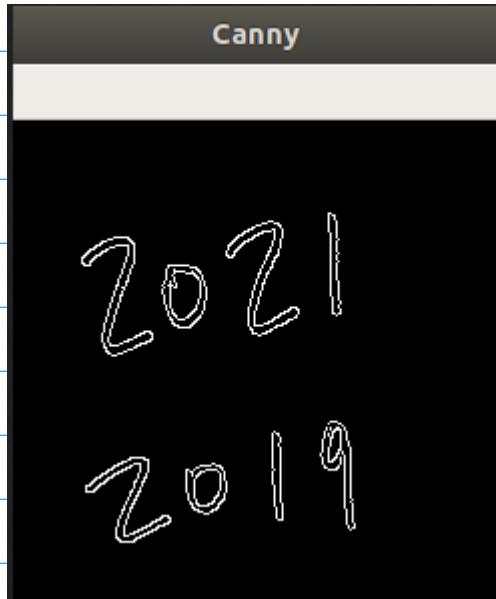
16

Example of A Binary Image
(This Binarized image is
Canny Edge Detection) Not Commonly used

From Eqn(4) Example:

2022S-108c-example-binary.pdf

Binarization Image.



By Eqn(4)

Binary Image

2022S-108-Intro-Binary.pdf

Operators for
Preprocessing (nb)

Sept 20 (Tuesday)

Note: 1° Next Class we will have in-class exercise, Regnies Run OpenCV Program, a) Have your Laptop Ready to Run OpenCV Program; b) Bring A blank Printer Paper to the Class, to Be used in the exercise; c) Bring your Smart phone; d) Check to Make sure your Video is working.

2° A New Homework will be posted on CANVAS tomorrow, please check.

3° For the Semester-Long Team Project, we will post the Rubrics & Requirements on Line (on CANVAS).

Pattern Recognition For Binary Images

The tool box for pattern recognition for binary images

- 1. Size
- 2. Moments
 - \bar{x}
 - \bar{y}
 - x^k
 - y^k etc.
- 3. Perimeter
- 4. Orientation
- 5. Compositions of the above
 - Perimeter and moments: vector
- 6. Invariant operators
 - size invariant
 - orientation invariant
 - illumination invariant

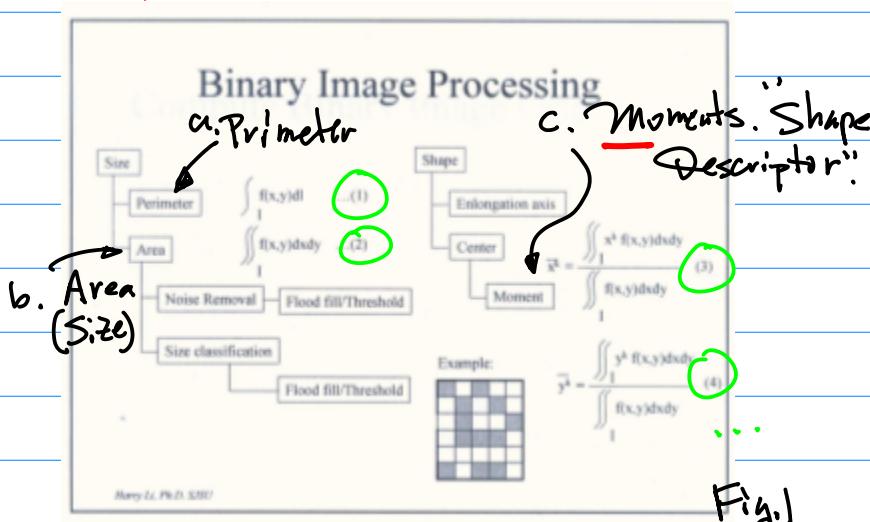
Biologically inspired techniques

- Rule 1. Proximity
- Rule 2. Similarity
- Rule 3. Closure
- Rule 4. Good continuation
- Rule 5. Symmetry
- Rule 6. Simplicity

Note: 'Proximity' usage for clean up binary image and remove noise, as well as growing boundary points per 'good continuation' rule to form a better edge map.

Note: Similarity defines a interesting question, how to describe one object is similar, or somewhat similar to others, neural network and fuzzy logic may help.

20225-108-Intro-Binary.pdf


 $f(0,0) < T$, Hence $B(0,0) = 0$;
 $\text{for}(0,1)$, Similarly, $f(0,1) = 70$, Hence,

$$B(0,1) = 0$$

...

...

 $\text{for}(3,0)$, $f(3,0) = 21 \geq T$, Therefore

$$B(3,0) = 255$$

e.g. often this
255 in Hand
Calculation

Can be replaced by 1.
(Scaling factor)

Example: Calculation of Binary Image.
Based on Eqn (4)

$$B(x,y) = \begin{cases} 255 & \text{if } I_{gray}(x,y) \geq T \\ 0 & \text{o/w} \end{cases} \dots (4)$$

 $\text{for}(3,1)$, $f(3,1) = 209 \geq T$, Therefore
 $B(3,1) = 255$

20225-108c-example-binary.pdf

	7	$f(x,y)$
70	70	80
75	85	140
90	210	101
211	209	115

Given an image $f(x,y)$

Find a proper Threshold

 T , use the T to

Perform Binarization.

Selection of "T" is based on Heuristics,
e.g. Human Expert Knowledge.

if $f(x,y)$, at Location (x,y) , is
greater to Equal to T .

$$f(x,y) \geq T ?$$

$\text{for}(0,0)$, $f(0,0) \geq 135$; $\therefore f(0,0) = 70$,

		K_x
K_y	0	0 0 0 255
0	0	0 255 0 0
255	255	0 0

No. of pix \leq
Per Row
 \times
No. of
Rows per
frame.

Example: Calculate the Size, e.g. Area, of the Binary Image $B(x,y)$.

Definition 1. (AREA/Size)

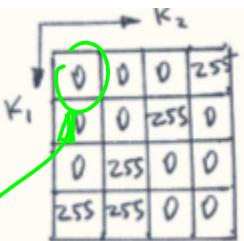
$$\iint_{\Omega} B(x,y) dx dy \dots (1)$$

Digitized Image / Discrete Function

$$\iint_{\Omega} \rightarrow \sum_{y=0}^{M-1} \sum_{x=0}^{N-1}$$

Hence,

$$\sum_{y=0}^{m-1} \sum_{x=0}^{N-1} B(x, y) \dots (1-b)$$



$$\bar{y}^k$$

$$\frac{\iint_{\Omega} y^k B(x, y) dx dy}{\iint_{\Omega} B(x, y) dx dy} \dots (3)$$

where $k=1, 2, \dots, K$. When $k=1$, \bar{y} is a Centroid along y -axis.

In Addition, Extend Eqn. (2) & (3) to the following.

Sol:

$$\sum_{y=0}^{m-1} \sum_{x=0}^{N-1} B(x, y) = \sum_{y=0}^{m-1} \left(\sum_{x=0}^{N-1} B(x, y) \right) m_{pq}(x, y) = \frac{\iint_{\Omega} (x-\bar{x})^p (y-\bar{y})^q B(x, y) dx dy}{\iint_{\Omega} B(x, y) dx dy} \dots (4)$$

$$= \sum_{y=0}^{m-1} [B(0, y) + B(1, y) + \dots + B(N-1, y)] \Big|_{N=4}$$

$$= \sum_{y=0}^{m-1} [B(0, y) + B(1, y) + \dots + B(3, y)]$$

$$= B(0, 0) + B(1, 0) + B(2, 0) + B(3, 0)$$

$$+ B(0, 1) + B(1, 1) + B(2, 1) + B(3, 1)$$

+ ...

$$+ B(0, 3) + B(1, 3) + B(2, 3) + B(3, 3)$$

$$= 5 \times 255 \quad (\rightarrow \text{Normalize it!})$$

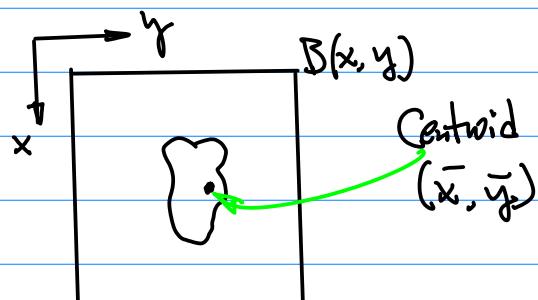
Hence, the size is 5.

Example: Computation of moments

Definition 2.

$$\bar{x}^k = \frac{\iint_{\Omega} x^k B(x, y) dx dy}{\iint_{\Omega} B(x, y) dx dy} \dots (z)$$

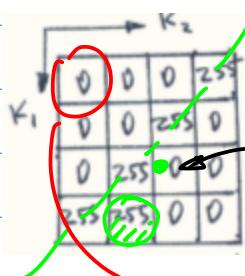
where $k=1, 2, \dots, K$; when $k=1$, \bar{x} "Centroid" along x -axis



where $T, q = 0, 1, 2, \dots, K$
 "Hu-moments" → Very Good Way to Describe Object Shape

Example: Find Centroid (\bar{x}, \bar{y}) of A given object

Since we now start from $(1,1)$,
we have



Visual Inspection
for (\bar{x}, \bar{y}) .

Sept. 22 (Thu).

Example: Pre-Processing Technique.

\bar{x}, \bar{y} Centroid.

From the previous Note.

$$\bar{x}^k = \frac{\iint_{S_k} x^k B(x, y) dxdy}{\iint_{S_k} B(x, y) dxdy} \dots (z)$$

$$B(1,4) + 2 \cdot B(2,4) + 3 \cdot B(3,4) + 4 \cdot B(4,4)$$

$$= 14$$

Therefore

For Visualizing
the Result

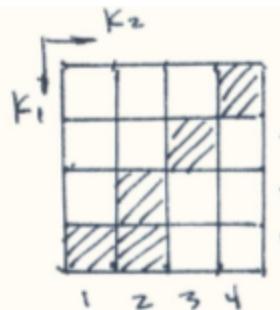
$$\bar{x} = 14 / A = 14 / 5 = 2.8 \approx 3$$

Rewrite Eqn (z) as

$$\bar{x} = \frac{\sum_{y=0}^3 \sum_{x=0}^3 x \cdot B(x, y)}{\sum_{y=0}^3 \sum_{x=0}^3 B(x, y)}$$

\downarrow

$\text{AREA} = 5$



Keep this for
Continued Computation
in your program.

Now,

$$\begin{aligned} & \sum_{y=0}^3 \sum_{x=0}^3 x \cdot B(x, y) \\ &= \sum_{y=0}^3 [0 \cdot B(0, y) + 1 \cdot B(1, y) + \\ & \quad 2 \cdot B(2, y) + 3 \cdot B(3, y)] \end{aligned}$$

Note: for Smaller Images, like the one we have
for the Simplicity of our discussion we
often to make the index x and y
starts from $(1,1)$ Not $(0,0)$.

Similarly,

$$\bar{y} = \frac{\sum_{y=0}^3 \sum_{x=0}^3 y B(x, y)}{\sum_{y=0}^3 \sum_{x=0}^3 B(x, y)}$$

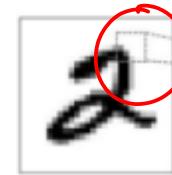
See the handout on the Class
github, we have

$$\bar{y} = \frac{12}{5} = 2.4 \approx 2$$

For Visualization

Purpose Only.

Example: Consider 2D Convolution
of a given Image.



INPUT
(28 x 28 x 1)

$m \times n$, $m = n = 28$

A single channel.
"A grayscale image"

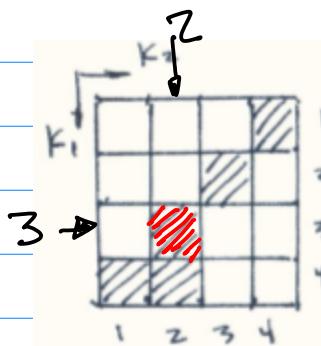


Fig. 1.

20	20	0	0
20	20	0	0
10	10	0	0
10	10	0	0

$I(k, h)$ $(m \times n)$ ($m = n = 4$)

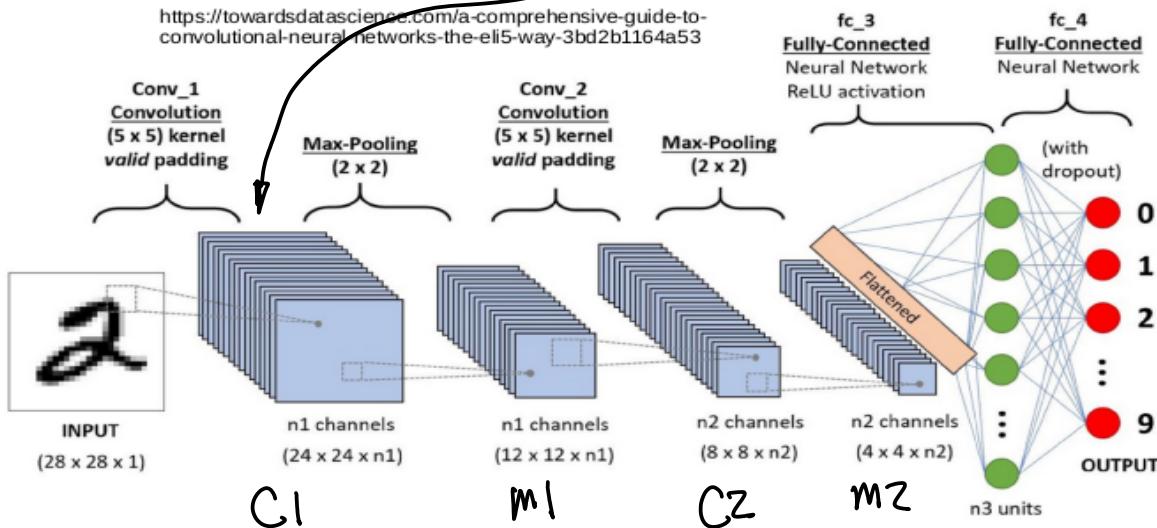
1	0	-1
1	0	-1
1	0	-1

A Kernel
 $K \times K$

$g(x, y)$

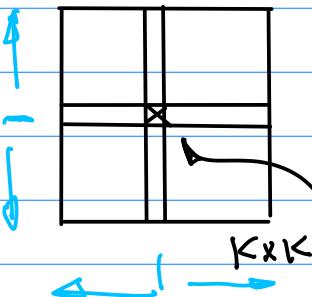
Now, consider Convolution Technique.

Illustration of A CNN for Digits Recognition



About the Kernel:

$K = \text{"Odd" Number.}$



Pixel of Interest

Fig. 3

$$\sum_{v=0}^{N-1} \sum_{u=0}^{M-1} I(u,v) g(x-u, y-v)$$

... (z)

a. Summation Index U, V ,

for $u=0, 1, 2, \dots, M-1$,
 $v=0, 1, \dots, N-1$.

Resolution of The Image $I(x,y)$
 $i \leq M \times N$.

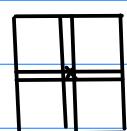
b. Kernel $g(x,y)$ is K -by- K ($K \times K$)
 in Size, K is always an "odd"
 Number.

The center pixel of the Kernel \rightarrow
 pixel of Interest;

"Odd" Number Makes No. of col. (Right
 & Left) equal, No. of Rows (Above
 & Beneath) equal \rightarrow Counting the
 Rows, Col. in a

"Un-Biased" fashion.

c. $g(x,y) \stackrel{(1)}{\rightarrow} Egn(z) \stackrel{(2)}{\rightarrow} ③$
 Originally $g(u,v) \rightarrow g(-u,-v)$
 Outside $Egn(z)$ $\downarrow ④$

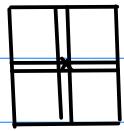


for x, y

$g(c_1-u, c_2-v)$

$\downarrow ⑤$

$g(x-u, y-v)$



for u, v

Fig. 1

Sept 27 (Tue).

Note: Check Homework on CANVAS.

Due A week from Today.

Oct. 4th.

Midterm will be given After the
 Hand written Student ID Recognition
 Project.

Example: Continued from the Z0
 Convolution Discussion.

openCV / deep-learning-2022s / 2022S-111-#2019S-23-2DConvolution-2019-2-4.pdf

Suppose we have an image $I(x,y)$,

And a Kernel $g(x,y)$ ($K \times K$).

Then, ZD Convolution is defined
 as

$$I(x,y) * g(x,y)$$

$$= \iint_{\mathbb{R}^2} I(u,v) g(x-u, y-v) du dv$$

... (1)

For Discrete Cases, we have

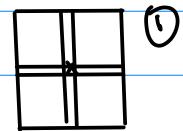
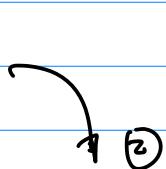
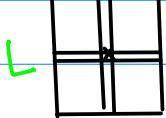
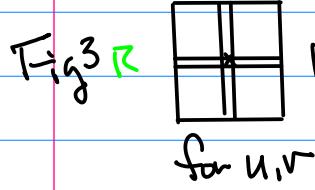
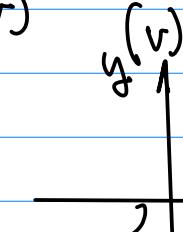
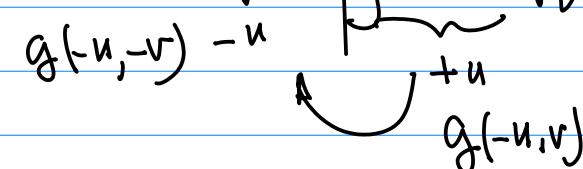
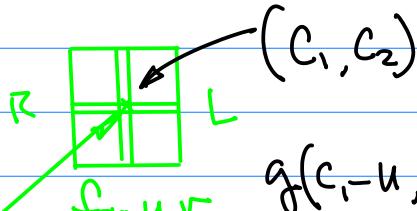
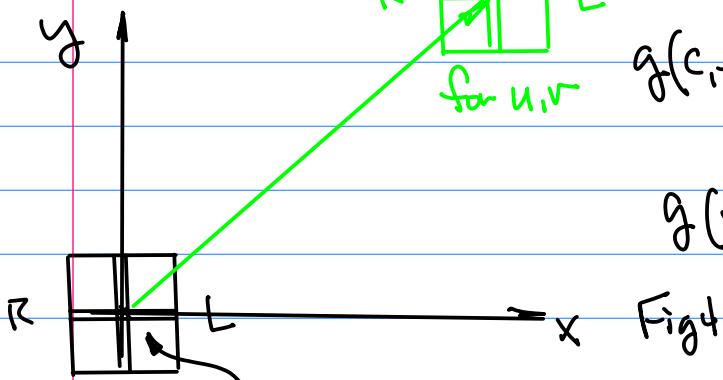
for x, y 

Fig. 2

for u, v for u, v $g(-u, -v)$  $g(-u, v)$

④

for u, v $g(c_1-u, c_2-v)$ Let $c_1 = x, c_2 = y$. $g(x-u, y-v)$ for u, v Without "Shifting" e.g.
 $g(-u, -v)$ or

$$g(c_1-u, c_2-v) \Big|_{\substack{c_1=0 \\ c_2=0}}$$

Fig 4

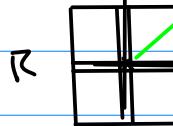
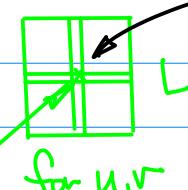
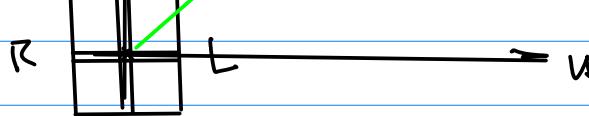
 v for u, v

Fig. 5

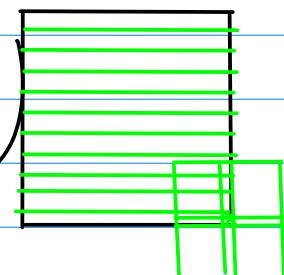
 (x, y) for u, v

Note: To perform 2D Convolution defined in Eqn(z) with the kernel $g(x, y)$.

"Slide" the Kernel at $(0, 0)$ of place $I(x, y)$

Top Left Corner.

Fig.8

 $(0, 0)$ 

Note: The Computation of 2D Convolution is defined as

Sum of Products. See Eqn(z)

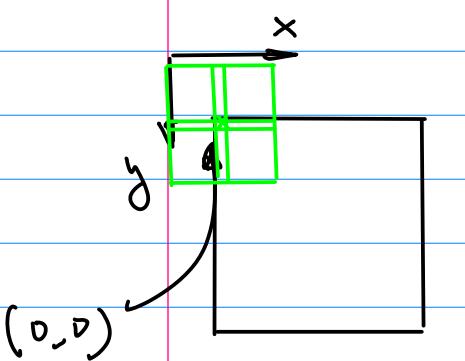
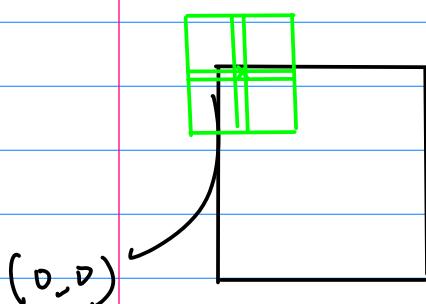


Fig.9

$g(x, y)$
Kernel of
 $K \times K$

Initial position of 2D Convolution.



Slide the Kernel
One pixel at Time
from Left to Right

Now, put All these Together.

Compute the convolution By Hand.

20	20	0	0
20	20	0	0
10	10	0	0
10	10	0	0

1	0	-1
1	0	-1
1	0	-1

A Kernel
 $K \times K$

 $I(x, y)$ $m \times n$

Continue the
Process, L to R,
Top to Bottom.

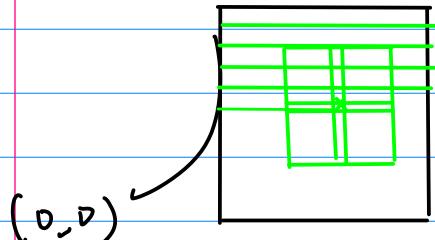


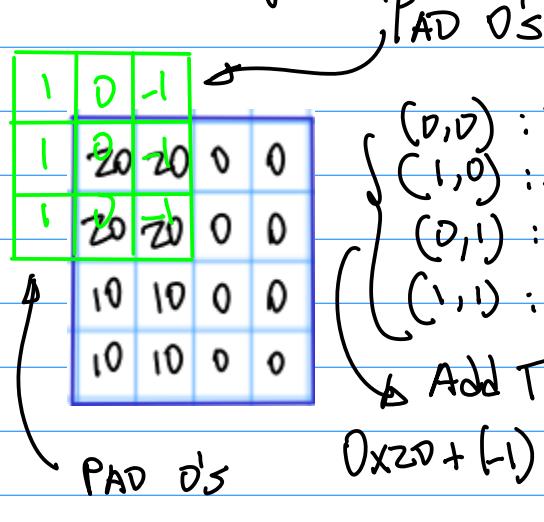
Fig.7

Fig.9

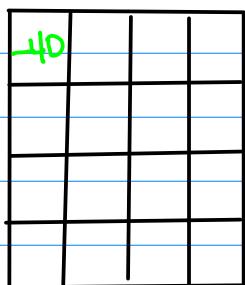
Step 1. Place $g(x,y)$ at $(0,0)$ on

Kernel

the Image.

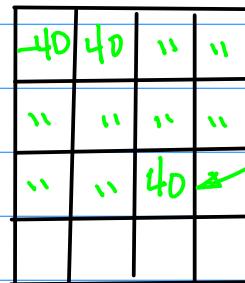


Convolution In Progress

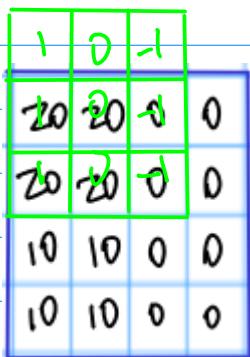


$$\begin{aligned} & 1 \times 20 + 0 \times 0 + (-1) \times 0 \\ & + 1 \times 20 + 0 \times 0 + (-1) \times 0 \\ & + 1 \times 20 + 0 \times 0 + (-1) \times 0 = 40 \end{aligned}$$

Convolution In Progress

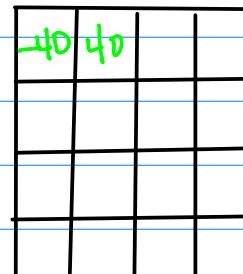


Step 2.



$$\begin{aligned} & 1 \times 0 + 0 \times 0 + (-1) \times 0 \\ & + 1 \times 20 + 0 \times 20 + (-1) \times 0 \\ & + 1 \times 20 + 0 \times 20 + (-1) \times 0 = 20 + 20 = 40 \end{aligned}$$

Convolution In Progress



Sept. 28.

Note: 1° Brief Update from

Each team, Status Activities.

On Thursday (A week from today)

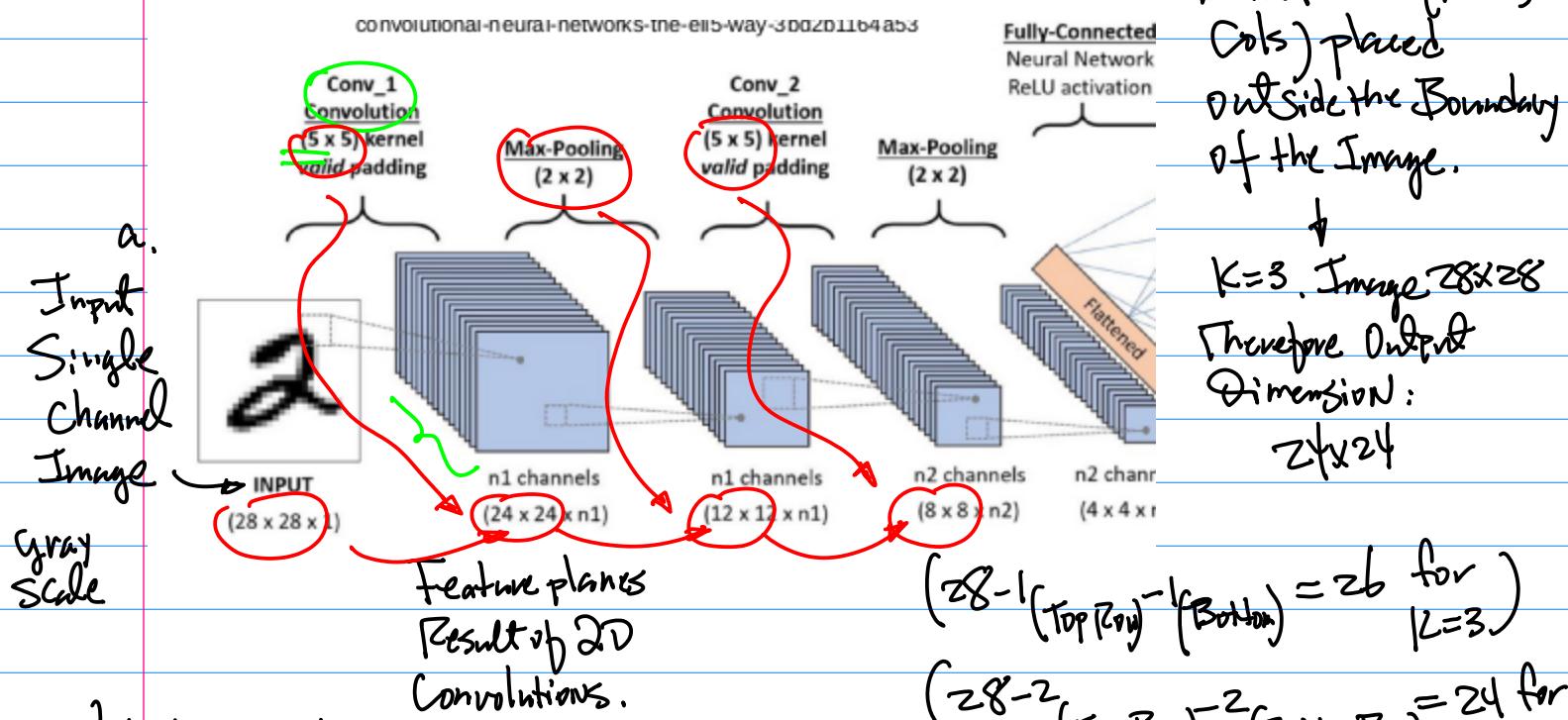
2° First Project On DCNN (MNIST)

for Handwritten Student ID

recognition System Design is

due on Oct. 10th.

Example:



Note 1. Resolution Change.

Input: 28×28 Output (Convolution) 24×24

$$(28 - 1)(\text{Top Row})^{-1}(\text{Bottom Row}) = 26 \text{ for } k=3$$

$$(28 - 2)(\text{Top Row})^{-2}(\text{Bottom Row}) = 24 \text{ for } k=5$$

$$\text{Image } M \times N, \text{ Kernel Size } k$$

$$M - \left(\frac{k-1}{2}\right) \times 2 = M - (k-1) \text{ Col.}$$

$$N - \left(\frac{k-1}{2}\right) \times 2 = N - (k-1) \text{ Row.}$$

Output ... (1)

$$[M - (k-1)] \times [N - (k-1)]$$

Kernel ($k \times k$) is placed inside
Image Plane. See

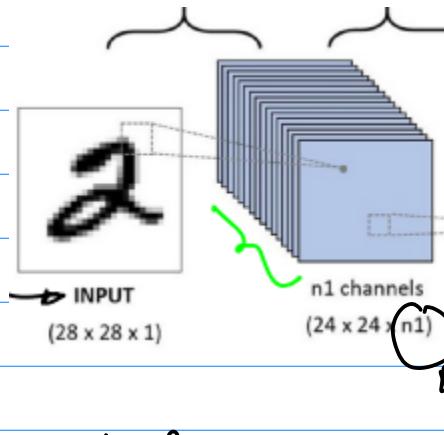
Fig. 1

+ Reduction of the
Rows (Top &
Bottom), Col.
(Left & Right)

Fig. 1.

in such a manner
None of them (Rows,
Cols) placed
Outside the Boundary
of the Image.+
 $K=3$. Image 28×28
Therefore Output
Dimension:
 24×24

Example:



Note 2. No. of Output layers is equal to No. of 2D Convolution.

Note 3. Max Pooling

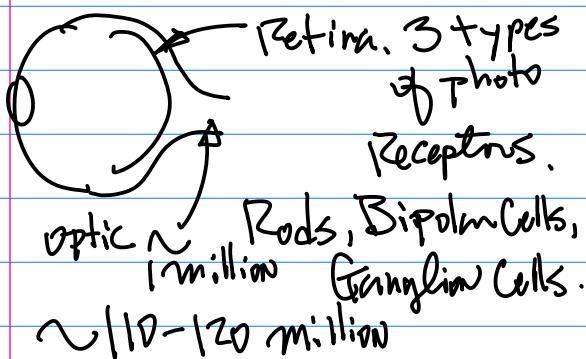
125	220
10	0

pixels

Max Pooling = 220

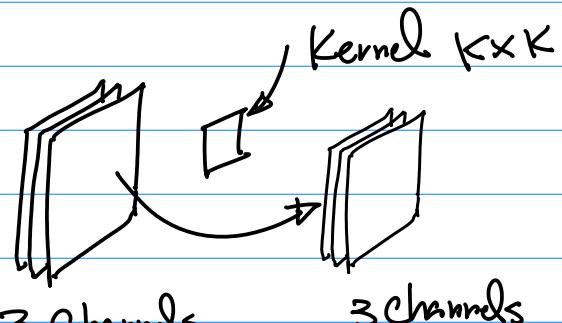
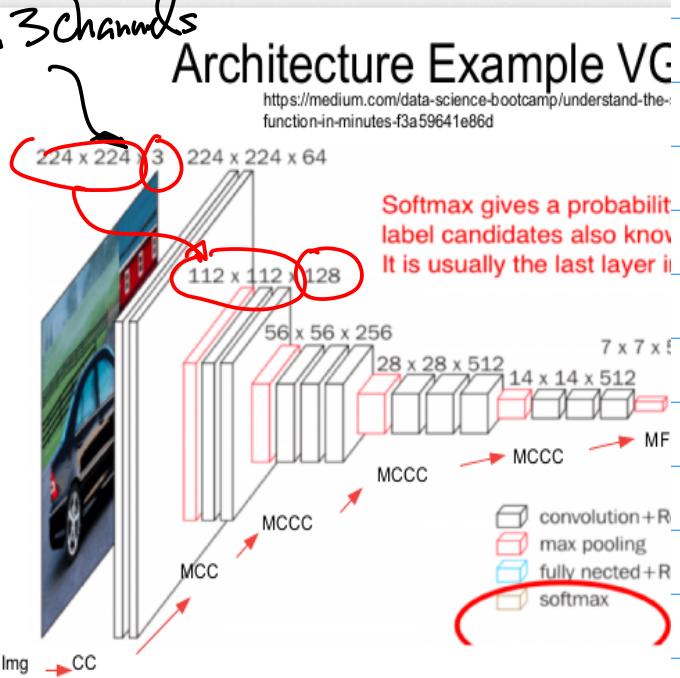
$\text{Max} \{ 125, 220, 10, 0 \}$

Biological Inspiration from Human Visual Perception System, Early Vision.



Change 2D Convolution "Step size" Stride to Penalize Both Function Convolution & Pooling.

a. 3 channels



3 channels with 1 kernel Convolution.

Optional (Not Required.)

Example:



Feature Extraction
"Edge" Map.

Boundary of the object.
Build A Kernel for the purpose of Extraction of Boundaries.

Object Boundaries \rightarrow Sudden Change of Image Intensity.

2022S-109-contour-intro-2022-2-28.pdf

Partial Derivative \hookrightarrow How to Detect Abrupt Change of the Intensity,

$$\frac{\partial}{\partial x} I(x, y) = \lim_{\Delta x \rightarrow 0} \frac{I(x + \Delta x, y) - I(x, y)}{\Delta x} \dots (z)$$

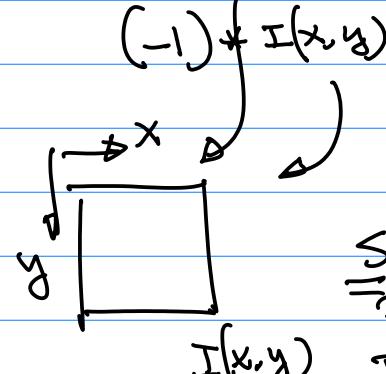
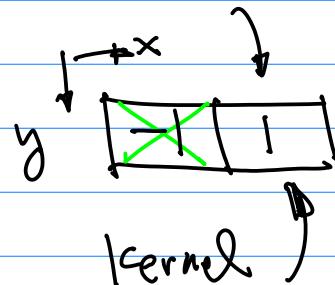
Similarly for y .

$$\frac{\partial}{\partial y} I(x, y) = \lim_{\Delta y \rightarrow 0} \frac{I(x, y + \Delta y) - I(x, y)}{\Delta y} \dots (z)$$

From Eqn(z), on Digital Image $\Delta x = 1$

$$\frac{\partial}{\partial x} I(x, y) = \lim_{\Delta x \rightarrow 0} \frac{I(x + \Delta x, y) - I(x, y)}{\Delta x}$$

$$\approx I(x+1, y) - I(x, y) = 1 * I(x+1, y) + (-1) * I(x, y)$$

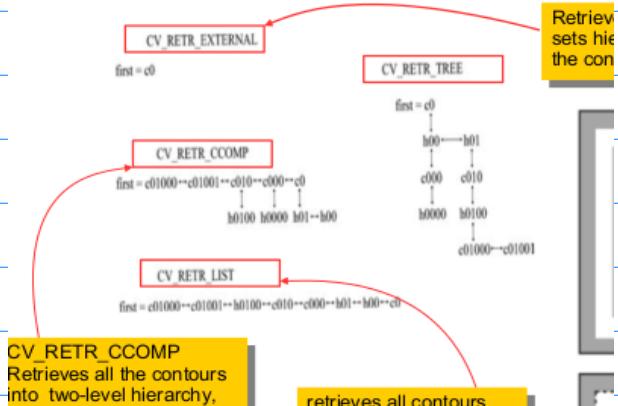


$\text{LoG}(x, y)$ Kernel, Laplace of Gaussian Kernel.

Now, consider Contour Analysis as Pre-processing Technique

Contours are Boundaries of the Objects, In OpenCV. They are Lists.

4 Contours N



Oct. 4th (Tue).

Handwritten Digits Recognition.

MNIST. Ref. from the Class

github

2022S-110-#lec5-1-mnist-hl-2021-3-1.pdf

2022S-110a-MNIST-digits-saveTrained.py

Background:

Step1. OpenCV \rightarrow Step2

= Python Program.

To Capture Web CAM

One Blank Printer

Paper + Marker

Write 4 Digits of yours

SID

Preprocessing Task.

Goal: To Build

Bounding Boxes

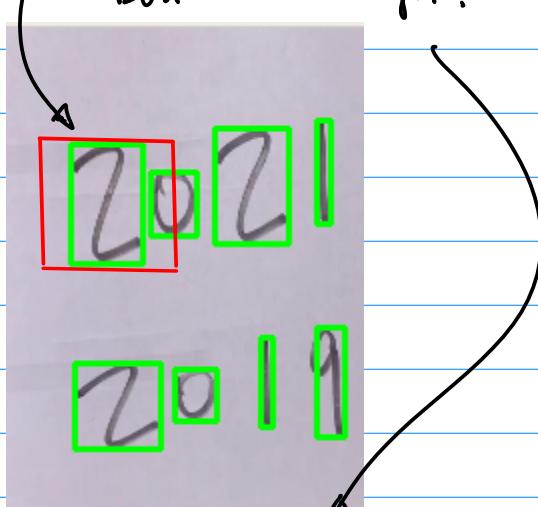
for the Digits.

(One Box for One

digit)

Step3. (Also Belongs to Preprocessing) Make Each Bounding Box to Become

- ① A Squared Small Image by the Bounding Box for Each Digit.



Without Distortion (e.g. without Changing the Aspect Ratio of the Image)

- ② To prepare the Squared Image to the right size, and Right format for MNIST Input.

1 Channel
 $\underbrace{28 \times 28}_{28 \times 28 \times 1}$

Step4. To Run the MNIST Sample Code. (Sample Code from google TIF Tutorial, or from the Class Recommended, or the

the Sample code from this class's github).

- ① The Sample code will get the Dataset for MNIST Convolutional Neural Network
- ② Read the Dataset → Collection of Small Image Annotation file
- ③ Will then Split the Dataset into Training set & Testing set. as well as its architecture

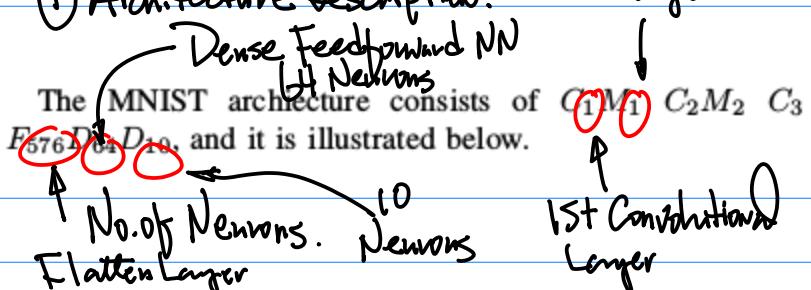
Step5. Save the trained DCNN. e.g. all Weights & Biases Once trained, are Saved into A file for the future use.

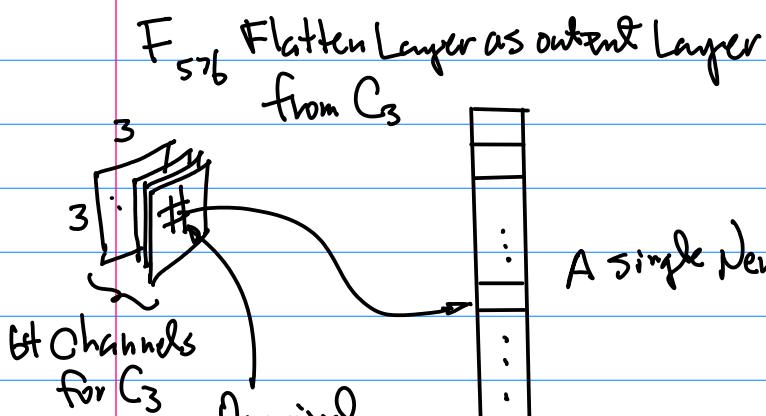
Step6. For Deployment purpose, Modify MNIST. (See Sample code), No Need to Download Dataset for Training, just Load Saved trained Result Back to the Defined Architecture.

Step7. Connect OpenCV Preprocessed Images (from Live Video, Web Cam) to the MNIST Code.

Example: On MNIST.

- ① Architecture Description.





Video Source from A Web CAM.

the 3rd part of the Homework:

Hand Calculation of (1) Binary Images. 1/2

A single Neuron. (2) 2D convolutions.

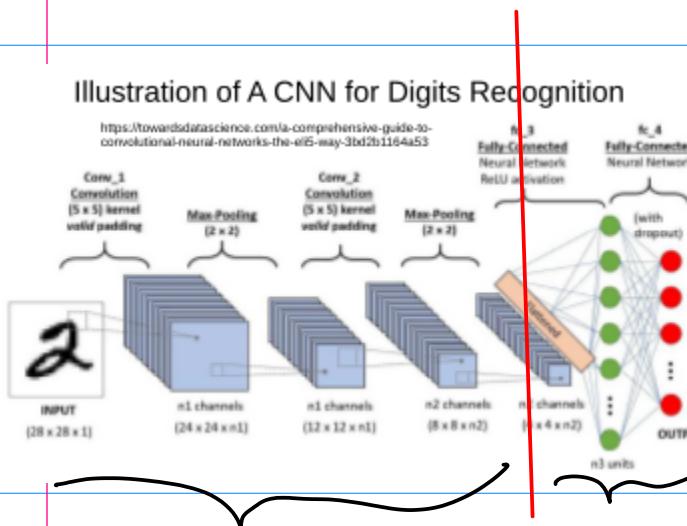
Team Project Quick Status checking,
Will continue on Thursday.

Fig.1

Oct.6 (Thu)

Note: 1° Project Announcement for
S19 Handwritten Digits Recognition
System, due Oct. 2nd, official
Announcement will be posted on CANVAS
Before End of the Day Friday;
2° Midterm after the project.

Example: DN MNIST.



Feature Extraction

Decision-Making

Note: N₃. of Neurons @ the Flatten Layer = 576



No. of the pixels @ C_3 Layer : $3 \times 3 \times 64$

Resolution

Channels / Layer
= 64

Homework Due A week

from Today, Oct. 11th.
(Tuesday)

2 parts. 1st Part: OpenCV Programming
Write A program to Capture Display Live

a.

Algorithm 1 Build MNIST convolutional NN

Require: import keras
 from keras import layers
 from keras import model

Ensure: Build MNIST convolutional NN layers sequentially

```
model = models.Sequential()
model.add(layers.Conv2D(32,(3,3), activation='relu', input_shape=(28,28,1)))
model.add(layers.MaxPooling2D((2,2)))
model.add(layers.Conv2D(64,(3,3),activation='relu'))
model.add(layers.MaxPooling2D((2,2)))
model.add(layers.Conv2D(64,(3,3),activation='relu'))
model.add(layers.Flatten())
model.add(layers.Dense(64,activation='relu'))
model.add(layers.Dense(10,activation='softmax'))
```

No. of the Neurons

Note: Eqn(1) "ReLU" is one of the very popular Activation Functions.

(4) Activation Function Softmax.

$$f(z_i) = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}} \quad \dots (2)$$

Where $\sum_{j=1}^K$ No. of Classes/Cats
 in the context of Handwritten Digits Recognition: $K=10$
 $(0, 1, 2, \dots, 9)$

$$e^{z_j}, \text{ for } j=1, 2, \dots, 10$$

we have

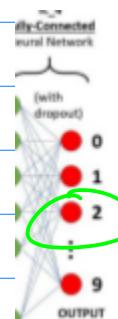
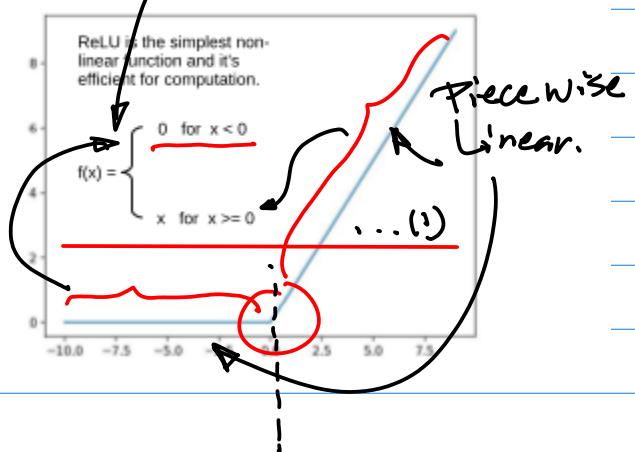
$$\sum_{j=1}^{10} e^{z_j} = e^{z_1} + e^{z_2} + \dots + e^{z_i} + \dots + e^{z_{10}}$$

Note, z_i is the output from the Neuron at Output layer, for $i=3$

① `models.Sequential`
 ② `model.add()`

`layers.Conv2D`
`layers.MaxPooling2D`
`layers.Flatten`
`layers.Dense`

③ Activation function.
 "ReLU", Math. Formulation

Rectified Linear Activation Function

Want to produce Output whose value can be formulated/integrated as probability value.

$$0 \leq \text{Prob}(x) \leq 1 \quad \dots (3a)$$

$$\int \text{Prob}(x) dx = 1 \quad \dots (3b)$$

$$\text{Now, } e^{z_1} + \dots + e^{z_{10}}$$

$$\frac{e^{z_i}}{e^{z_1} + \dots + e^{z_{10}}}$$

$$0 \leq \frac{e^{z_i}}{e^{z_1} + \dots + e^{z_{10}}} \leq 1$$

(match up to
Eqn 3a)

$$\frac{e^{z_1}}{e^{z_1} + \dots + e^{z_{10}}} + \frac{e^{z_2}}{e^{z_1} + \dots + e^{z_{10}}} + \dots +$$

$$\frac{e^{z_1}}{e^{z_1} + \dots + e^{z_{10}}} + \dots + \frac{e^{z_{10}}}{e^{z_1} + \dots + e^{z_{10}}}$$

$$= \frac{e^{z_1}}{e^{z_1} + \dots + e^{z_{10}}} = 1, \text{ Satisfies Eqn 3b}$$

Note: TensorFlow Keras Based
MNIST Sample Code.

- 2022F-105a-#6mnist-numerals-ch02 (copy...)
- 2022F-105b-#6mnist-numerals-ch02.py
- 2022F-105c-#load-deployment-MNIST.py

(5) Model.Save()

105a: \rightarrow 105b Saves the Trained Result (e.g. Both Architectural Information,
Baseline Code all weights/Biases)
For Training

```
-----save trained model-----
import h5py
model.save('harryTest.h5')
#-end
```

↓
105c.

Note: Save your Result in the following Deployment.

Naming Convention:

"FirstName_LastName-SID(4 Digits)-P1.h5"

① Will visit this topics later.

Oct. 11, 22

1. The optimizer is "rmsprop". RMSprop is a gradient based optimization technique used in training neural networks. This technique balances the step size, e.g., momentum, which makes the step size decrease for large gradients to avoid exploding, and increase the step for small gradients to avoid vanishing. So here is the background of it. "RMSprop is unpublished optimization algorithm designed for neural networks, first proposed by Geoff Hinton in lecture 6 of the online course Neural Networks for Machine Learning. RMSprop lies in the realm of adaptive learning rate methods, which have been growing in popularity in recent years, but also getting some criticism[6]. Its famous for not being published, yet being very well-known; most deep learning framework include the implementation of it out of the box" [Bushaev, 2018].

2. The loss function or the objective function is "categorical_crossentropy", which minimizes the loss function based on the probability of the likelihood of the predicted output class.

4 Different Types Contours Implementation in OpenCV.

② Will visit this later, Oct. 11, 22

2022S-109-contour-intro-2022-2-28.pdf

Motivation: Build Bounding Boxes → ROI (Region of Interest)

To Crop the Image
→
Further Processing

OpenCV Implementation for Contour Analysis.

4 Contours Models

Dr. EXTERNAL



CV_RETR_EXTERNAL

first = c0

CV_RETR_TREE

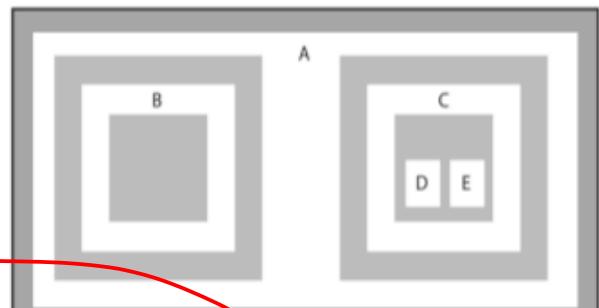
first = c0

b

TREE

h00—h01
h000—h010
h0000—h0100
c000—c0100
c0000—c01000

Retrieves only the external outer contours. It sets hierarchy[i][2]=hierarchy[i][3]=-1 for all the contours.



CV_RETR_CCOMP
Retrieves all the contours into two-level hierarchy, top-level for external boundaries and the 2nd level for the holes.

first = c01000--c01001--c0101--c000--c00
h0100--h0000--h01--h00

CV_RETR_LIST

first = c01000--c01001--h0100--c0101--c000--h01--h00--c0

retrieves all contours without any hierarchical relationships.

C. Notation

"C" — Contour
"h" — hole

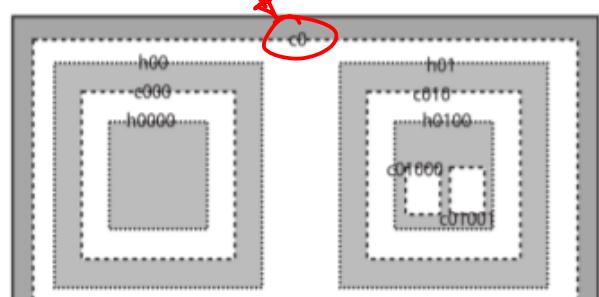


Fig. 1

Oct 11, 22.

Note: 1° Whitepaper on MNIST.

2022S-110-#lec5-1-mnist-hl-2021-3-1.pdf — lec5-1-mnist-hl-2021-3-1

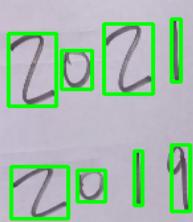
2° Midterm Exam Oct. 27

Thursday.

3° Two Topics in the white paper
to Be visited later.Example: Continuation on Contour
Analysis. PPT. (on github)

2022S-109-contour-intro-2022-2-28.pdf

Motivation: Using Preprocessing Techniques

We have covered in
the class.Find Bounding Boxes
for the Digits, see
Green Boxes in Fig.1.Fig.1. We will Need Contour Analysis
Technique.

From the Sample Code for Contour Analysis.

1. Line 24, Resize function Can Change the

```

23
24 img = cv2.imread(image, cv2.IMREAD_COLOR)
25 img = cv2.resize(img, (256, 256))
26 cv2.imshow('original image', img)

```

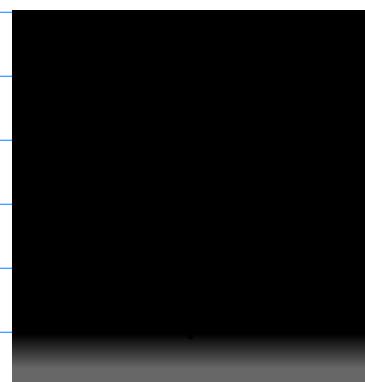


Fig.3a

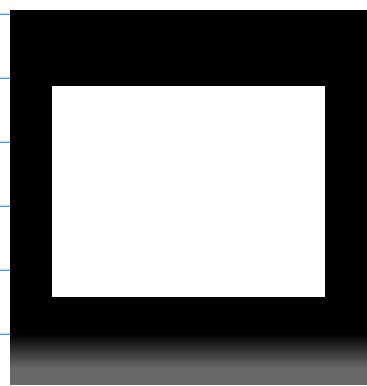
Add An
Object
(White)

Fig.3b

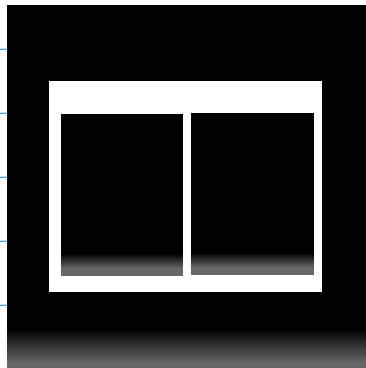
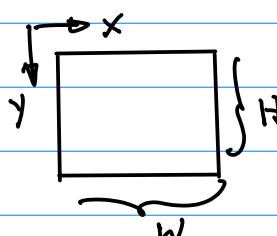
White Obj
has 2
Black "holes"

Fig.3c

Add 2
Foreground
White
Squares.

Fig.2



$$\text{Aspect Ratio} = \frac{W}{H} / W \neq H$$

... (!)

When $W=H=256$ in Line 24,
it changes the A.R.

Aspect Ratio of An Image. where

$$\text{Aspect Ratio} = \frac{W}{H}$$

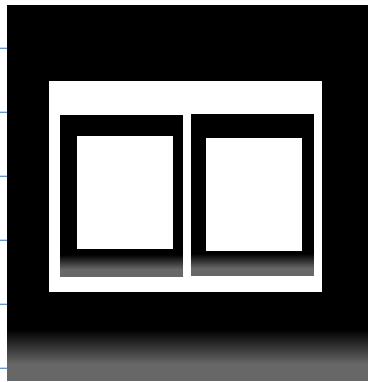


Fig. 3d

Add 2 Black
Squares As
Background

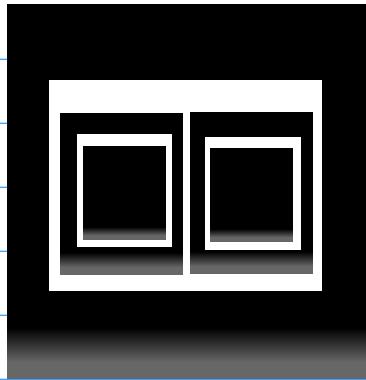
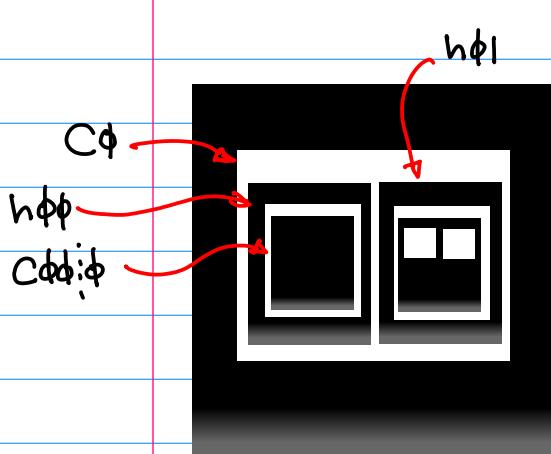


Fig. 3e



Add 2 Smaller
White Squares
to the Right Blank
Square

Fig. 3f

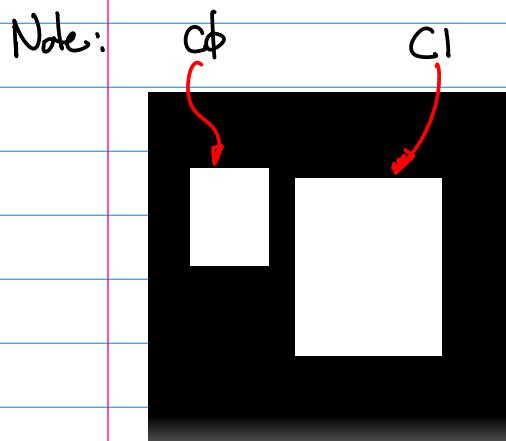
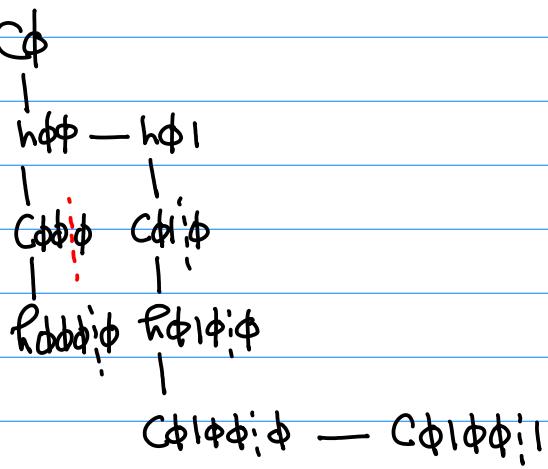


Fig. 3g



External Contour

```

30 thresh = cv2.Canny(imggray, 100, 200)
31 cv2.imshow('Canny', thresh)
32
33 contours, hierarchy = cv2.findContours(thresh, cv2.RETR_EXTERNAL, cv2.CHAIN_APPROX_SIMPLE)
34 # Draw contour
35 thresh1 = cv2.cvtColor(thresh, cv2.COLOR_GRAY2BGR)
36 cv2.drawContours(thresh1, contours, -1, (0, 255, 0), 5) # all contours
37 cv2.imshow('All contours', thresh1)
38 #Bounding box

```

From PPT, pp.8

Note: Use Size/Area of A Contour
to filter Out Random Noise

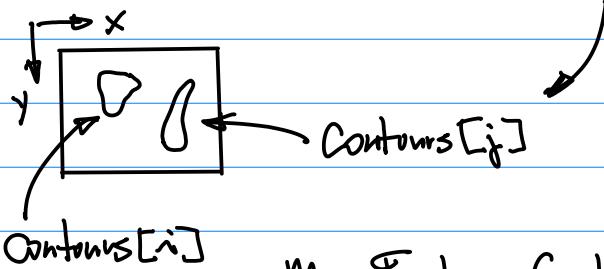
1. Moments

```

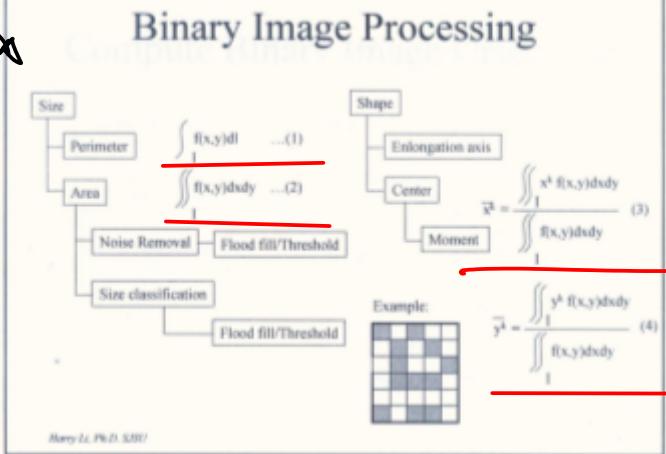
1 import cv2
2 import numpy as np
3
4 img = cv2.imread('star.jpg',0)
5 ret,thresh = cv2.threshold(img,127,255,0)
6 contours,hierarchy = cv2.findContours(thresh, 1, 2)
7
8 cnt = contours[0]
9 M = cv2.moments(cnt)
10 print M

```

To find moments of
Any Contours, i or j



Many Features Can be
derived from the moments.

**2. Contour Area**

```
area = cv2.contourArea(cnt)
```

Set Threshold T = 10 (pixels)

if $area[i] < T$

then $area[i]$ is Random Noise

Oct.13 (Th)

Example: About Resize the Image

And/or ROI Without Distortion
of Aspect Ratio.

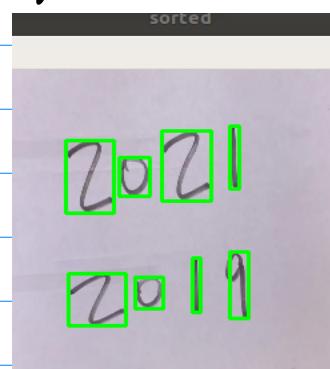


Fig.1a

Step1. Make the Bounding Box
Design ROI as A Square
Image without Distortion of A.R.

Use OpenCV find Contour function

→ Bounding Box → W, h (Width, Height)

$$\text{Max}\{W, h\} = h$$

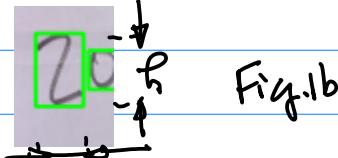


Fig. 1b

Use h Dimension "w" to Build Square Image.

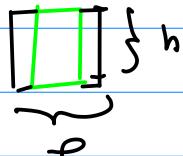


Fig. 1c

bitwiseAND

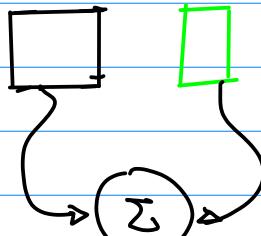


Fig. 1d

Ref. pp. 36.

Square Image.

[2022S-101-note-part2-cmpe258-2022-04-26.pdf](https://drive.google.com/file/d/1o1yfXzJLjyfCQHgkVqfDfBxGKUOOGIw/view?usp=sharing)

(For Yolo As Well)

Background:

Existing (Previously) Techniques.

① Sliding Windows
② Region Proposal

You Only Look Once:
Unified, Real-Time Object Detection Technique

Joseph Redmon*, Santosh Divvala*, Ross Girshick*, Ali Farhadi*†

University of Washington*, Allen Institute for AI†, Facebook AI Research†

<http://pjreddie.com/yolo/>

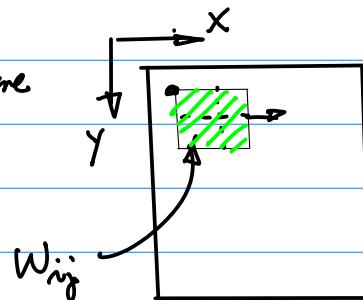
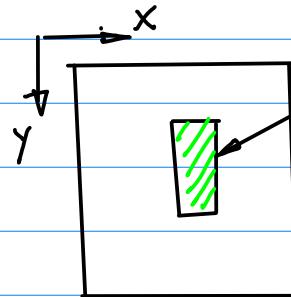


Fig. 2a

Region Proposal Technique Built for R-CNN.

Object Detection Technique is applied on the W_{ij} .
 $\phi(W_{ij})$

localized, And No "global Image $I(x, y)$ Information".



Bij, $\theta(Bij)$ object Detection & Recognition

Need/Lack of Global Image Information

Post Processing is Need.

Fig. 2b.

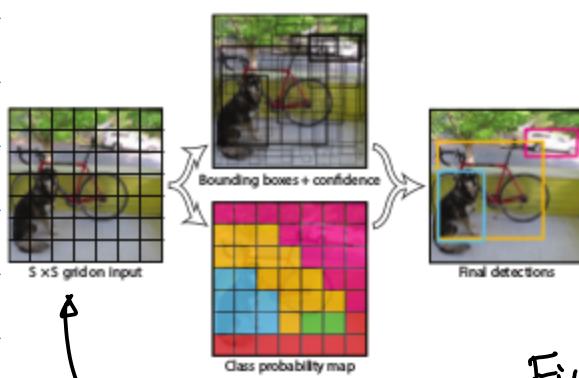
Now, Yolo for Object Detection & Recognition

Ref: 1. Yolo Paper

<https://github.com/opencv/opencv/blob/master/doc/deep-learning-2022s/2022S-112-yolo-paper.pdf>

2. Readme

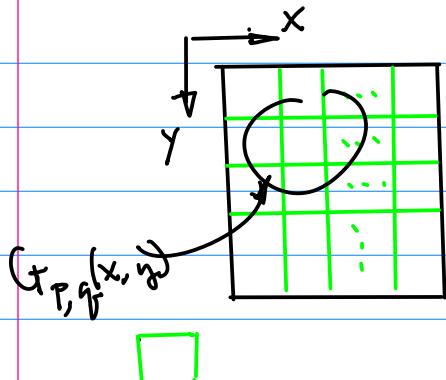
<https://github.com/opencv/opencv/blob/master/doc/deep-learning-2022s/2022S-113-README-yolo4-v2-yy-hl-2021-4-5.txt>



Note:



1. Image $I(x, y)$ is divided into $S \times S$ Grids. Denote it as $G_{p,q}(x, y), \dots (1)$ where $p, q = 0, 1, 2, \dots$ indicate the location of the grid. (See Ref. lecture Note, pp40)



2. Bounding Boxes on Each Grid.

$$B_{i,j}(x, y) \dots (2a)$$

$$\begin{matrix} \neq \\ i=x, j=y \end{matrix}$$

x, y coordinates of the top left corner of the Box

3. For Each Bounding Box, we define 5 parameters.

x, y (location), Top left corner)

w, h : width and height of the Box.

f : Confidence of the Box for its belonging to a certain Object

We can write

$$(x, y, w, h, f) \rightarrow B(x, y, w, h, f) \dots (2b)$$

Note: for Each Grid Cell,

1. It predicts Bounding Boxes and their Confidence Score
2. It computes Conditional probability

$$\text{Prob(Class}_i|\text{Object}) \dots (3)$$

The probability of a given object which belongs to Class i , for $i=1, 2, \dots, M$;

4. The Entire Image $I(x, y)$ is employed to generate Class Probability Distribution Map (Fig. 2c), to be used in Eqn Below

$$\text{Pr(Class}_i|\text{Object}) * \text{Pr(Object)} * \text{IOU}_{\text{pred}}^{\text{truth}} = \text{Pr(Class}_i) * \text{IOU}_{\text{pred}}^{\text{truth}} \quad (1)$$

5. IOU (Intersection of Union). \Rightarrow p40 Tref

Example: Illustration of IOU

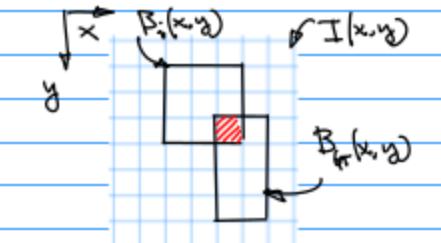


Fig.3

Fig.5

$$\text{IOU} = \frac{\text{Intersection}}{\text{Union}} \dots (7)$$

Where Bounding Box $B_i(x, y)$ is a prediction.

Bounding Box $B_{GT}(x, y)$ is the Ground Truth

Red Region: Prediction matches the ground truth; \rightarrow Intersection

$$B_i(x, y) \cap B_{GT}(x, y) \dots (4-a)$$

Entire Area of $B_i(x, y)$ and $B_{GT}(x, y)$

$$B_i(x, y) \cup B_{GT}(x, y) \dots (4-b)$$

Oct 18th (Tue)

(Continued) from the Discussion of Notation.

From Fig. 2c, pp. 3b.

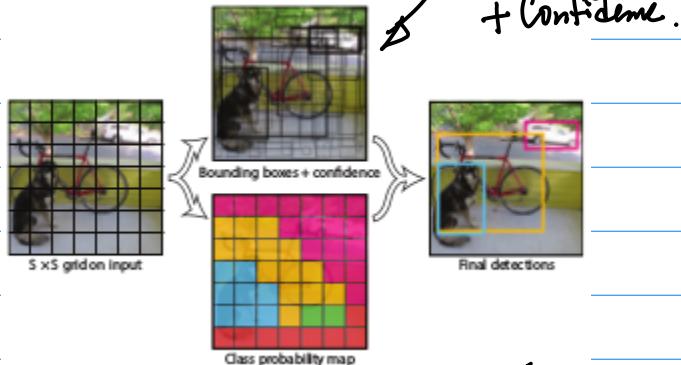
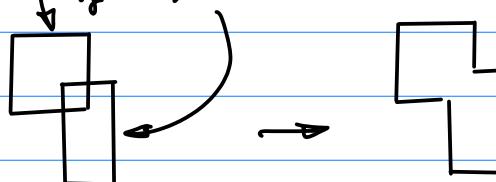
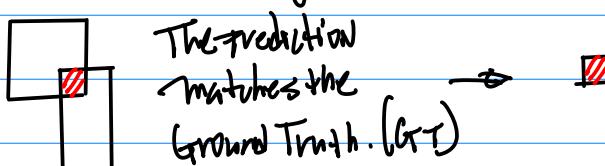


Fig. 1

The Union of $B_i(x, y)$, $B_{GT}(x, y)$



The Intersection of $B_i(x, y)$, $B_{GT}(x, y)$



Confidence of the Prediction

By Bounding Box $B_{ij}(x, y)$ is defined as a ratio:

$$C = \frac{\text{Intersection}}{\text{Union}}$$

(IoU)

Confidence, denoted as η

Sometimes,

Example: Given $B_{ij}(x, y)$ and $B_{GT}(x, y)$ in Fig 5, pp. 3b.

Calculate Confidence "C" (η)

Sol
First, Find the Union of $B_{ij}(x, y)$ and $B_{GT}(x, y)$.

$$\begin{aligned} B_{ij}(x, y) \cup B_{GT}(x, y) &= \\ N[B_{ij}(x, y)] + N[B_{GT}(x, y)] - \\ N[B_{ij}(x, y) \cap B_{GT}(x, y)] &= \\ 3 \times 3 + 4 \times 2 - 1 \times 1 &= 9 + 8 - 1 = 16 \end{aligned}$$

Note,

$$N[B_{ij}(x, y) \cap B_{GT}(x, y)] = 1$$

Hence

$$C = \frac{1}{16}$$

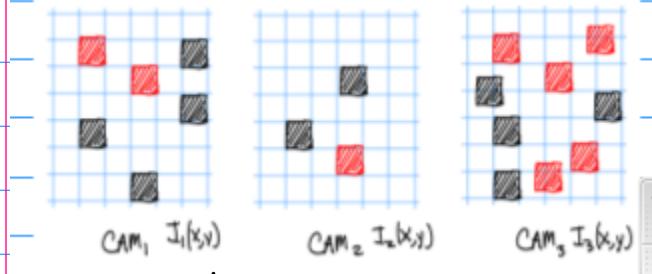
Note: $0 \leq C \leq 1$, similar to the probability.

The Consider Class probability map.
 (Same size as the Image $I(x,y)$)

2022S-101-note-part2-cmpe258-2022-05-3.pdf

Background on Bayesian Theorem.

Example, PP.39



Camera 1, $I_1(x,y)$, Camera 2, ...
 $I_2(x,y)$,

Red Squares : Pedestrian ; Black Squares :
 Vehicles

- Event of having Pedestrian Appeared in the Images

$$R = R_{I_1} + R_{I_2} + R_{I_3} \dots \text{... (1)}$$

- Probability of the event R .

$$\text{Prob}(R) = \text{Prob}(R_{I_1}) + \text{Prob}(R_{I_2}) + \text{Prob}(R_{I_3}) \dots \text{... (2)}$$

Where $I_1 \cap I_2 \cap I_3 = \emptyset$ (Empty set)

- Consider

$$\text{Prob}(R_{I_1}) = \text{Prob}(R/I_1) \text{Prob}(I_1)$$

Pedestrian(s) in $I_1(x,y)$... (3a)

$$\text{Prob}(R_{I_2}) = \text{Prob}(R/I_2) \text{Prob}(I_2) \dots \text{... (3b)}$$

$$\text{Prob}(R_{I_3}) = \text{Prob}(R/I_3) \text{Prob}(I_3) \dots \text{... (3c)}$$

$E_{\text{inf}}(z)$ and $E_{\text{fr}}(3a) - (3c)$:

$$\text{Prob}(R) = \sum_{i=1}^3 \text{Prob}(R/I_i) \text{Prob}(I_i) \dots \text{... (4)}$$

Now, consider the formulation Below:
 (See Fig.1, PP.37)

$$\underbrace{\text{Pr(Class}_i|\text{Object}) * \text{Pr(Object)}}_{\overline{\text{A}}} * \text{IOU}_{\text{pred}}^{\text{truth}} = \underbrace{\text{Pr(Class}_i)}_{\overline{\text{B}}} * \text{IOU}_{\text{pred}}^{\text{truth}} \quad \text{... (1)}$$

Bounding Box Confidence.

$$\text{Prob}(\text{Class}_i) = \text{Prob}(\text{Class}_i|\text{Object}) \text{Prob}(\text{Object}) \dots \text{... (5)}$$

Note: The Discussion and Mathematic Formulation, Equations, Formula are required.

Implementations: {
 Code (github)
 Architecture
 Loss function}

Homework: Due Oct 27th(Th).

Installation & Test Run Yolo4.

- Download Yolo4 (Readme from the github). Run the code.

2022F-106-YOLO4-README-Tiny-Yolo4-G...

2. Use your Smart phone to Record

~15sec. Video Clips for the testing of Yolo4. (Such as Traffic, Vehicles, Pedestrians),

3. Submission:

(1) Screen Capture of the Success execution of the code;

(Identifier, Such as folder name etc.).

(2) Video Clips saved After the processing

Oct. 20.

Note: Homework Using Yolo.

Updated Readme (github link was updated).

Example: Architecture of Yolo, Ref:

2022F-108a-Yolo-architecture-loss-function-2022-10-10.pdf

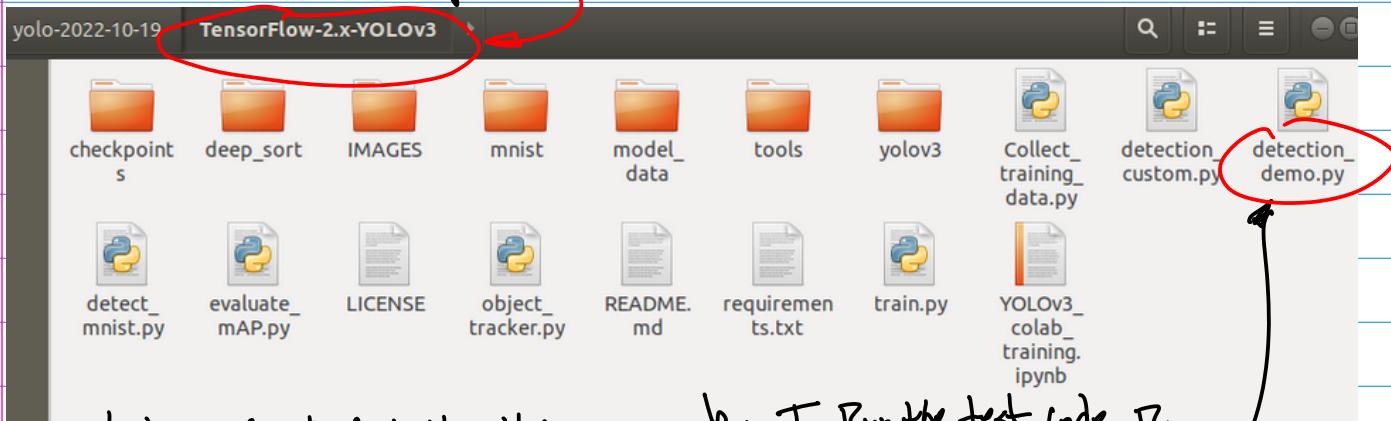
Architecture Discussion, Fig.1

Note: Midterm Exam Schedule

Nov. 3rd. Review Session will be provided;

2022F-106-README-Tiny-Yolo4-GPU-Ubu...

a. Down Load from the github



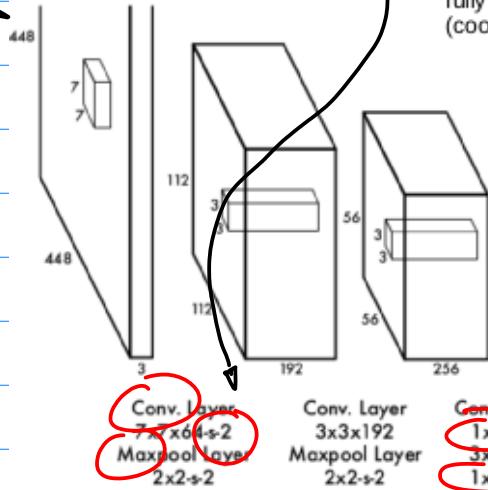
marked as v3, But follow the readme Document, modify the Configuration file (.py) to choose Yolo v4;

b. To Run the test code, Run Python Code

c. Create "yaml" file for Conda Environment Setup.

a. $448 \times 448 \times 3$

1. Input image size: $448 \times 448 \times 3$.
Resolution reduction for feature extraction/abstraction Pooling and convolution with stride = 2;



b. Convolution with

Stride = 2. "Scan" the input layer at every two pixels from Left to Right, then at every other row from Top to Bottom.

Base Line Yolo Architecture

Design guideline: The block of convolutional layers to extract image features, the fully connected layers to predict the output probabilities and the locations (coordinates).



Hand Calculation
Is Required.

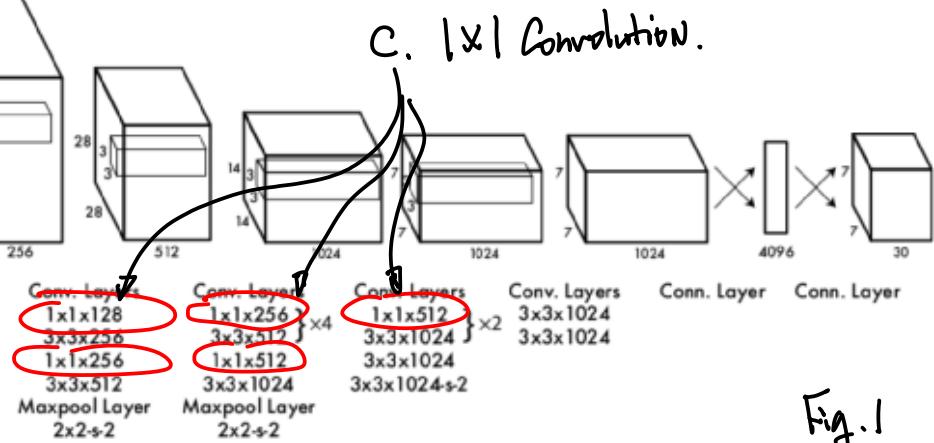


Fig. 1

The Motivation for the

2-Stride, or K-Stride, Convolution

is to combine Convolution with
SubSampling, e.g. Reduction of
the feature layer. Example Below,

See Stride pattern in Green.

Consider 3×3 Kernel Convolution

20	20	0	0
20	20	0	0
10	10	0	0
10	10	0	0

1	0	-1
1	0	-1
1	0	-1



$\frac{3 \times 3 \times 1}{\text{Channel}}$

Reduce 3×3 kernel to 1×1

Kernel below, then convolve with the
image $I(x,y) \rightarrow$ Output Resolution
 $M \times N$

25

is $M \times N$ 

20	20	0	0
20	20	0	0
10	10	0	0
10	10	0	0

Fig. 2

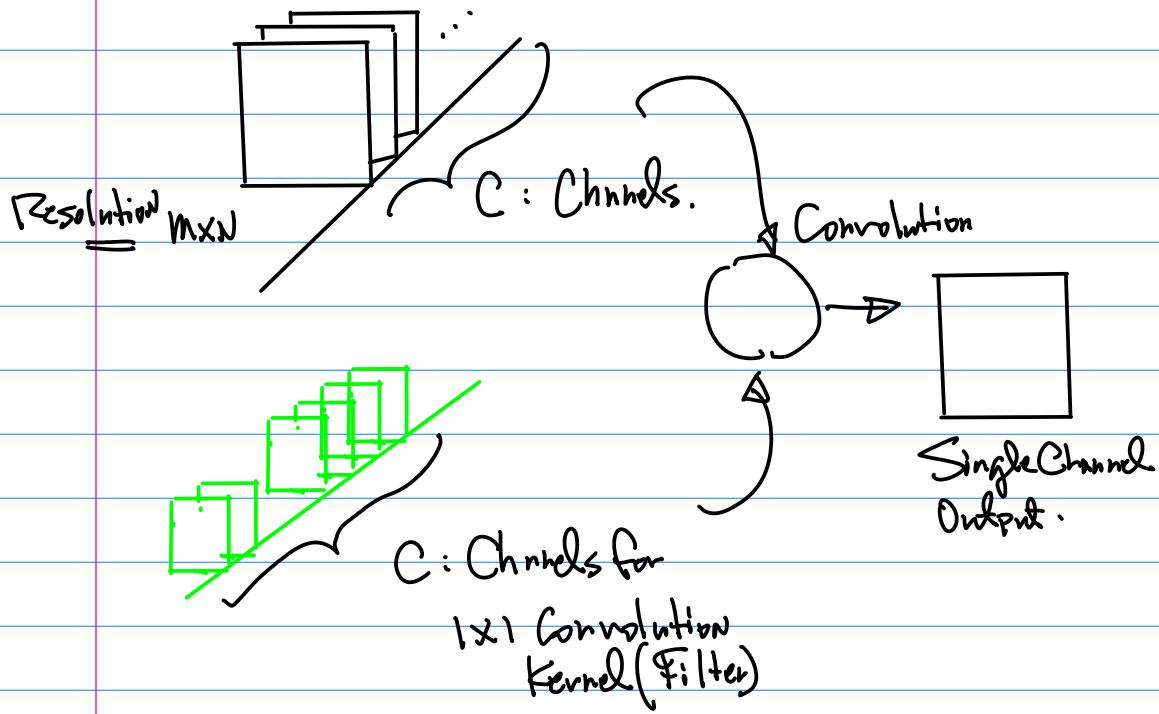
Image

x-y Two Dimensional Space
"Spatial"
V.S. Temporal (Time)

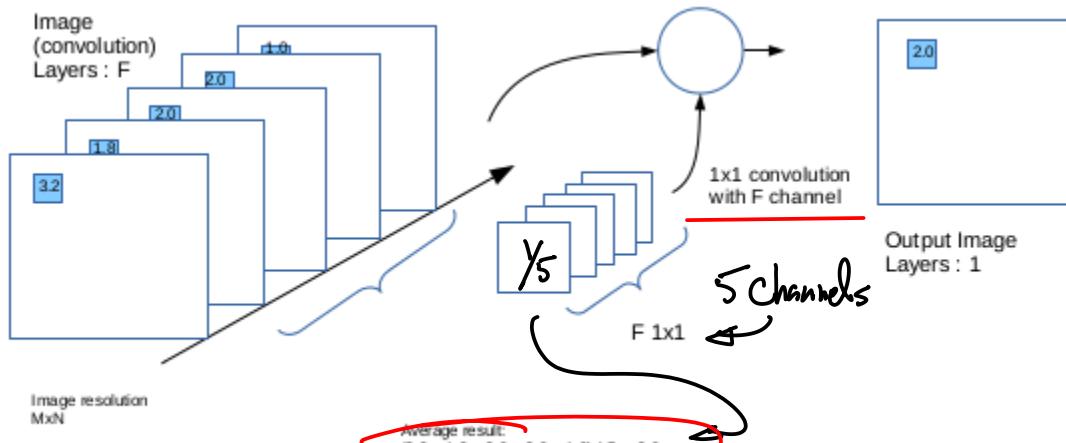
Space (Spatial) \rightarrow Space in
in 2D 3D and Beyond.

for Example for Image

Note: 1° 1×1 Convolution Reduces the
Number of Layers of the Input.
2° Pooling.



Feature
Layers.



I.i. Ph.D

Example: a) A loss function can be defined by subtracting the output (function) from the ground truth. \rightarrow Square it to prevent from possible cancellations when summed up together

Location Based Loss function. C)

Loss Function for YOLO

b) One Loss function
 $f_{\text{loss}} = \sum_{i \in I} f_{\text{loss},i} \dots (1)$

$$f_{\text{loss}} = \alpha_1 f_{\text{loss},1} + \alpha_2 f_{\text{loss},2} + \dots + \alpha_k f_{\text{loss},k}$$

Where $\alpha_1 + \alpha_2 + \dots + \alpha_k = 1$

Note: If we want emphasize Location Prediction, $\rightarrow f_{\text{loss},1} \rightarrow \alpha_1 \uparrow$ Bigger α_1

d) Geometric Shape (Always in Rectangle Shape, Size) Loss Function

$$\lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} \left[(\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2 \right]$$

Location (x_i, \hat{x}_i) of Bounding Box

$$\lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right]$$

$$+ \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} \left[(\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2 \right]$$

$$+ \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} (C_i - \hat{C}_i)^2$$

$$+ \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{noobj}} (C_i - \hat{C}_i)^2$$

$$+ \sum_{i=0}^{S^2} \mathbb{1}_i^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2$$

$\mathbb{1}_{ij}^{\text{obj}} = \begin{cases} 1 & \text{when } B_{ij} \text{ exists} \\ 0 & \text{o/w (otherwise)} \end{cases}$

C. From Eqn Below.

$$\Pr(\text{Class}_i | \text{Object}) * \Pr(\text{Object}) * \text{IOU}_{\text{pred}}^{\text{truth}} = \Pr(\text{Class}_i) \quad (1)$$



Fig 1.a

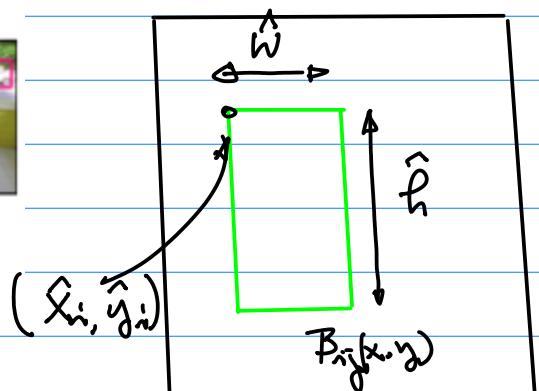


Fig 1.b

Note: Midterm Exam Scheduled

on the 3rd, Nov.

2022S-114c-KmeanCluster-v3-2022-4-19.pdf

$$+ \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} (C_i - \hat{C}_i)^2 \\ + \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{noobj}} (C_i - \hat{C}_i)^2$$

$$\arg \min_S \sum_{i=1}^k \sum_{x \in S_i} \|x - \mu_i\|^2 = \arg \min_S \sum_{i=1}^k |S_i| \text{Var } S_i$$

... (1)

Loss function Based on the Confidence of Bounding Box

Loss function of the Class probability.

Feature vector $\vec{x} = (x_1, x_2, \dots, x_N)$ with Dimension N. $\rightarrow \vec{x}$
for Example N=2.

a set of observations (x_1, x_2, \dots, x_n) ,

Ground Truth

$$+ \sum_{i=0}^{S^2} \mathbb{1}_i^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2$$

Estimated Probability

Prob(c)

Class, example, Traffic Signs, Pedestrian, Vehicles etc.

1st Sub: For Experiment 1.

so, for k-th Experiment,
we have $\vec{x}_k = (x_{k1}, x_{k2})$. etc.

$\mu_i = (\mu_{i1}, \mu_{i2}, \dots, \mu_{iN})$, mean of the given class i

A Vector, Same dimension as the feature Vectors

for N=2, $\mu_i = (\mu_{i1}, \mu_{i2})$

i index for Class i.

2022S-114c-Kmean-handCalculat...

2022S-114c-KmeanCluster-v3-20...

Background: A Tool to Create Probability Distribution Map. (In Fig 1a, Pg 42)

$$\arg \min_S \sum_{i=1}^k \sum_{x \in S_i} \|x - \mu_i\|^2 = \arg \min_S \sum_{i=1}^k |S_i| \text{Var } S_i$$

then,

$$\sum_{x \in S_i}$$

Comparison, for the purpose
of Classifying / Grouping a feature
Vector to Class i .

Mean (AVG)

 μ_i : feature Vector Distribution

Take all feature
vectors \mathbf{x}_j , as
long as they are

from the Class S_i ,
Class i

features: Edge Comp. from
Edge Detection; Colour
Distribution; Contours

Graphically

μ_i is a
Cluster

Then, To make sure the
minimization

Argmin

to cover all classes

$$\sum_{i=1}^K \text{Total K classes.}$$

$$\mathbf{x} - \mu_i$$

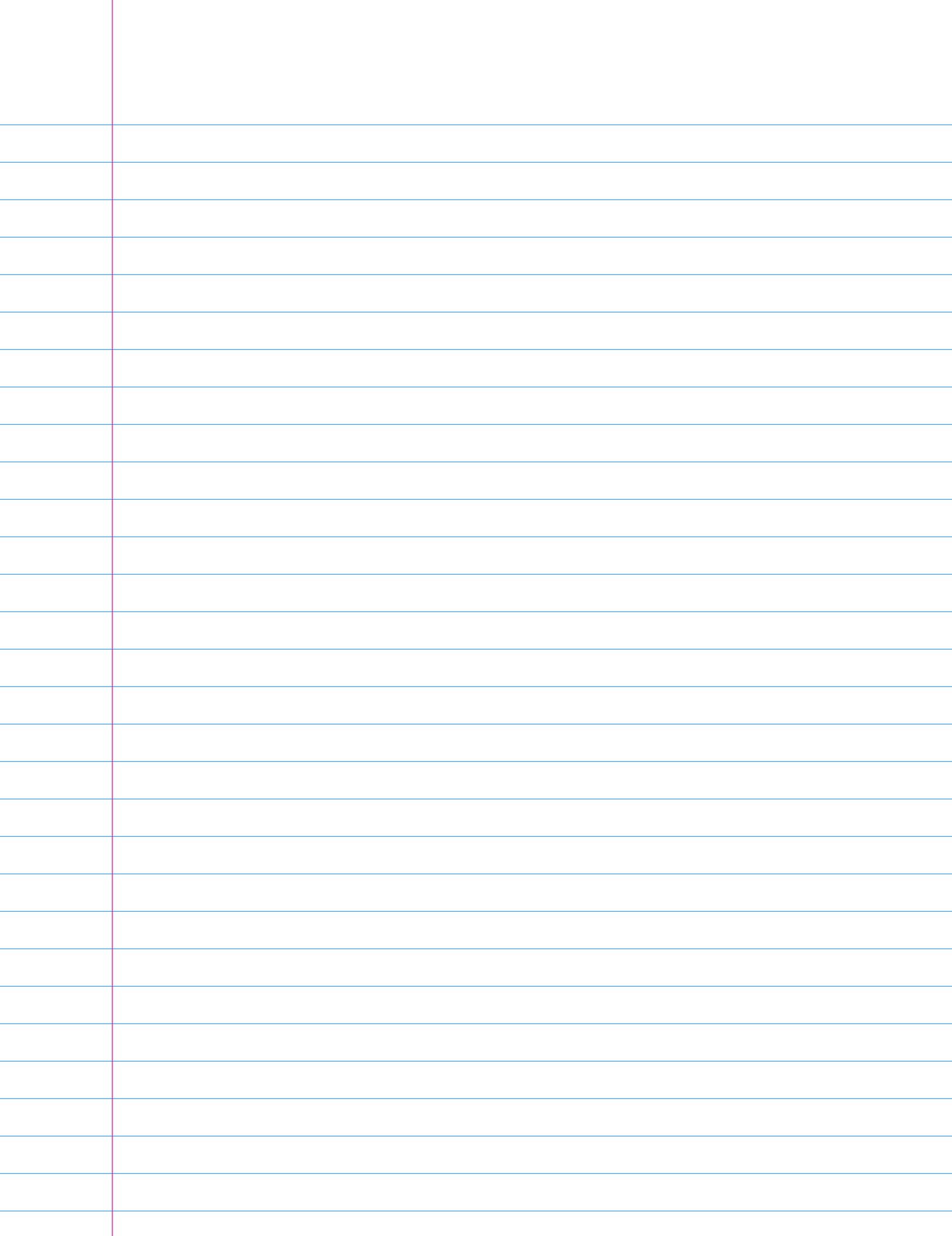


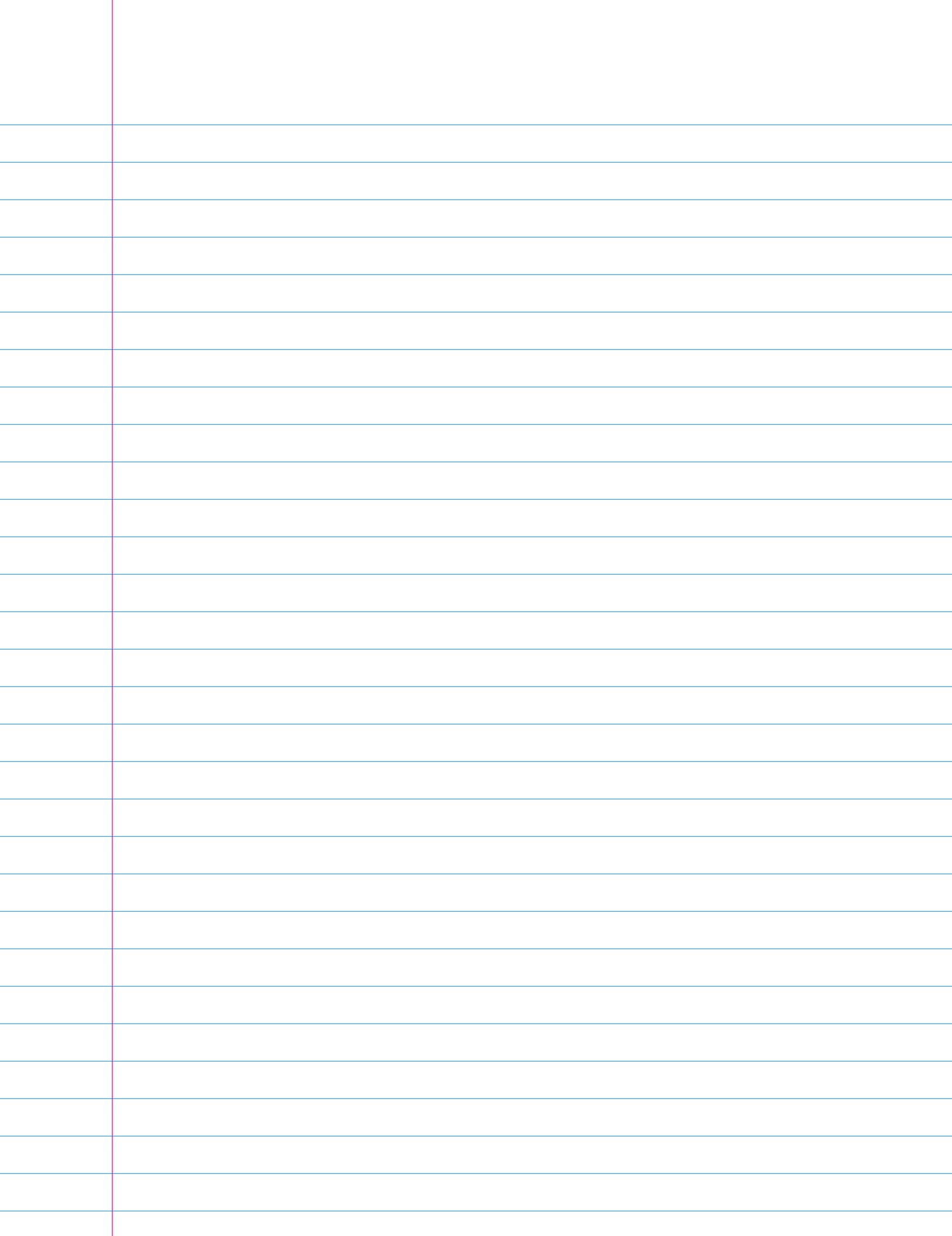
Minimize the Classification Error (Distance)
By Grouping Feature Vector to its Right
(Optimal) Cluster (Class)

"Argmin" Or "Min" Minimization



$\arg \min_S$ for All the possible
Classes

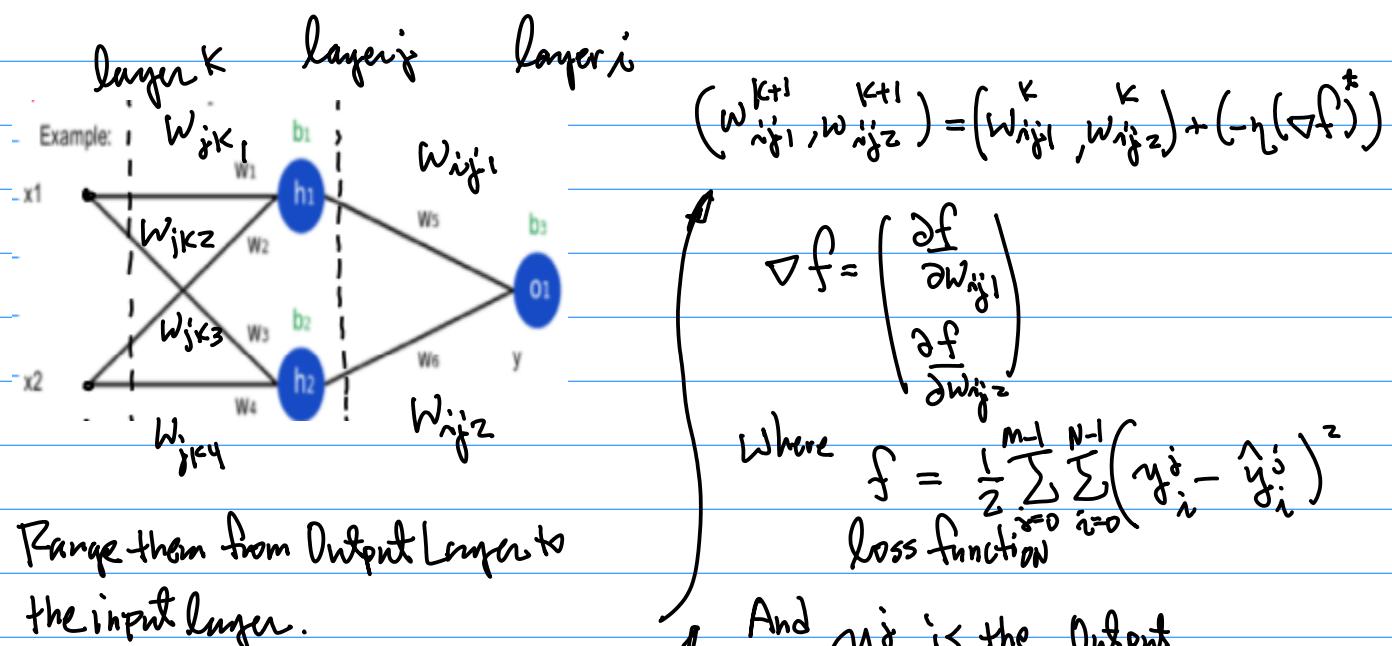
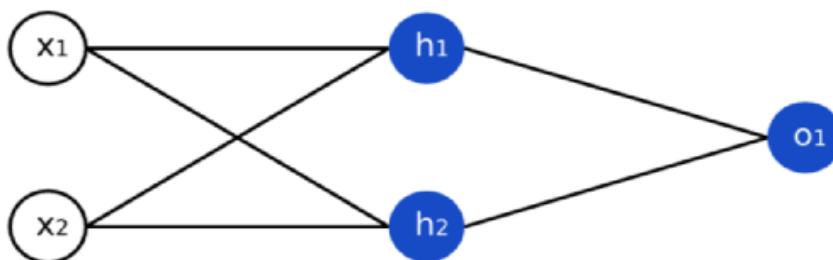




Feed Forward NN

<https://victorzhou.com/blog/intro-to-neural-networks/>

Input Layer Hidden Layer Output Layer



$w_{ijf} \rightarrow w_{jik}$, Scalable

$$(x_1^{k+1}, x_2^{k+1}) = (x_1^k, x_2^k) + [-\eta(\nabla f)^t]$$

$$\nabla f = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_i} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

y_i^j i'th output @ the last layer.
@ Experiment j
Drop j, Rewrite Subscript.

$$y_{ijk} = f\left(\sum_k w_{ijk} x_{ijk} + b_{ij}\right)$$

$$= f\left(h(w_{ijk}, x_{ijk}, b_{ij})\right)$$

$$\frac{\partial}{\partial w_{ijk}} L = \frac{\partial}{\partial w_{ijk}} \left(\frac{1}{2} \sum_m \sum_k (y_{ijk}^m - \hat{y}_{ijk}^m)^2 \right)$$

$$\rightarrow 2 \cdot \frac{1}{2} \cdot \underbrace{\left(\sum_m \sum_k (y_{ijk}^m - \hat{y}_{ijk}^m) \right)}_{S} \frac{\partial}{\partial w_{ijk}} (-\hat{y}_{ijk}^m) = \leq \frac{\partial}{\partial w_{ijk}} f\left(\sum_k w_{ijk} x_{ijk} + b_{ij}\right)$$

$$= -S f'(.) \frac{\partial}{\partial w_{ijk}} \left(\sum_k w_{ijk} x_{ijk} + b_{ij} \right)$$

$$= -S f'(.) \frac{\partial}{\partial w_{ijk}} \sum_k w_{ijk} = -S f'.$$