CSE 575: Statistical Machine Learning (Spring 2021)
Instructor: Nupur Thakur

Neural Networks & Deep Learning

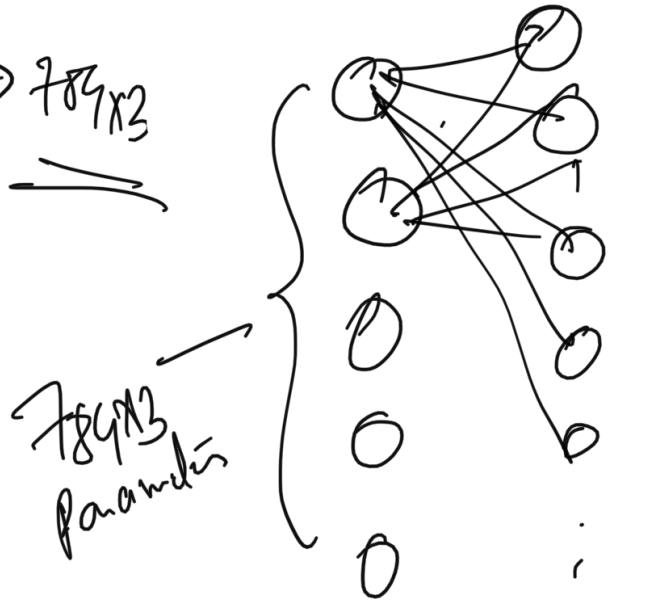
Table of contents

FC (fully connected).

Number of parameters in dense layers

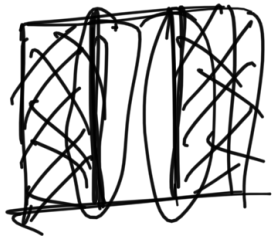
224 x 224

28 x 28 x 3 ⇒ 784 x 3

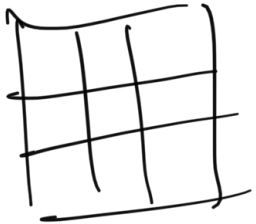


huge no. of parameters: 784 x 3 x 1000.

Convolutional Neural Network (CNN)

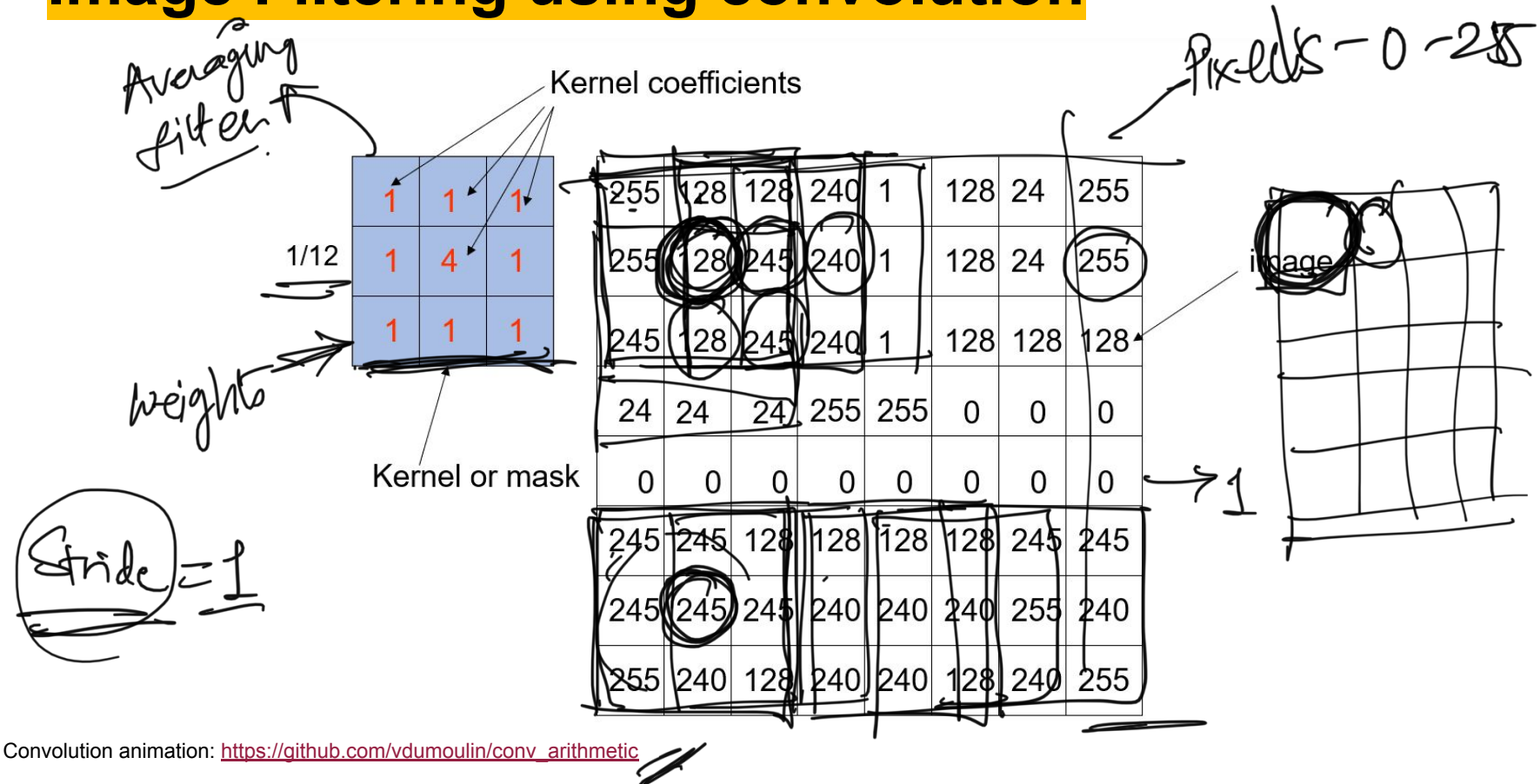


- Basic building block is the convolution for image processing
- Weight sharing - reduces parameters
- Learn the kernels based on the task.



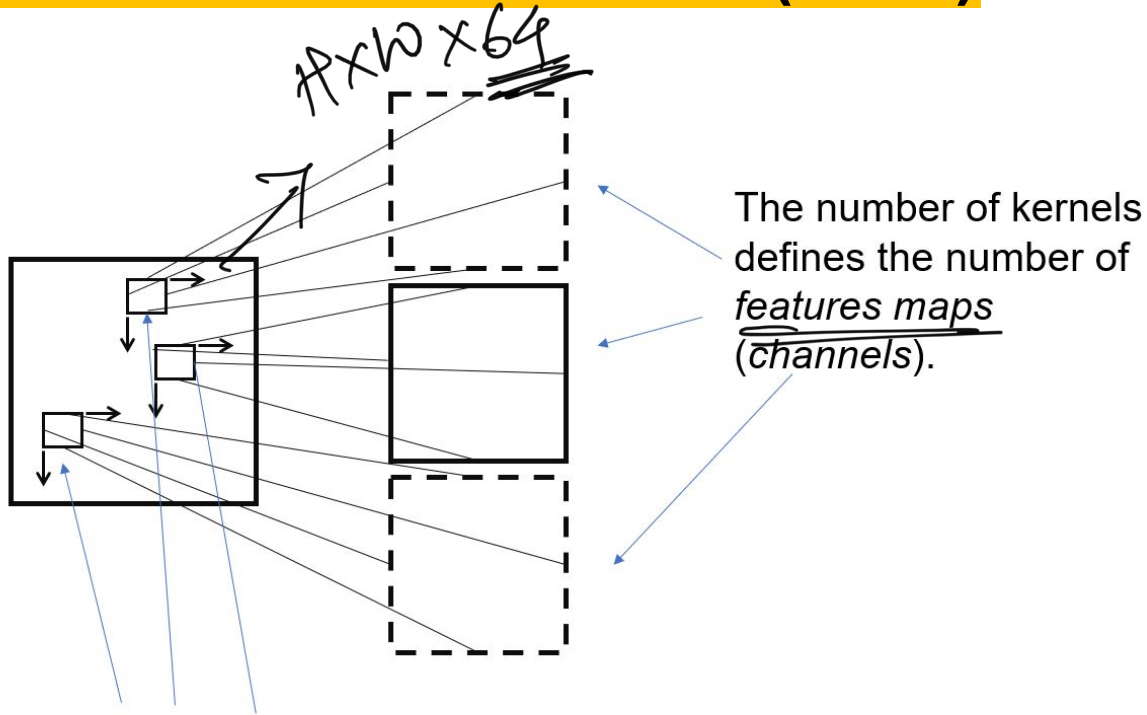
Edge (Vertical).

Image Filtering using convolution



Convolution animation: https://github.com/vdumoulin/conv_arithmetic

Convolutional Neural Network (CNN)



The sizes of the kernels define the *receptive fields*.

Summary of CNNs : https://cs231n.github.io/convolutional-networks/

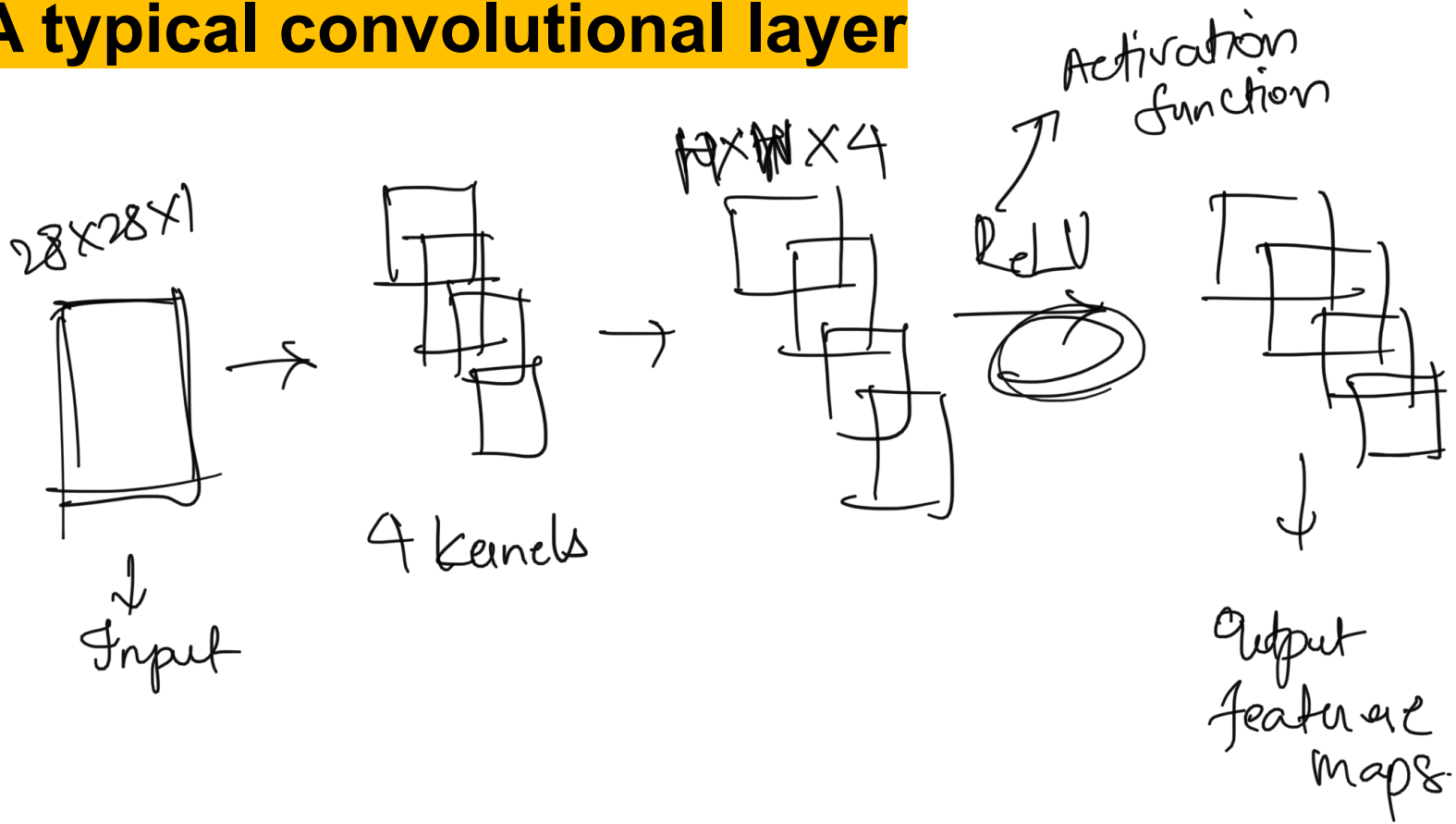
Padding in convolutional layer

- Zero-padding
  - Equally value the boundary pixels.
- Same / less reduced size of the output.

255	128	128	240	1	128	24	255
255	128	245	240	1	128	24	255
245	128	245	240	1	128	128	128
24	24	24	255	255	0	0	0
0	0	0	0	0	0	0	0
245	245	128	128	128	128	245	245
245	245	245	240	240	240	255	240
255	240	128	240	240	128	240	255

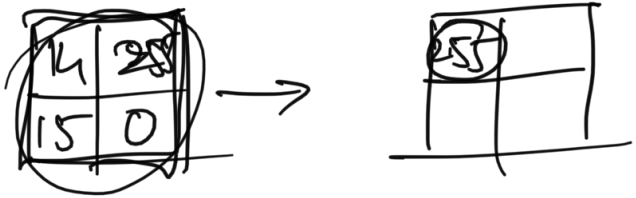
Summary of CNNs : https://cs231n.github.io/convolutional-networks/

A typical convolutional layer



Summary of CNNs : https://cs231n.github.io/convolutional-networks/

Pooling in CNN

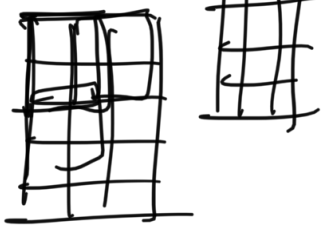


- Used to reduce the size of the output, thereby reducing the number of parameters and preventing overfitting.
- Done on every channel along the height and width of the output. So, the number of channels remain the same.

— Max Pooling

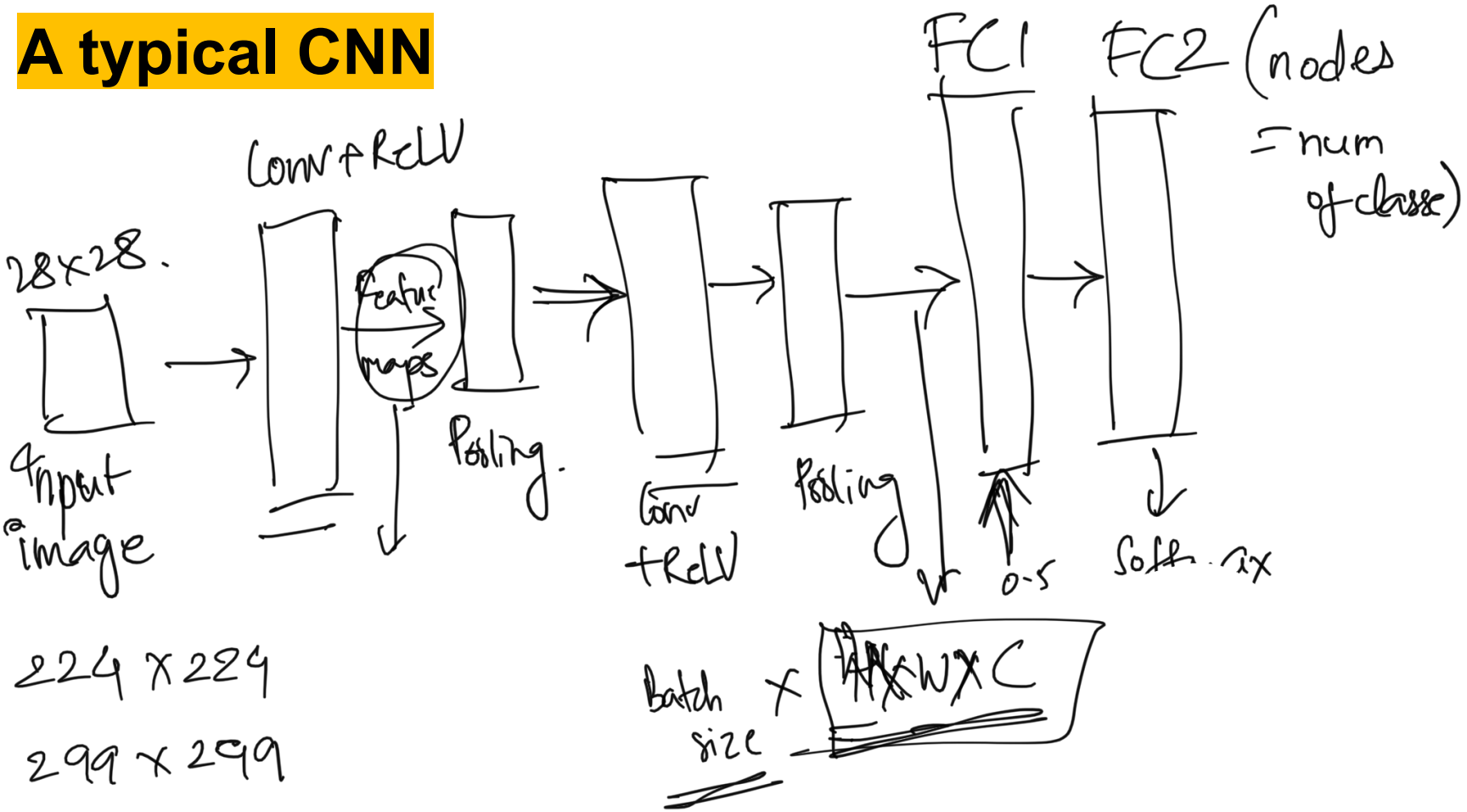
— Average Pooling.

— Kernel size, type of pooling, stride.



28×28 ⇒ 784

A typical CNN



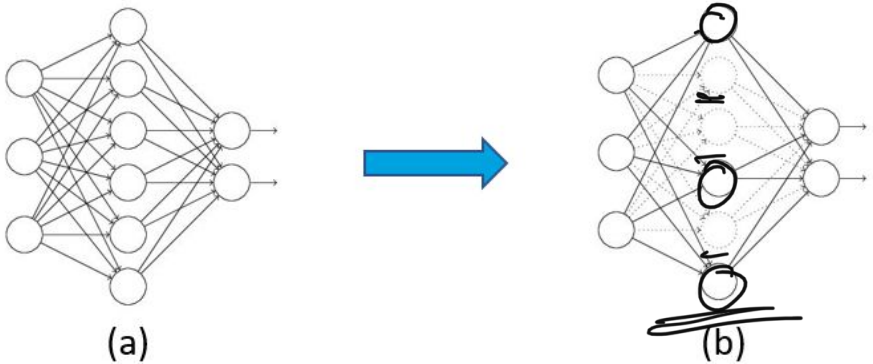
Why is regularization necessary?

- Huge parameter space - poor solutions
- Convergence to local minimum which leads to good performance on training data only.
- To overcome -
  - L2 norm of weights
  - Dropout
  - Batch normalization.

Dropout

- Obtain (b) by randomly deactivate some hidden nodes in (a).
- Reducing co-adaptation of neurons

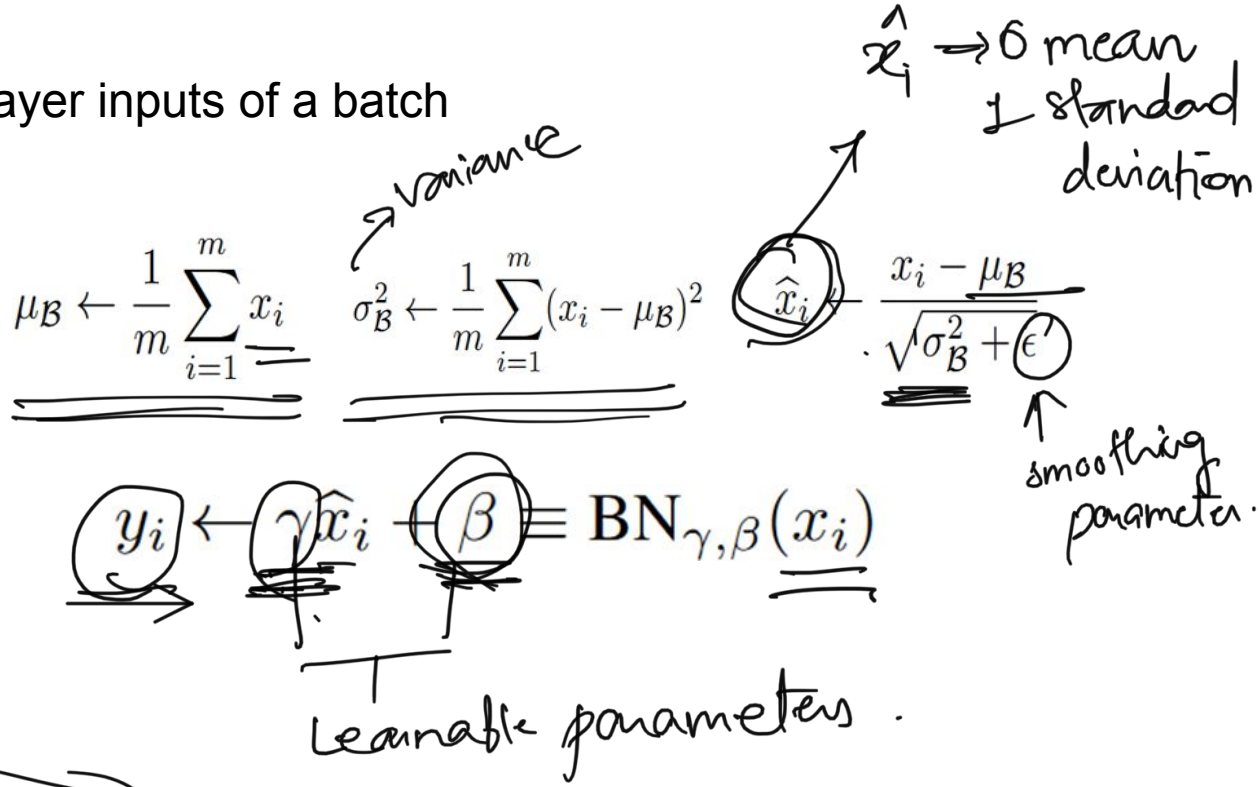
Probability of dropping a neuron. (0 - 1) - (p)



Refer for details on Dropout : https://jmlr.org/papers/volume15/srivastava14a/srivastava14a.pdf

Batch Normalization

- Normalizes layer inputs of a batch

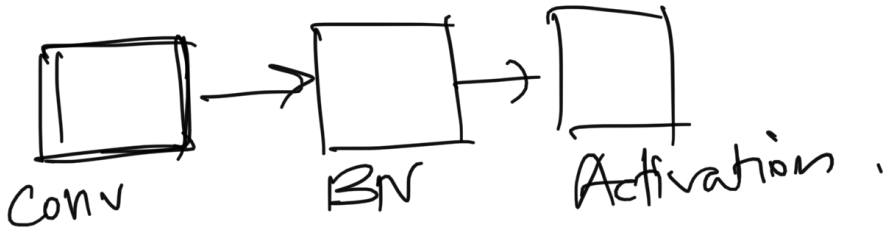


Paper about BN: https://arxiv.org/pdf/1502.03167.pdf

Batch Normalization - Advantages

- Use of higher learning rates
- Covariate shift is reduced
- Avoids vanishing gradients

- Makes the network less sensitive to weight initialization.



Paper about BN: https://arxiv.org/pdf/1502.03167.pdf

Questions?