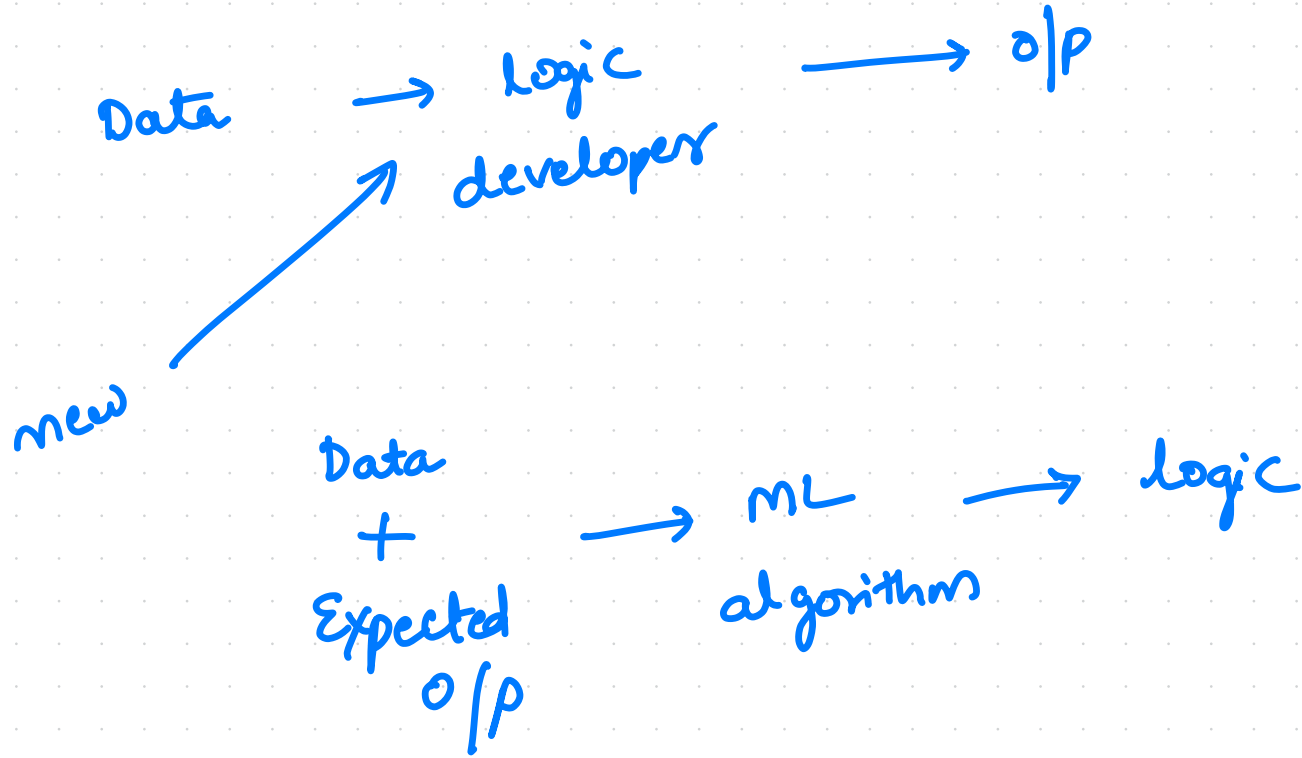




Deep Learning



Deep learning

Overfit

- Regularization : $L1$ & $L2$
- Drop out ✓
- Early stopping ✓

Choose optimizers efficiently,
Normalization, Initializers

Underfit

- Increase Complexity
- Use pre-trained



$$w = w - \eta \frac{\partial L}{\partial w}$$

↑
↑
small

$$\frac{\partial L}{\partial w_5}$$

$$\frac{\partial L}{\partial w_4} = \frac{\partial L}{\partial w_5} \frac{\partial w_5}{\partial w_4}$$

$$10 - 0.000001 \rightarrow 9.999999$$

Vanishing
gradients

Exploding
Gradients

Initializers

Glorot & He initialization

l_1 & l_2

Batch
32

l_2 regularization
→ Forward pass

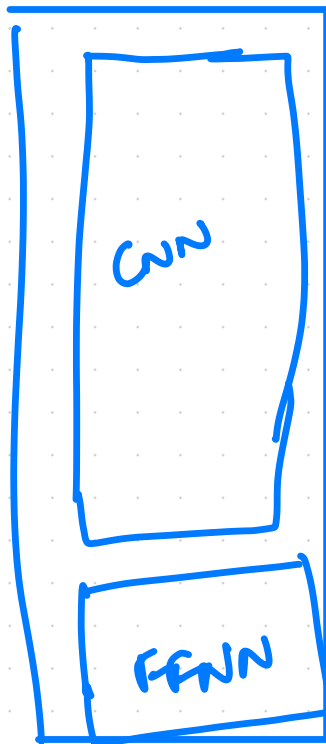
→ Loss + l_2 loss

↖ l_1 loss

$l_1 - l_2$ (h)

Academia :

Dataset A



CNN
+

Pooling

Car or
Bus

Feature
Extractor

classification

Transfer Learning

man/
Bike or
Bi Cycle /
women

Dataset B

Freeze
layers

CNN
Data
set
A

fine tuning

DLNN

4 neurons

✓ softmax

Imagenet Competition



Labelled Dataset

classification

object Detection

Segmentation

Human pose
detection

AlexNet

Image: 224 (height) × 224 (width) × 3 (channels)

Convolution with 11 × 11 kernel + 4 stride: 54 × 54 × 96

↓ ReLu

Pool with 3 × 3 max. kernel + 2 stride: 26 × 26 × 96

↓ ReLu

Convolution with 5 × 5 kernel + 2 pad: 26 × 26 × 256

Pool with 3 × 3 max. kernel + 2 stride: 12 × 12 × 256

Convolution with 3 × 3 kernel + 1 pad: 12 × 12 × 384

↓ ReLu

Convolution with 3 × 3 kernel + 1 pad: 12 × 12 × 384

↓ ReLu

Convolution with 3 × 3 kernel + 1 pad: 12 × 12 × 256

↓ ReLu

Pool with 3 × 3 max. kernel + 2 stride: 5 × 5 × 256

↓ flatten

Dense: 4096 fully connected neurons

↓ ReLu, dropout p=0.5

Dense: 4096 fully connected neurons

↓ ReLu, dropout p=0.5

Dense: 1000 fully connected neurons

Output: 1 of 1000 classes

Layer	# filters / neurons	Filter size	Stride	Padding	Size of feature map	Activation function
Input	-	-	-	-	227 x 227 x 3	-
Conv 1	96	11 x 11	4	-	55 x 55 x 96	ReLU
Max Pool 1	-	3 x 3	2	-	27 x 27 x 96	-
Conv 2	256	5 x 5	1	2	27 x 27 x 256	ReLU
Max Pool 2	-	3 x 3	2	-	13 x 13 x 256	-
Conv 3	384	3 x 3	1	1	13 x 13 x 384	ReLU
Conv 4	384	3 x 3	1	1	13 x 13 x 384	ReLU
Conv 5	256	3 x 3	1	1	13 x 13 x 256	ReLU
Max Pool 3	-	3 x 3	2	-	6 x 6 x 256	-
Dropout 1	rate = 0.5	-	-	-	6 x 6 x 256	-

Layer	# filters / neurons	Filter size	Stride	Padding	Size of feature map	Activation function
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
Dropout 1	rate = 0.5	-	-	-	6 x 6 x 256	-
Fully Connected 1	-	-	-	-	4096	ReLU
Dropout 2	rate = 0.5	-	-	-	4096	-
Fully Connected 2	-	-	-	-	4096	ReLU
Fully Connected 3	-	-	-	-	1000	Softmax

classification

Softmax

