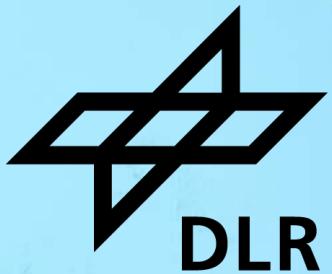


INTRODUCTION TO DEEP LEARNING

PART II – ADVANCED CONCEPT AND CNN

Auliya Fitri, Sai Vemuri, Sreerag Naveenachandran

**Machine Learning Group
Institute of Data Science**



Schedule



Date	Time	Activity
13.11.2025 Day 1	09:00 - 10:00	Introduction and basics
	10:00 - 10:30	Hands-on I
	10:30 - 10:45	Coffee break
	10:45 - 11:45	Advanced concept and Convolutional Neural Network
	11:45 - 12:15	Hands-on II
	12:15 - 12:30	Recap Day 1
14.11.2025 Day 2	09:00 - 10:00	Deep Generative Model
	10:00 - 10:30	Hands-on III
	10:30 - 10:45	Coffee break
	10:45 - 11:45	Transformer
	11:45 - 12:15	Hands-on IV
	12:15 - 12:30	Code and knowledge sources + closing

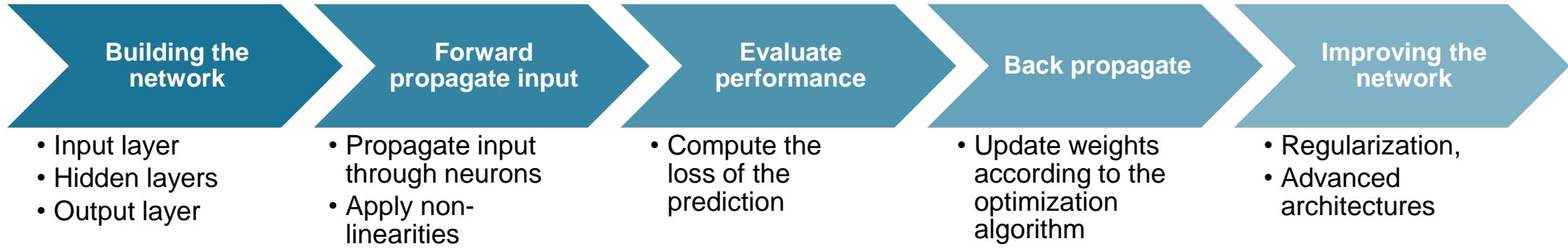


II. Advanced Concept and Convolutional Neural Network

- Regularization
- Variants of Neural Networks
- Convolutional Neural Networks (CNN)

Inspired by lectures from Paris Saclay and MIT; images taken from these, if not noted otherwise

Neural network concepts

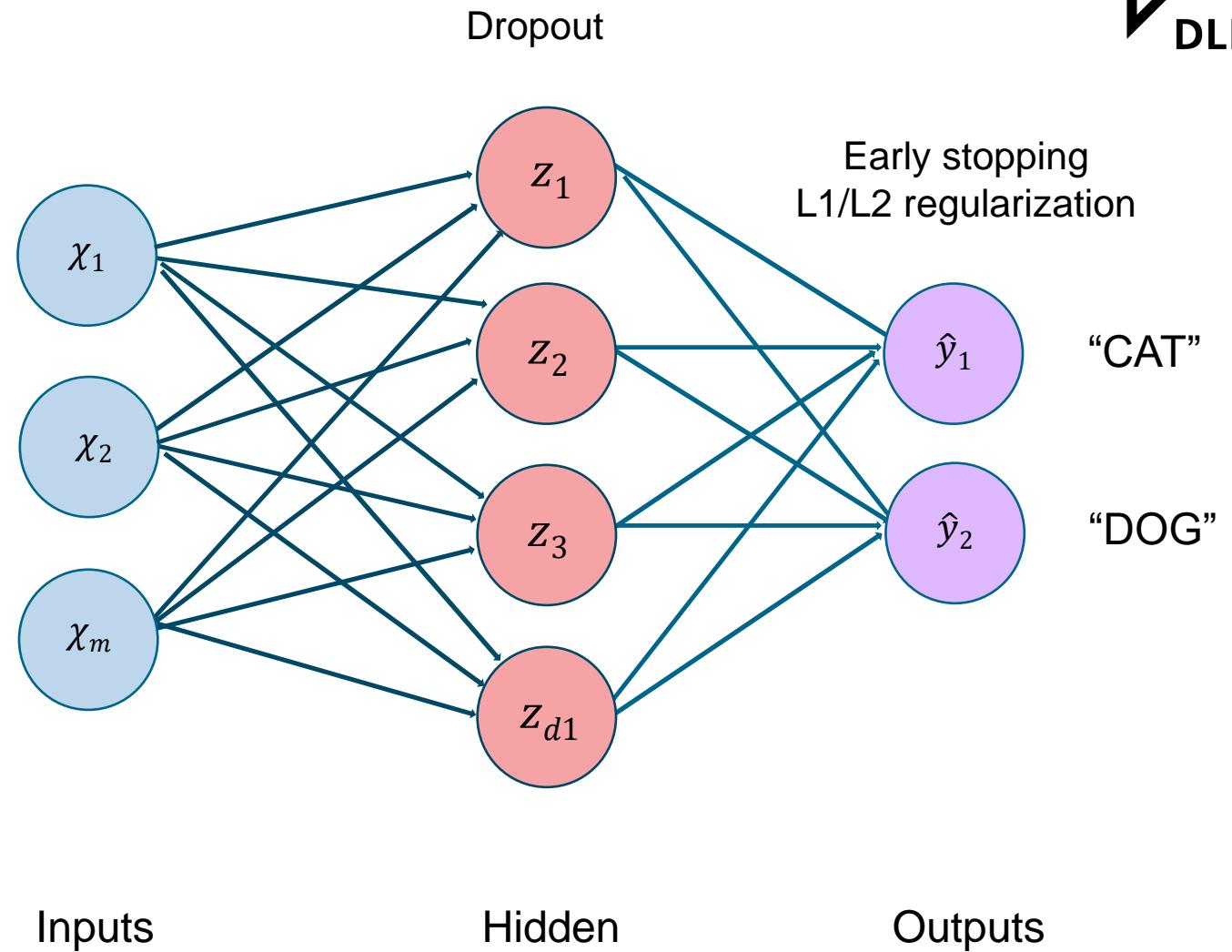


Improve the Network: Regularization

Data Augmentation

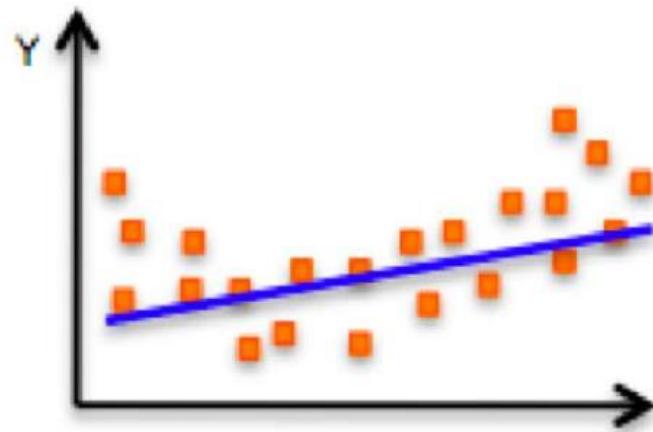


$$\begin{bmatrix} \chi_1 \\ \chi_2 \\ \vdots \\ \chi_m \end{bmatrix}$$



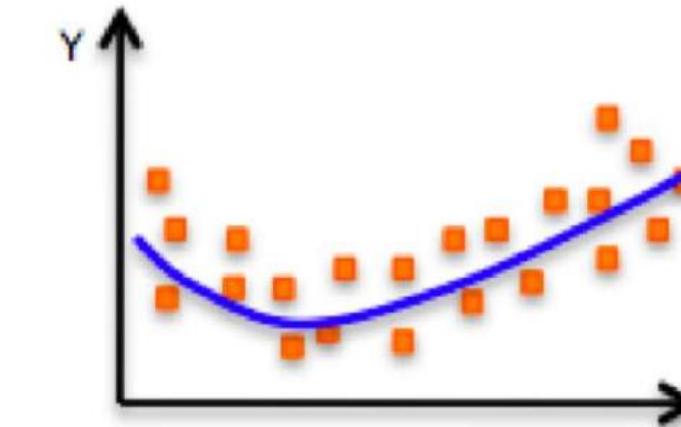
Regularization

The overfitting problem

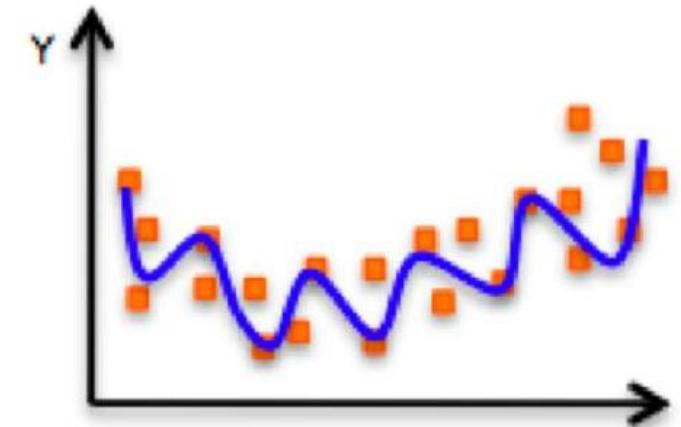


Underfitting
Model does not have capacity
to fully learn the data

= High bias



Ideal fit ← → **Tradeoff**

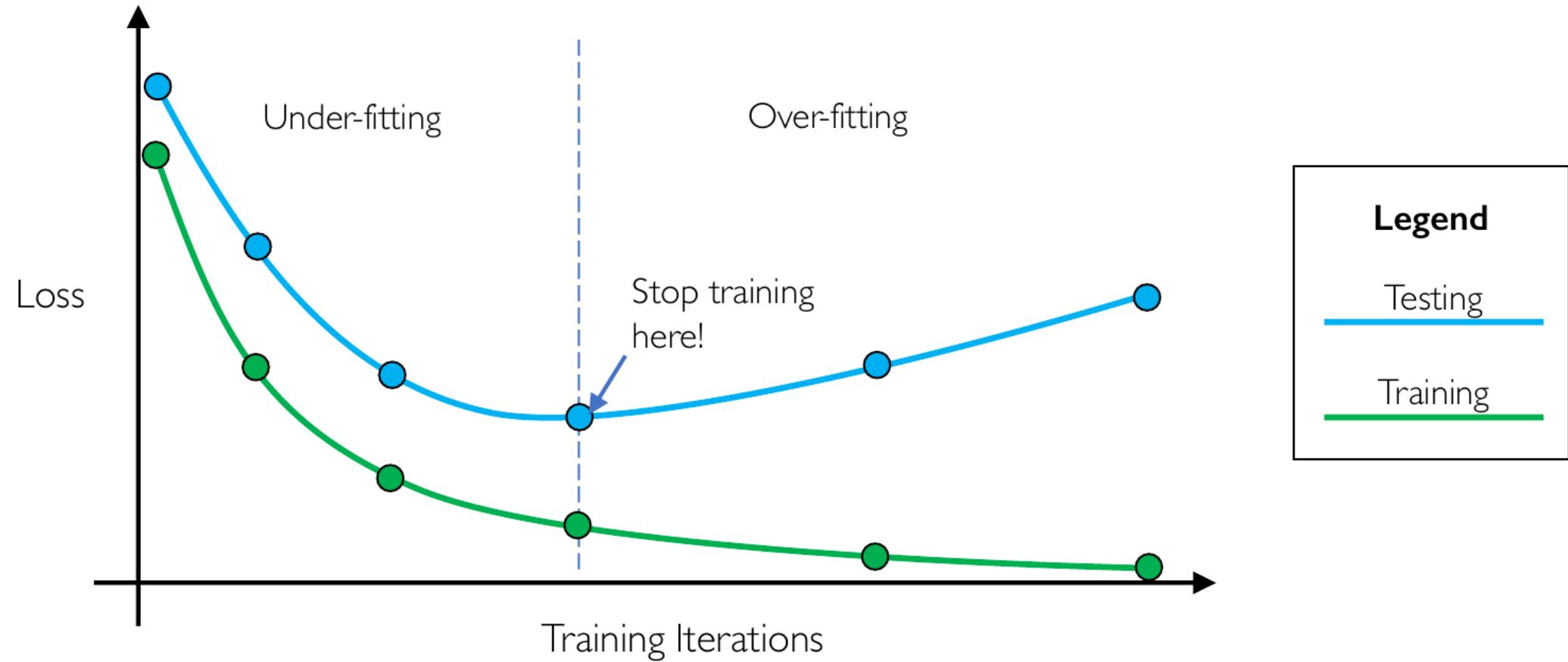


Overfitting
Too complex, extra parameters,
does not generalize well

= High variance

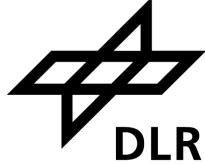
Regularization

Preventing overfitting: Early stopping



Regularization

Preventing overfitting: l_1 regularization



We want to find the network weights that **achieve the lowest loss**

$$\mathbf{W}^* = \underset{\mathbf{W}}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^n \mathcal{L}(f(\mathbf{x}^{(i)}; \mathbf{W}), y^{(i)}) + \frac{\lambda}{2} \sum_l |\mathbf{w}_l|$$

→ Leads to sparser weights (more zeroes in weights) that are not too adapted to the data at hand

Regularization

Preventing overfitting: Data augmentation



Original



Flip



Random crop



Contrast



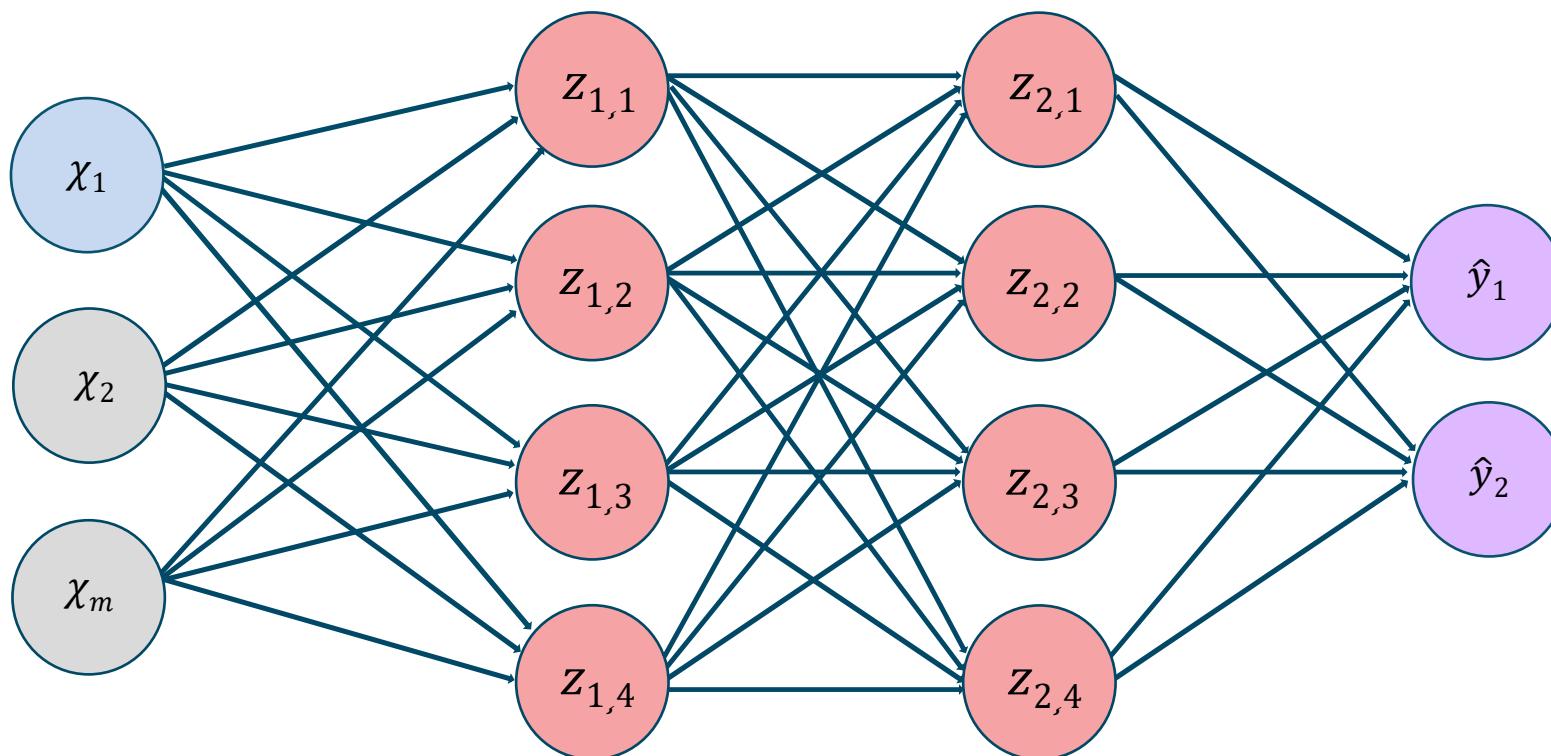
Tint



Regularization

Preventing overfitting: Dropout

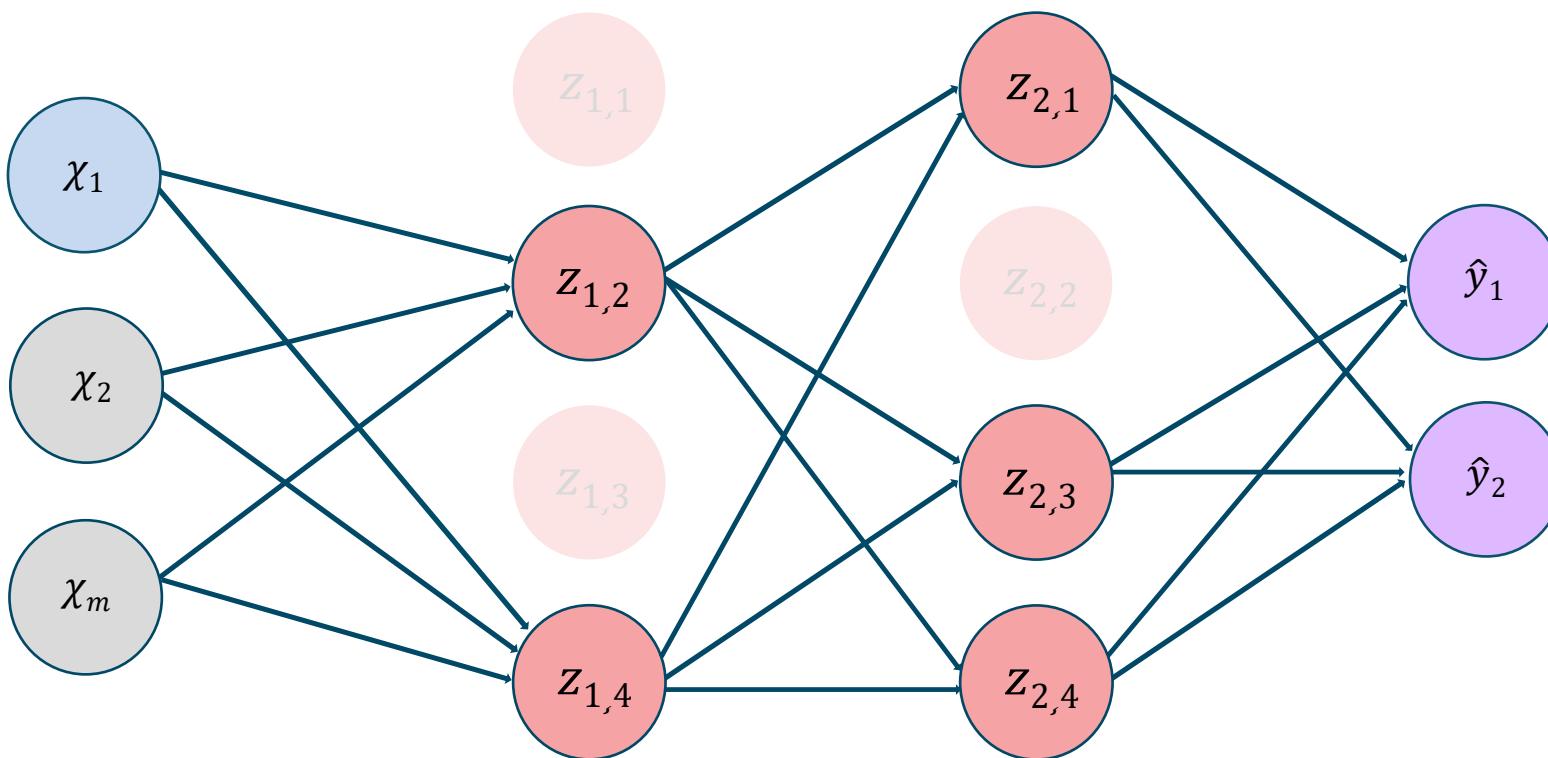
During training, randomly set some activations to 0



Regularization

Preventing overfitting: Dropout

During training, randomly set some activations to 0



Other ways to improve the network

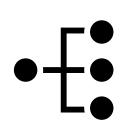


Data preprocessing

Data Cleaning (lack/noise)

Data Transformation (normalization, ...)

Data Reduction (aggregation, ...)



Network initialization

Random Initializations (e.g. Glorot)



Batch normalization

standardizes layer inputs to stabilize learning process & reducing training epochs



Optimizers

(Stochastic methods like Adam)



Training Schedules



Hyperparameters tuning



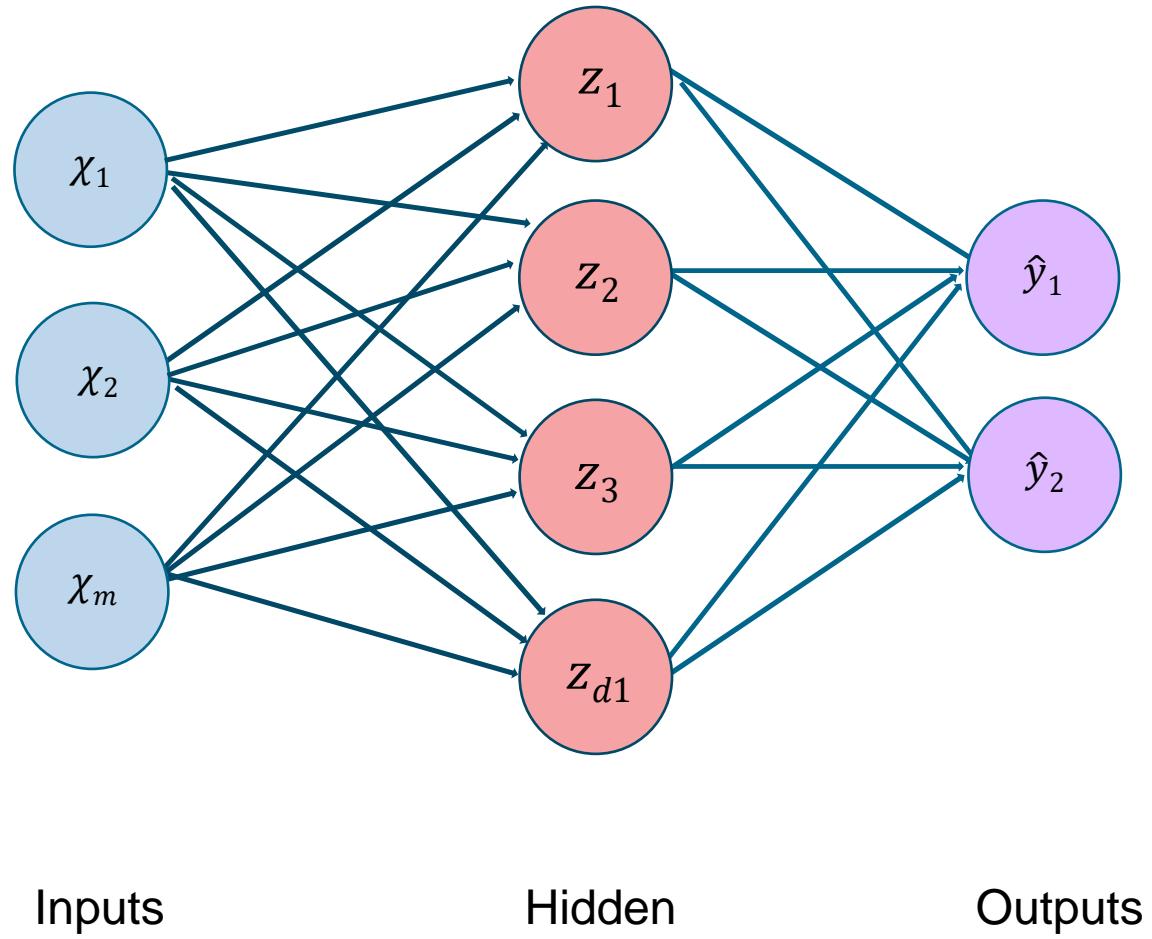
Minibatch composition

(shuffle data, take only a subset, ...)

Variants of Neural Networks



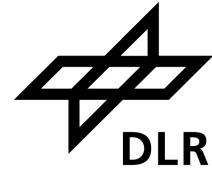
$$\begin{bmatrix} \chi_1 \\ \chi_2 \\ \vdots \\ \chi_m \end{bmatrix}$$



“CAT”

“DOG”

Types of Neural Networks Layer

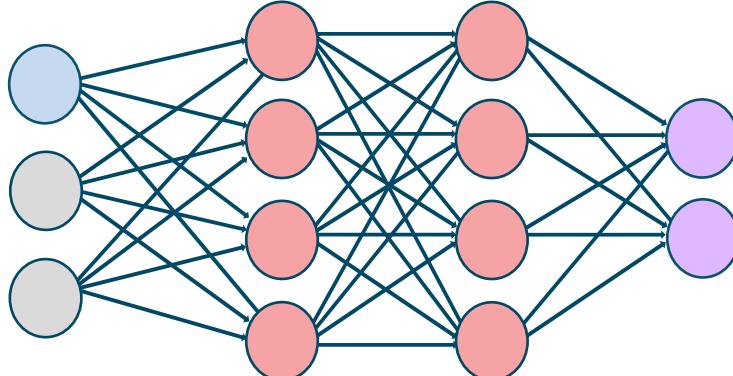


Application	Layer Type			
	Fully Connected (dense)	Convolution	Deconvolution	Recurrent
Image Classification	✓	✓		
Image segmentation		✓	✓	
Text processing	✓			✓
Speech recognition	✓			✓
Time series	✓			✓
Image-to-image translation (GAN)	✓	✓	✓	
Autoencoders	✓	✓		
Deepfake	✓	✓		✓

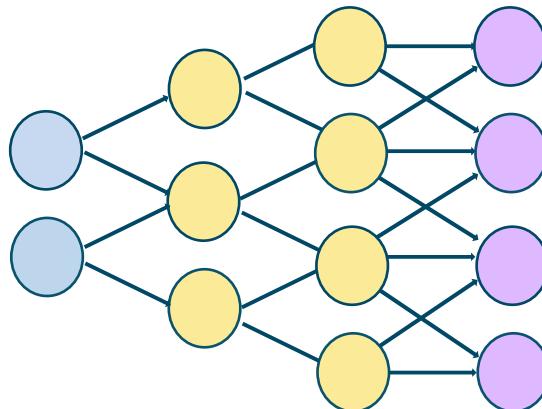
Variants of Neural Network - Examples



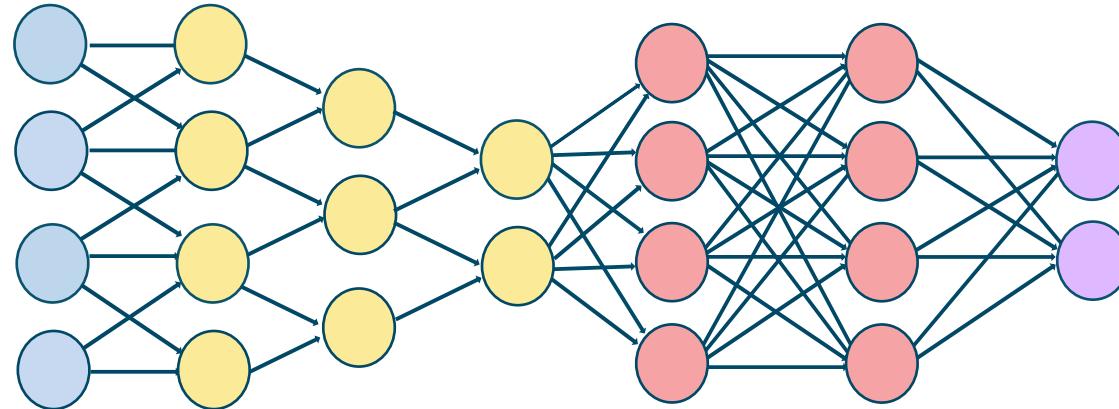
Deep Feed Forward (DFF) (fully connected)



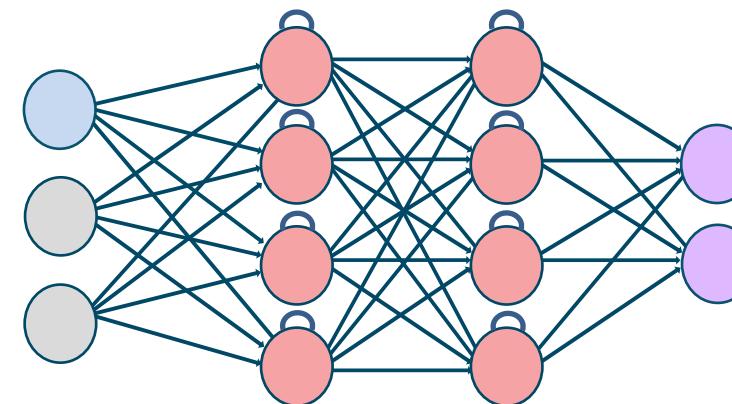
Deconvolution network (DN)



Deep convolution network (DCN)

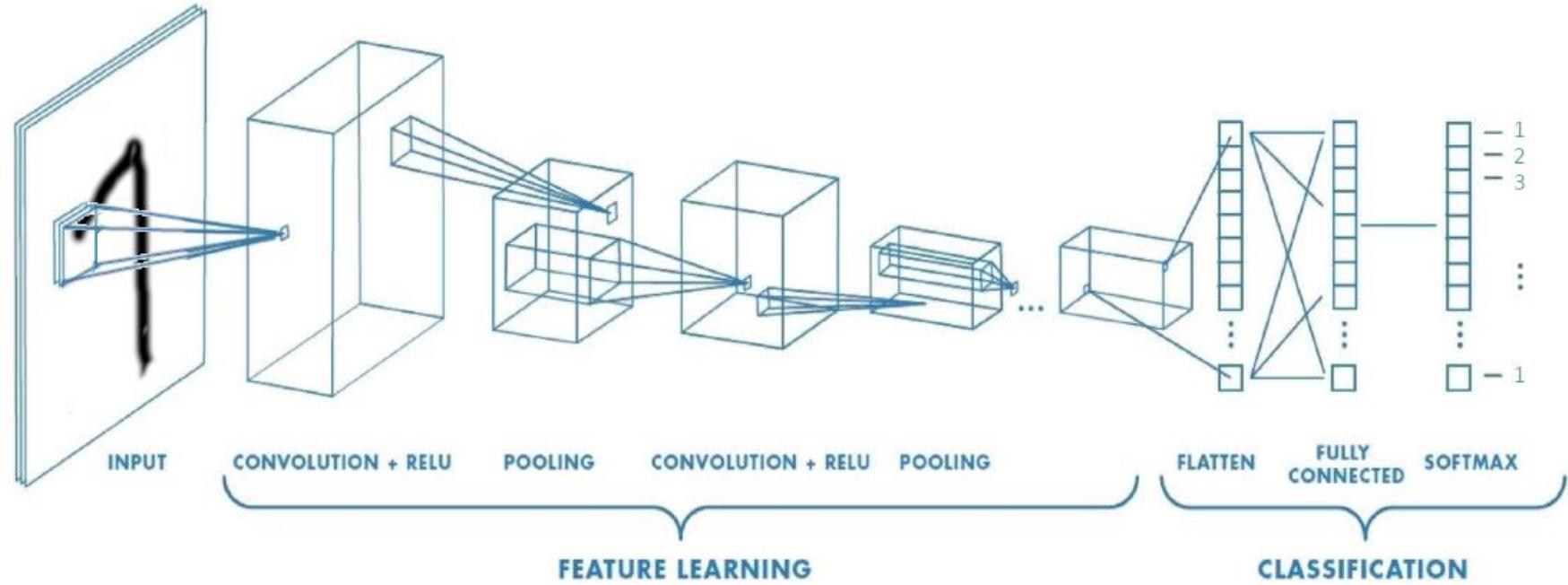


Recurrent neural network (RNN)

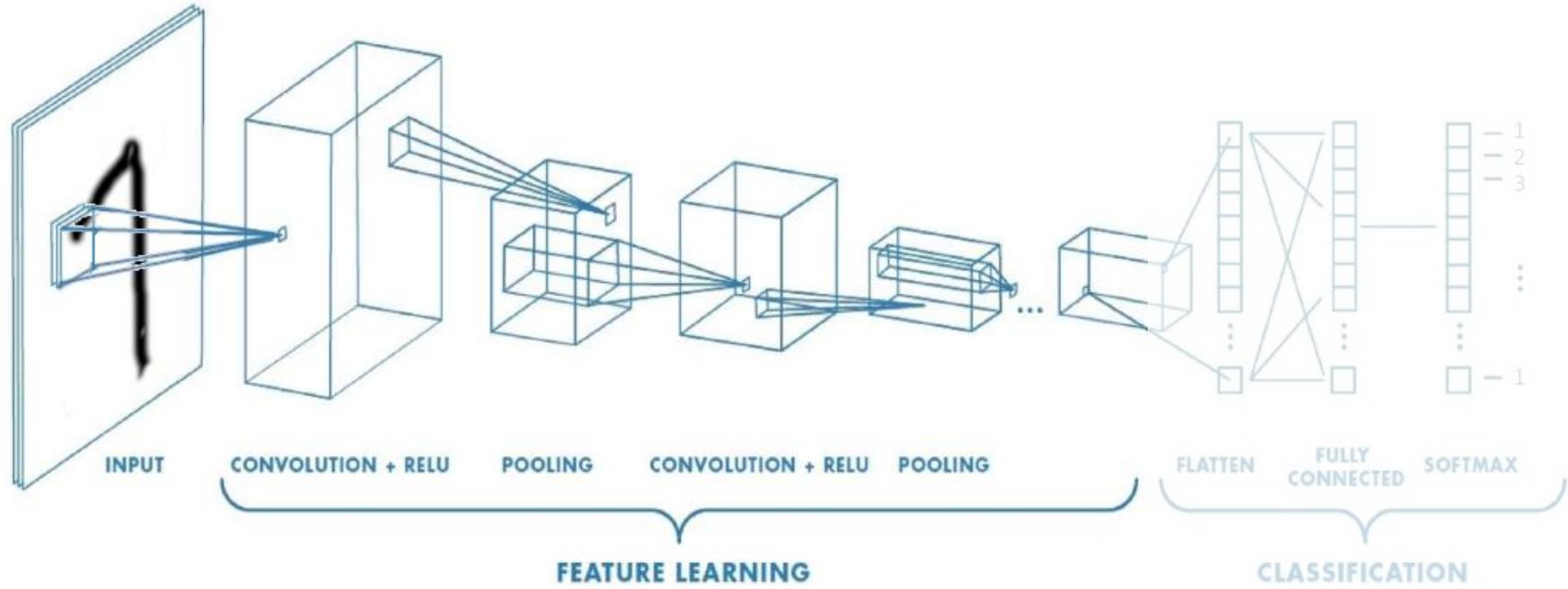


● = Input
● = Output

Convolutional Neural Networks



Feature Learning



1. Learn features in input image through **convolution**
2. Introduce **non-linearity** through activation function (real-world data is non-linear!)
3. Reduce dimensionality and preserve spatial invariance with **pooling**

Convolutional Neural Networks

Feature extraction with convolutions



Original



Sharpen



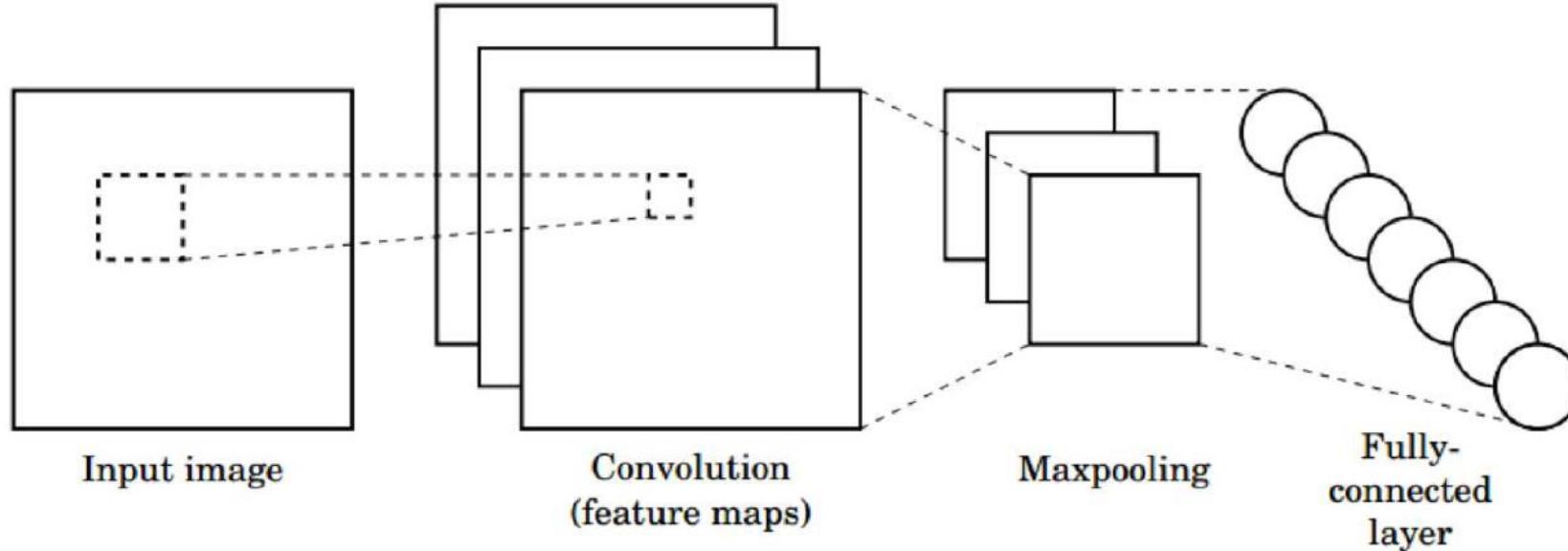
Edge Detect



“Strong” Edge
Detect

Convolutional Neural Networks

Building a CNN



- 1. Convolution:** Apply filters with learned weights to generate feature maps.
- 2. Non-linearity:** Often ReLU.
- 3. Pooling:** Downsampling operation on each feature map.

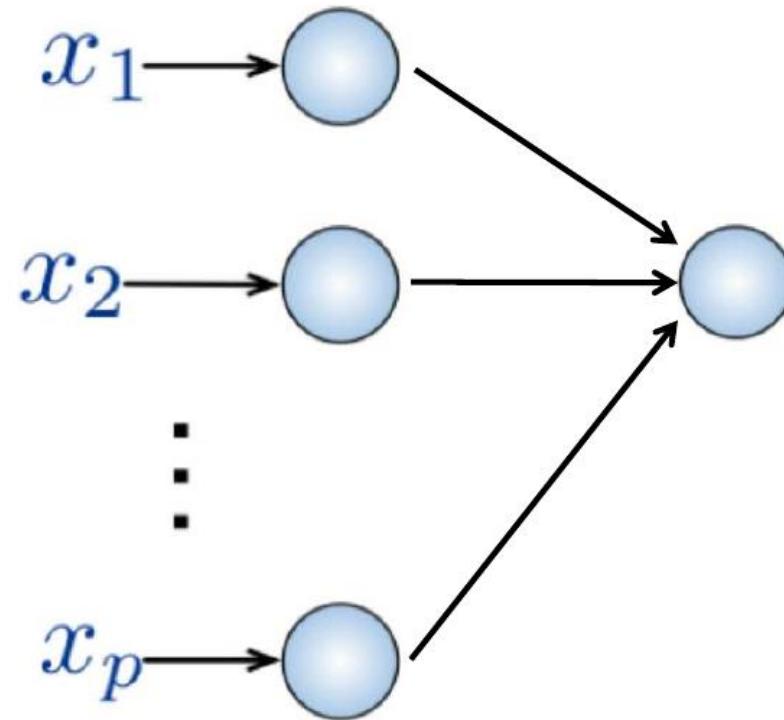
Train model with image data.
Learn weights of filters in convolutional layers.

Convolutional Neural Networks

Learning on image data

Input:

- 2D image
- Vector of pixel values



Fully Connected:

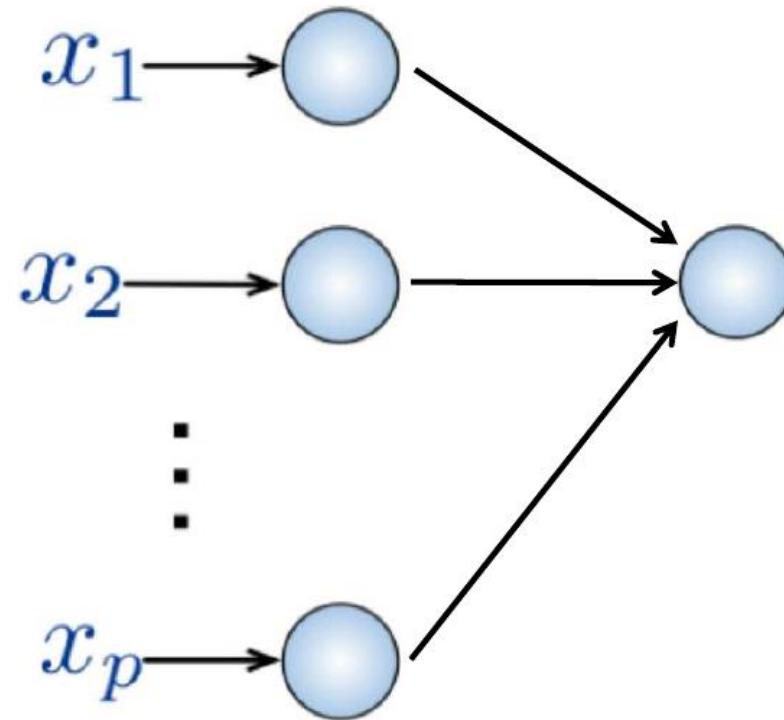
- Connect neuron in hidden layer to all neurons in input layer
- No spatial information!
- And many, many parameters!

Convolutional Neural Networks

Learning on image data

Input:

- 2D image
- Vector of pixel values



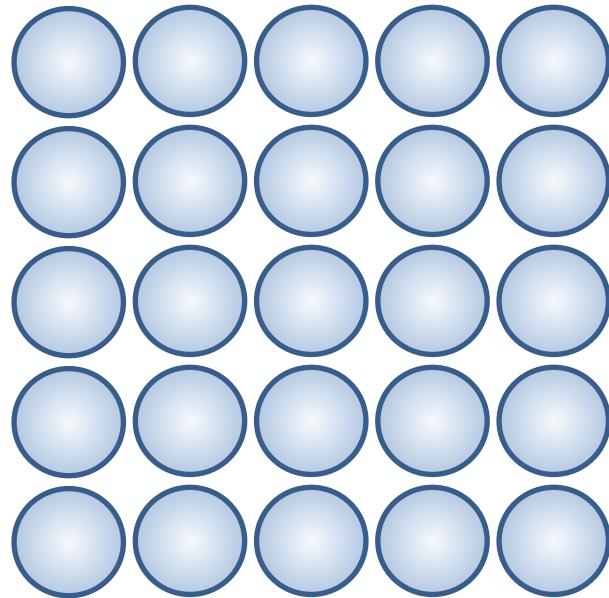
Fully Connected:

- Connect neuron in hidden layer to all neurons in input layer
- No spatial information!
- And many, many parameters!

How can we use **spatial structure** in the input to inform the architecture of the network?

Convolutional Neural Networks

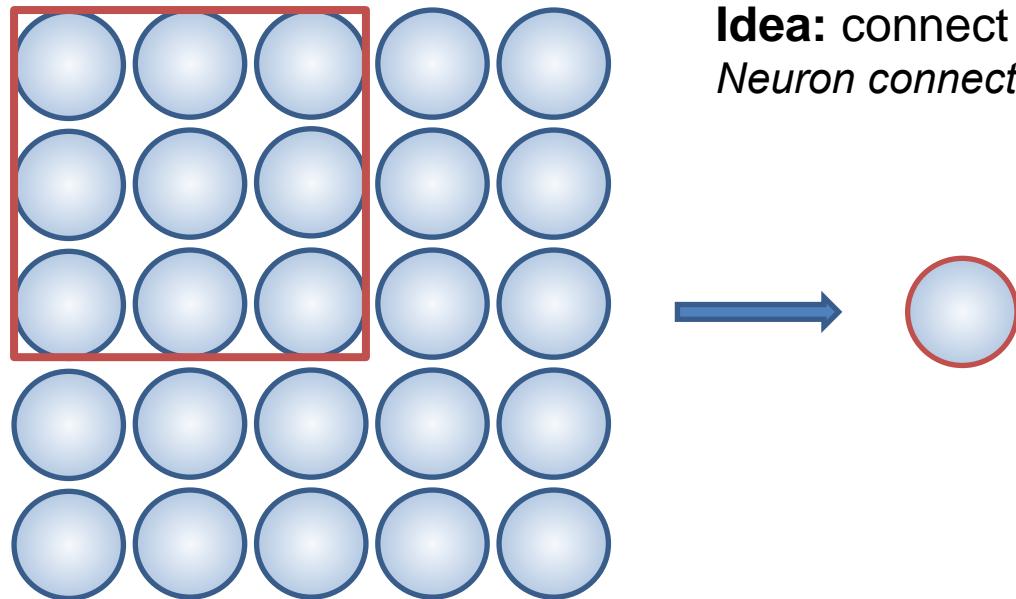
Learning on image data



Input: 2D image (5x5 or 256x256 or ...)
Array of pixel values

Convolutional Neural Networks

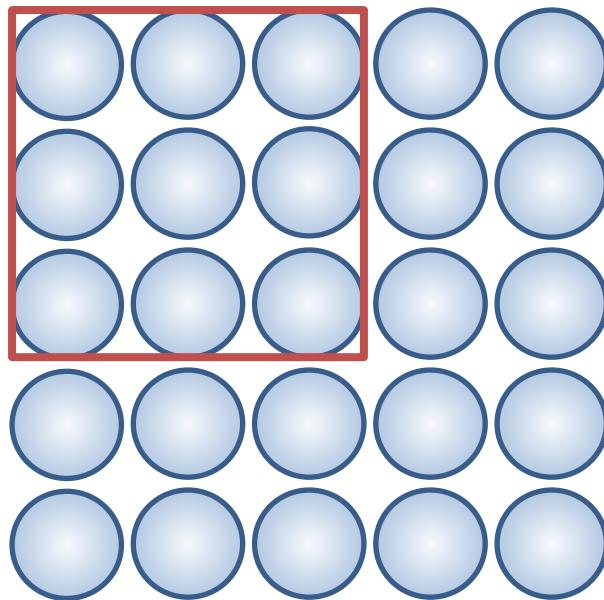
Learning on image data



Idea: connect patches of input to neurons in hidden layer
Neuron connected to the region of input only „sees“ these values.

Convolutional Neural Networks

Learning on image data



3x3 filter: matrix
of weights W_{ij}

For a neuron in hidden layer:

- Take inputs from patch
- Compute weighted sum
- Apply bias

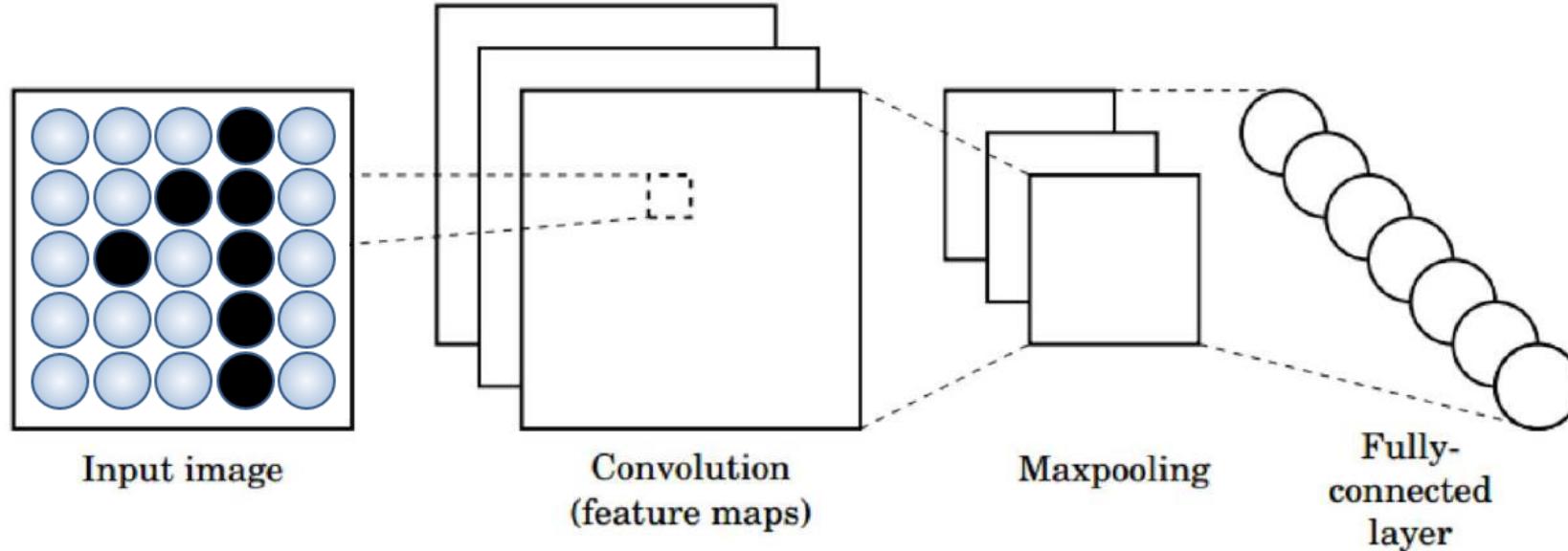
$$\sum_{i=1}^3 \sum_{j=1}^3 W_{ij} X_{i+p,j+q} + b$$

For neuron (p,q) in hidden layer

- 1) Applying a window of weights
- 2) Computing linear combinations
- 3) Activating with non-linear function

Convolutional Neural Networks

Building a CNN - Example



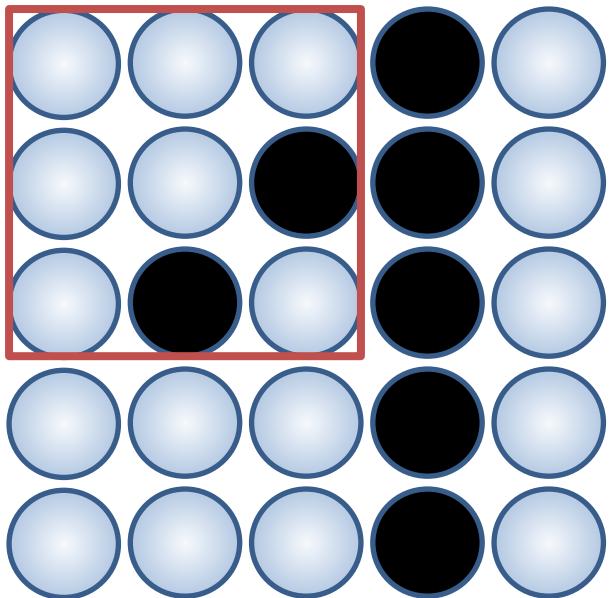
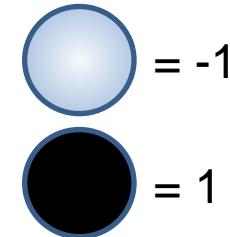
- 1. Convolution:** Apply filters with learned weights to generate feature maps.
- 2. Non-linearity:** Often ReLU.
- 3. Pooling:** Downsampling operation on each feature map.

Train model with image data.

Learn weights of filters in convolutional layers.

Convolutional Neural Networks

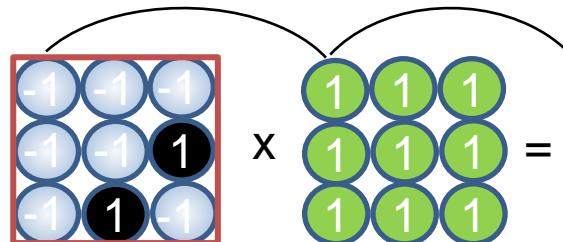
Learning on image data - Example



- Connect patch in input layer to a single neuron in subsequent layer
- Use a **sliding window** to define connections
- Ex: extracting feature with **filter** f_{i_0}



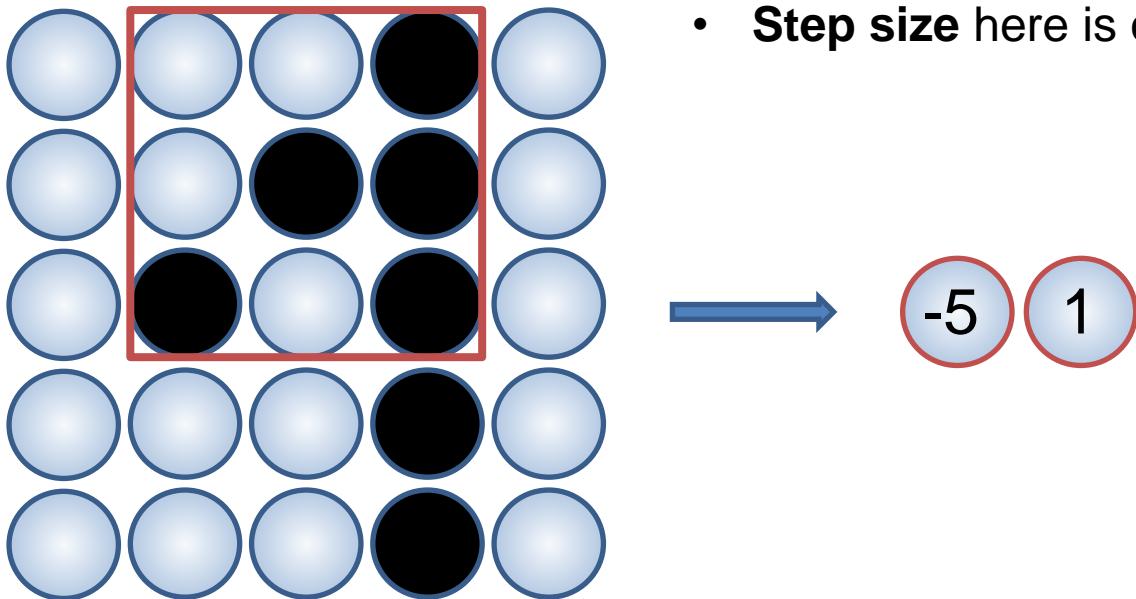
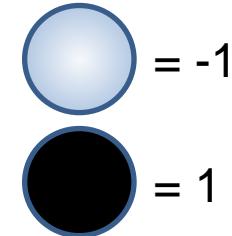
Multiply same indices of clipping and filter and sum it up.



The diagram shows the 3x3 input patch from the previous step being multiplied by a 3x3 filter. The input patch is shown with its values: -1, -1, -1 in the first row; -1, -1, 1 in the second row; and -1, 1, -1 in the third row. The filter is shown with its values: 1, 1, 1 in the first row; 1, 1, 1 in the second row; and 1, 1, 1 in the third row. Arrows point from each value in the input patch to its corresponding value in the filter. The resulting calculation is displayed as $= (-1)+(-1)+(-1)+(-1)+(-1)+1+(-1)+1+(-1) = -5$.

Convolutional Neural Networks

Learning on image data - Example

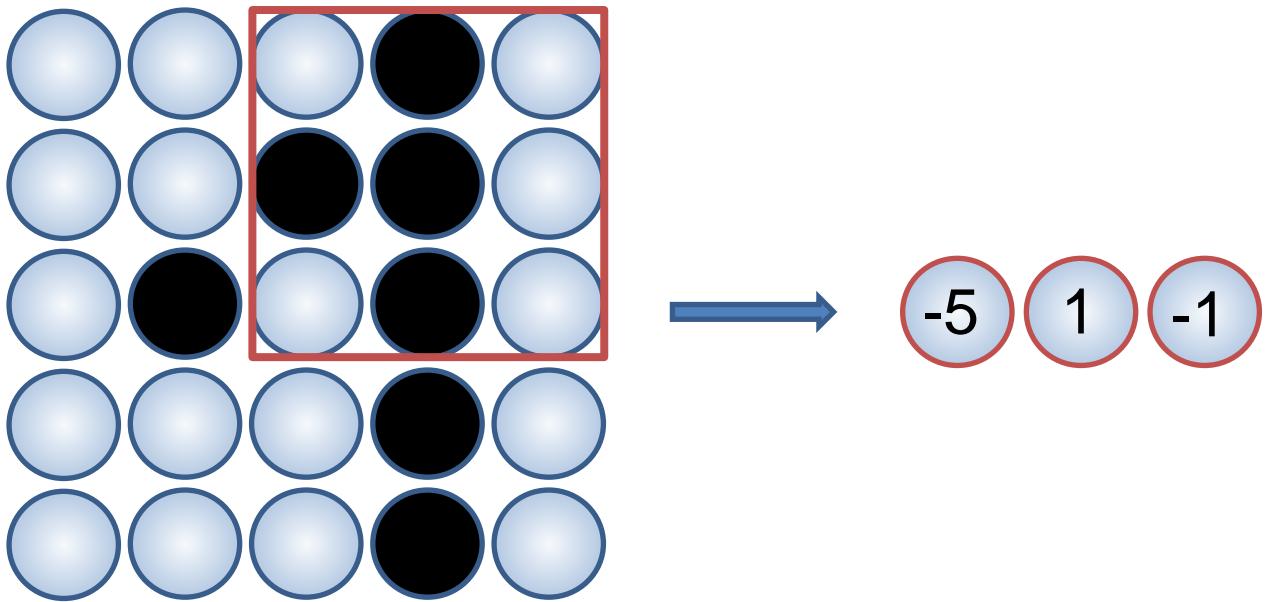


$$\begin{bmatrix} -1 & -1 & 1 \\ -1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = 1$$

Convolutional Neural Networks

Learning on image data - Example

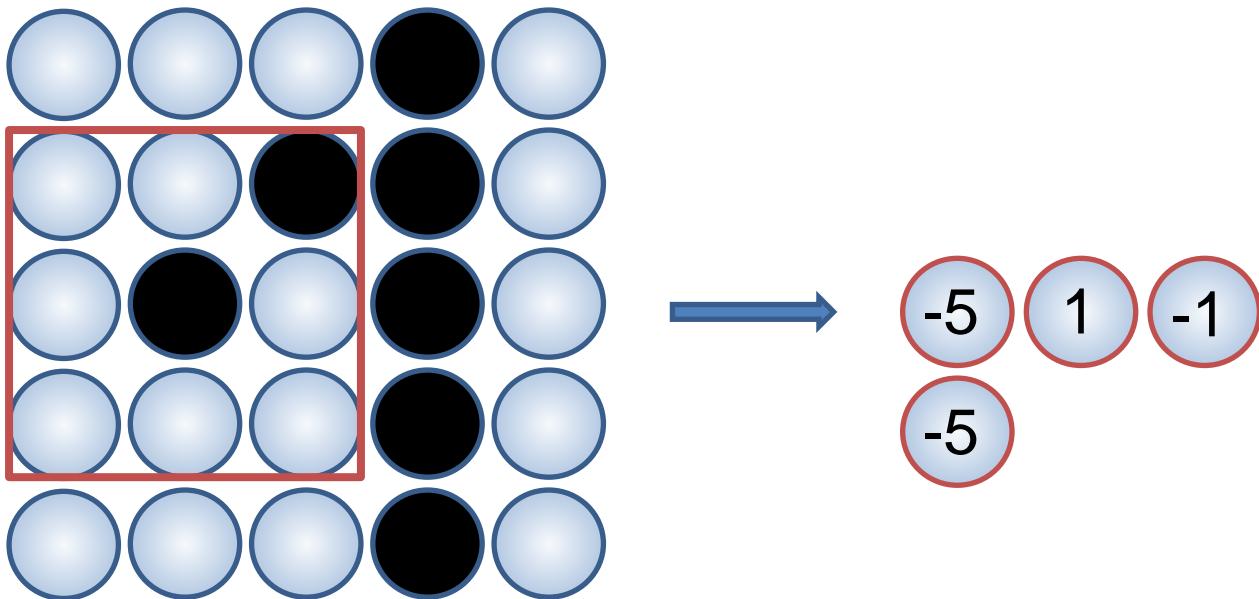
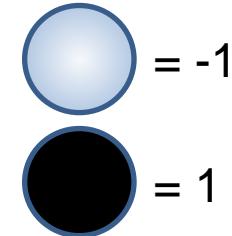
Light blue circle = -1
Black circle = 1



$$\begin{bmatrix} -1 & 1 & -1 \\ 1 & 1 & -1 \\ -1 & 1 & -1 \end{bmatrix} \times \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = -1$$

Convolutional Neural Networks

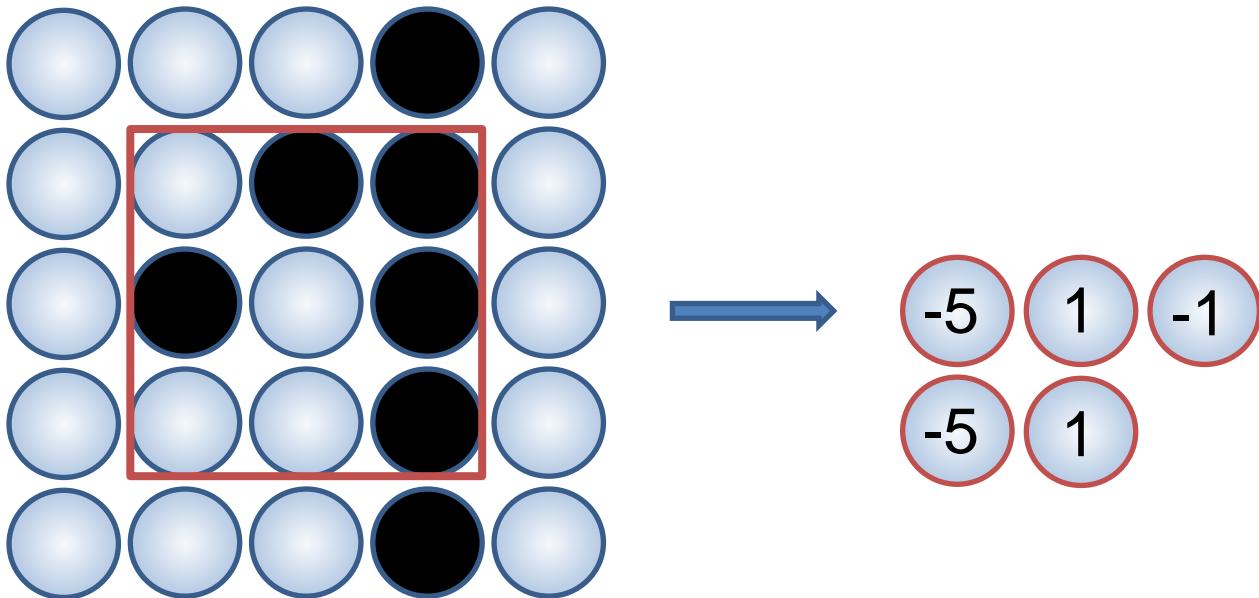
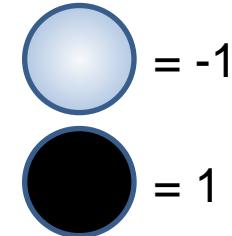
Learning on image data - Example



$$\dots \times \begin{matrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{matrix} = -5$$

Convolutional Neural Networks

Learning on image data - Example

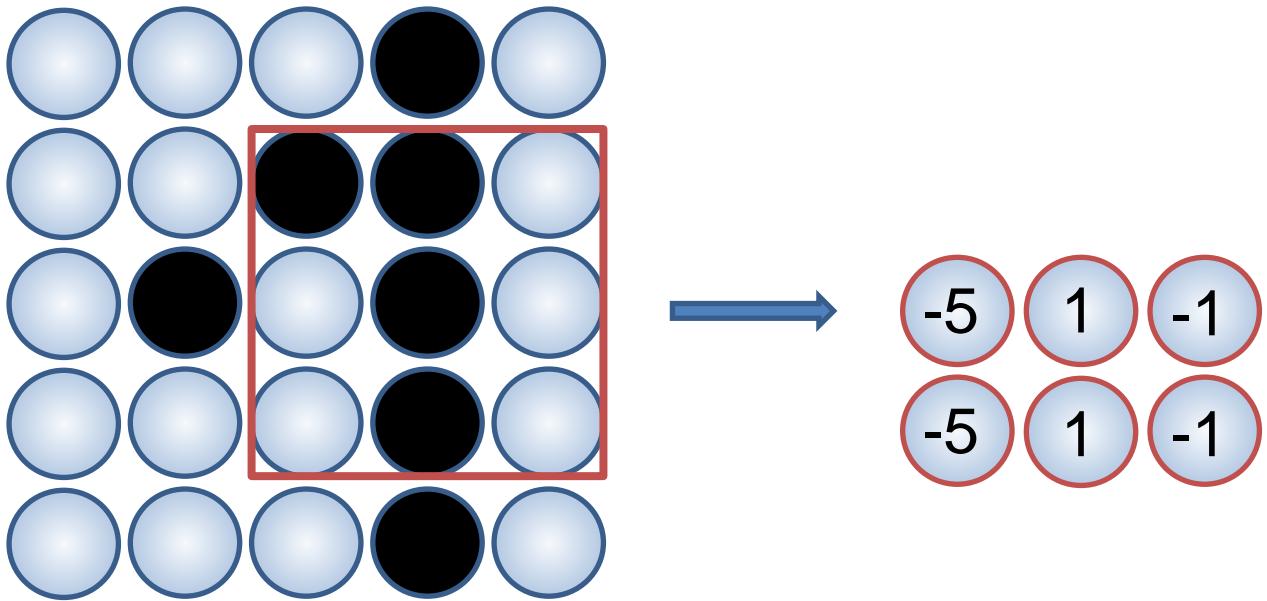


$$\dots \times \begin{matrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{matrix} = 1$$

Convolutional Neural Networks

Learning on image data - Example

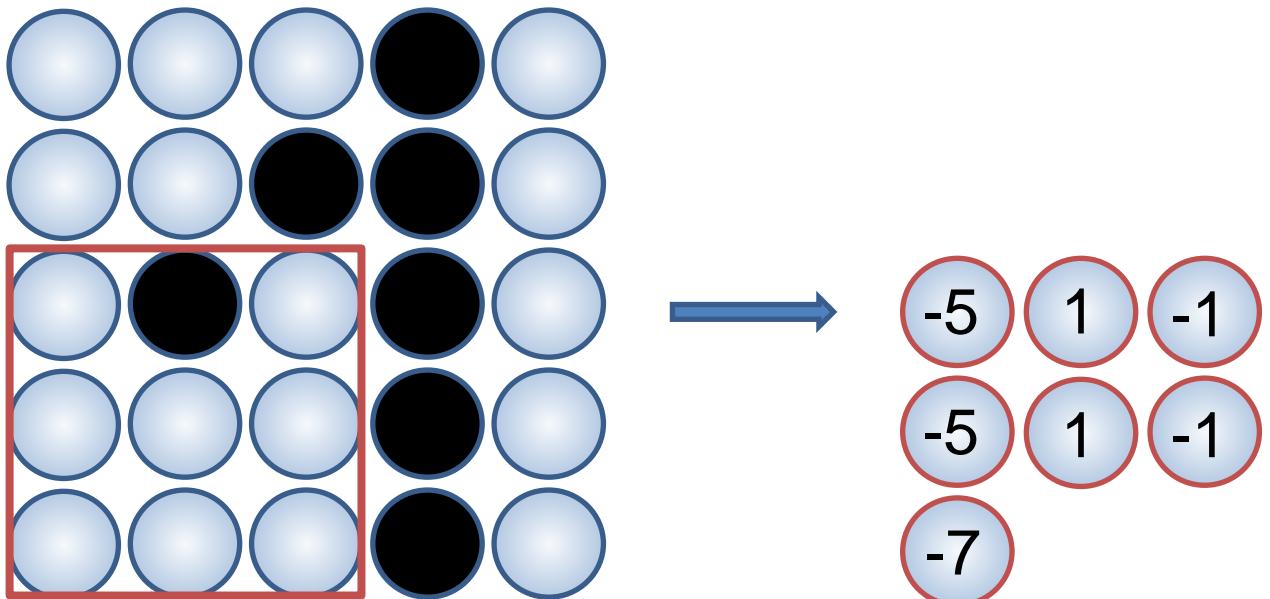
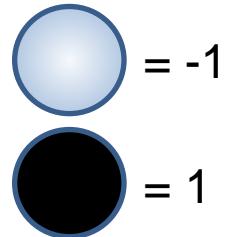
Light blue circle = -1
Black circle = 1



$$\dots \times \begin{matrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{matrix} = -1$$

Convolutional Neural Networks

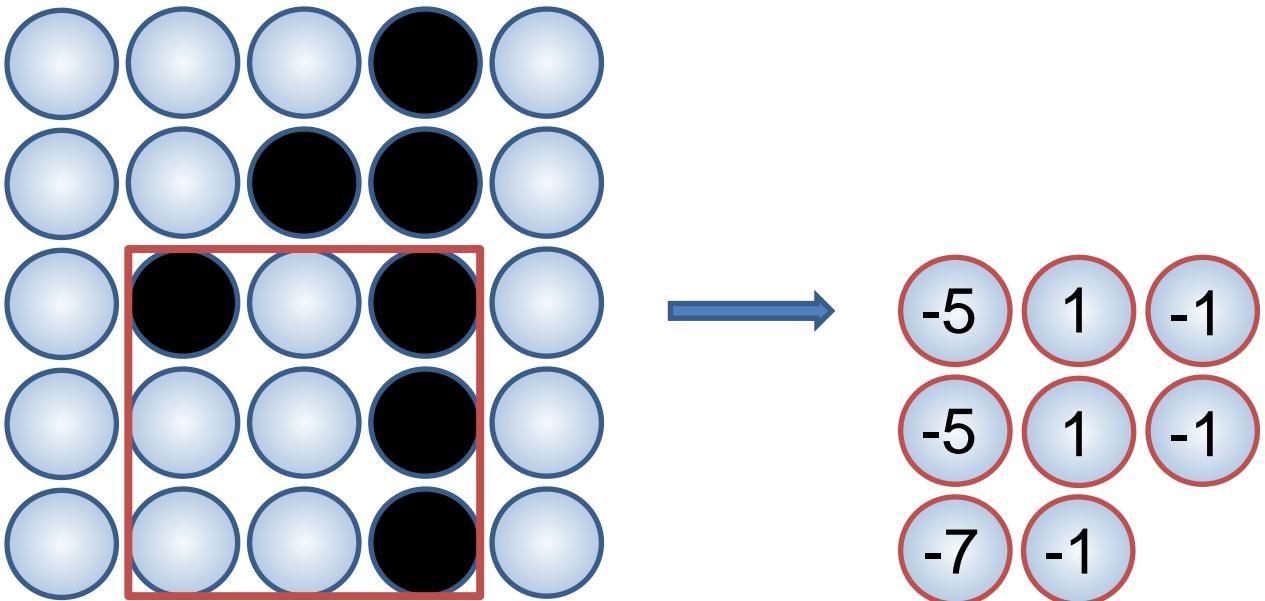
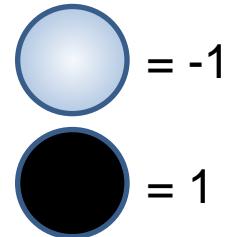
Learning on image data - Example

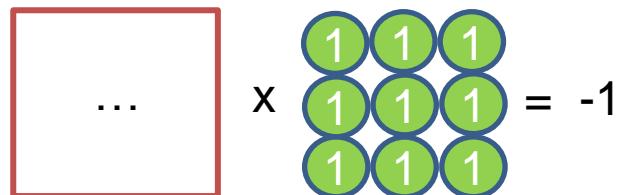


$$\dots \times \begin{matrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{matrix} = -7$$

Convolutional Neural Networks

Learning on image data - Example



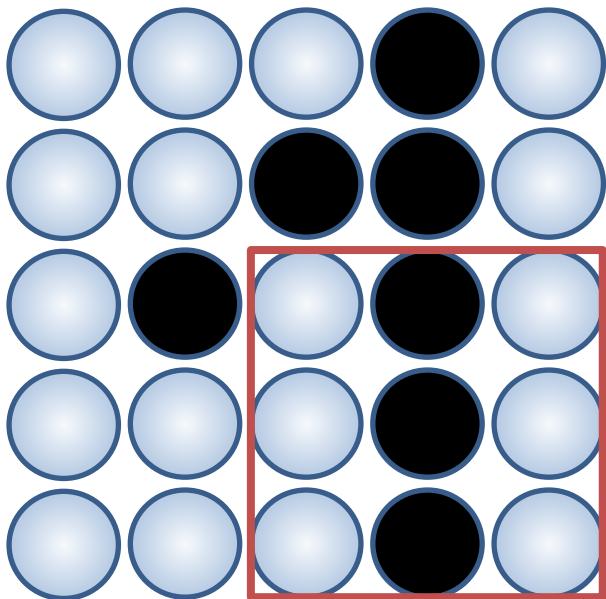


The diagram shows a red box containing three dots (...), representing the bias term, multiplied by a 3x3 kernel of all ones (1, 1, 1; 1, 1, 1; 1, 1, 1). The result is labeled = -1.

$$\dots \times \begin{matrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{matrix} = -1$$

Convolutional Neural Networks

Learning on image data - Example



- This results in a “feature map” → here : 9 neurons for the 5x5 image



$$\begin{matrix} -5 & 1 & -1 \\ -5 & 1 & -1 \\ -7 & -1 & -3 \end{matrix}$$

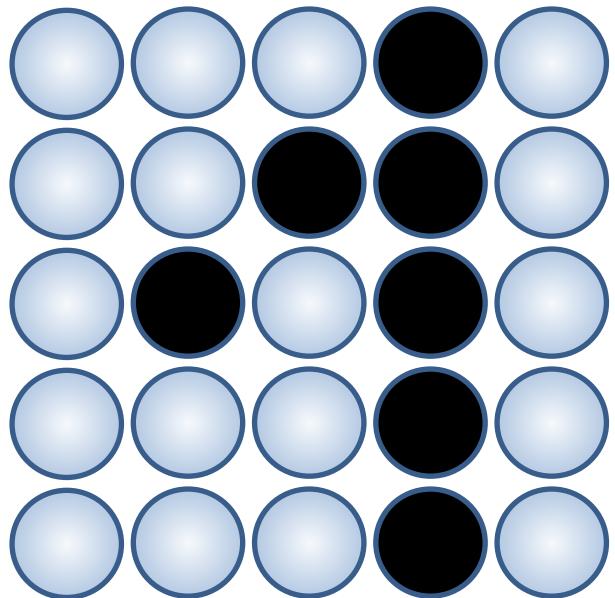
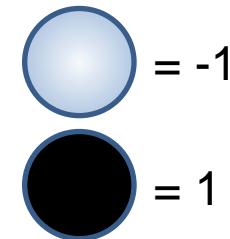
- Filter of size $3 \times 3 \rightarrow 9$ separate weights
- Scan step wise the image with the filter
- That's **Convolution**

$$\dots \times \begin{matrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{matrix} = -3$$

$$\begin{matrix} \text{light blue circle} & = -1 \\ \text{black circle} & = 1 \end{matrix}$$

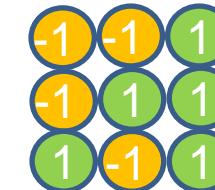
Convolutional Neural Networks

Learning – Feature Extraction



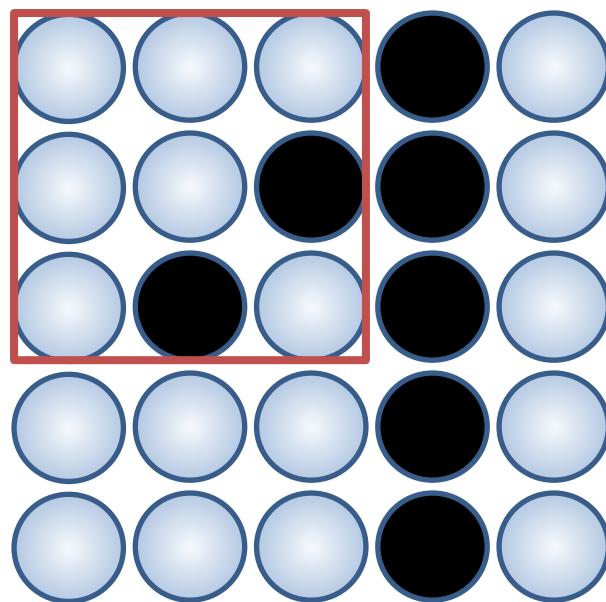
- Repeat the scan process with **multiple filters**

Next: extracting feature with filter $f_{i_1} =$



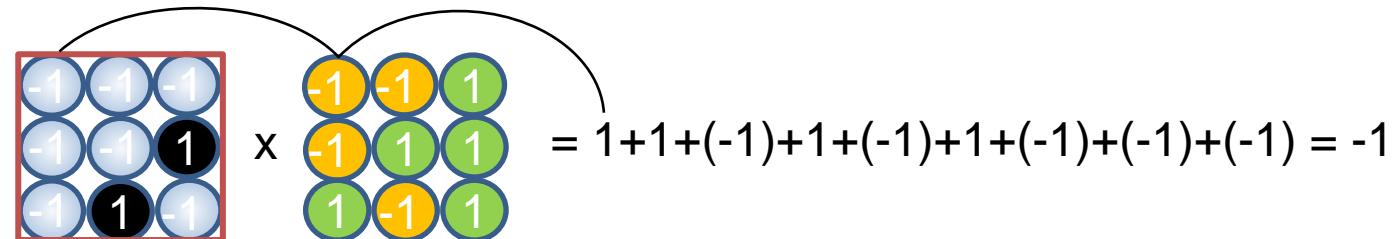
Convolutional Neural Networks

Learning – Feature Extraction



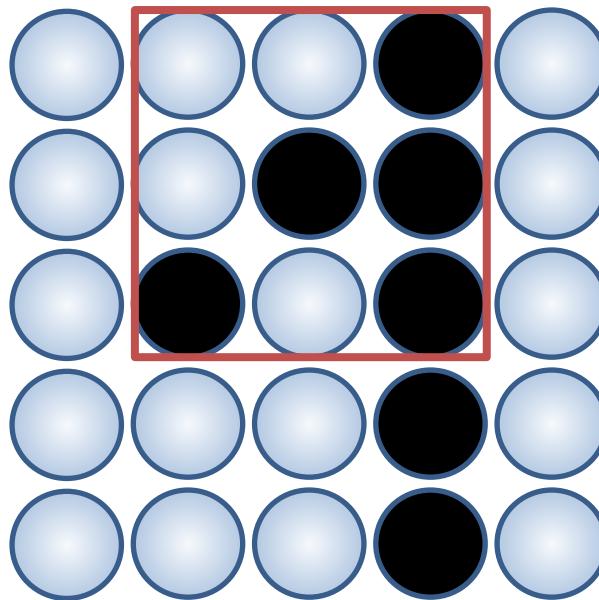
Low score !
barely matches

-1

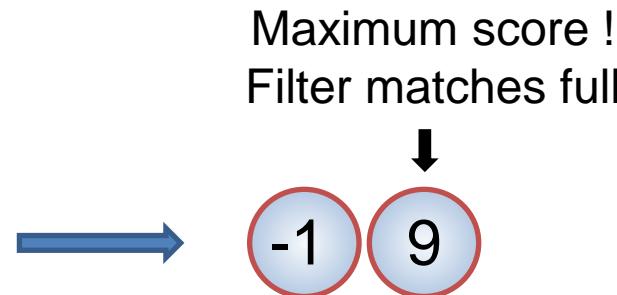

$$\begin{matrix} -1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{matrix} \times \begin{matrix} -1 & -1 & 1 \\ -1 & 1 & 1 \\ 1 & -1 & 1 \end{matrix} = 1+1+(-1)+1+(-1)+1+(-1)+(-1) = -1$$

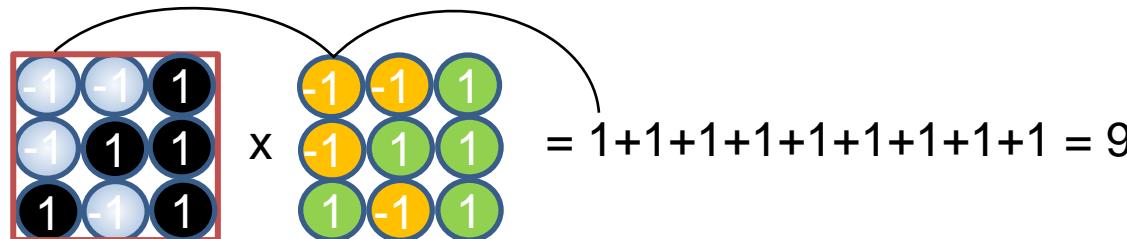
Convolutional Neural Networks

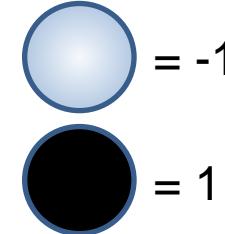
Learning – Feature Extraction



Maximum score !
Filter matches full



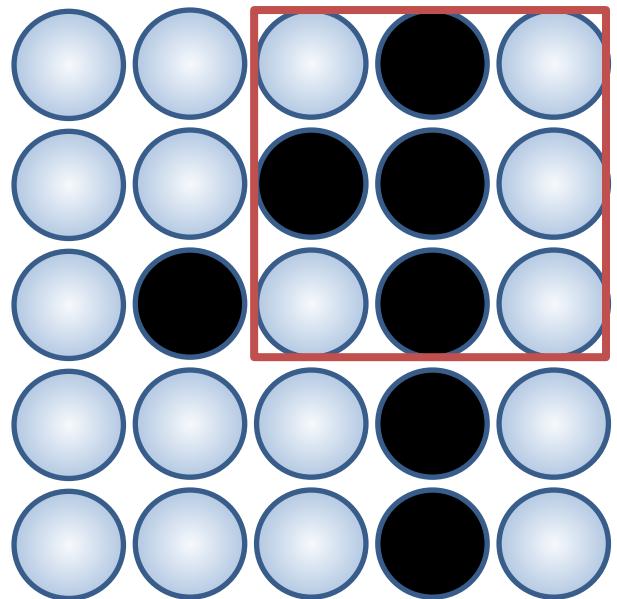
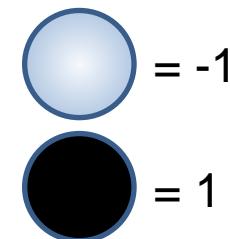

$$\begin{bmatrix} -1 & -1 & 1 \\ -1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix} \times \begin{bmatrix} -1 & -1 & 1 \\ -1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix} = 1+1+1+1+1+1+1+1 = 9$$



 = -1
 = 1

Convolutional Neural Networks

Learning – Feature Extraction



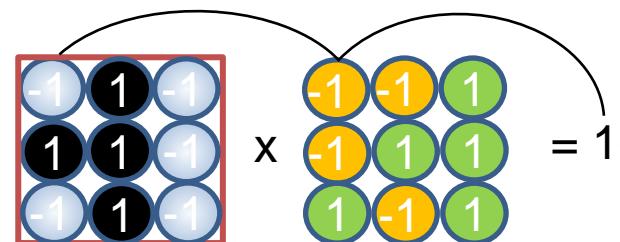
Very low score !
Many mismatches



A blue arrow points from the input image to the output scores. Below the arrow, the text "Very low score ! Many mismatches" is displayed, indicating that the current feature extraction is failing to correctly identify the black circle.



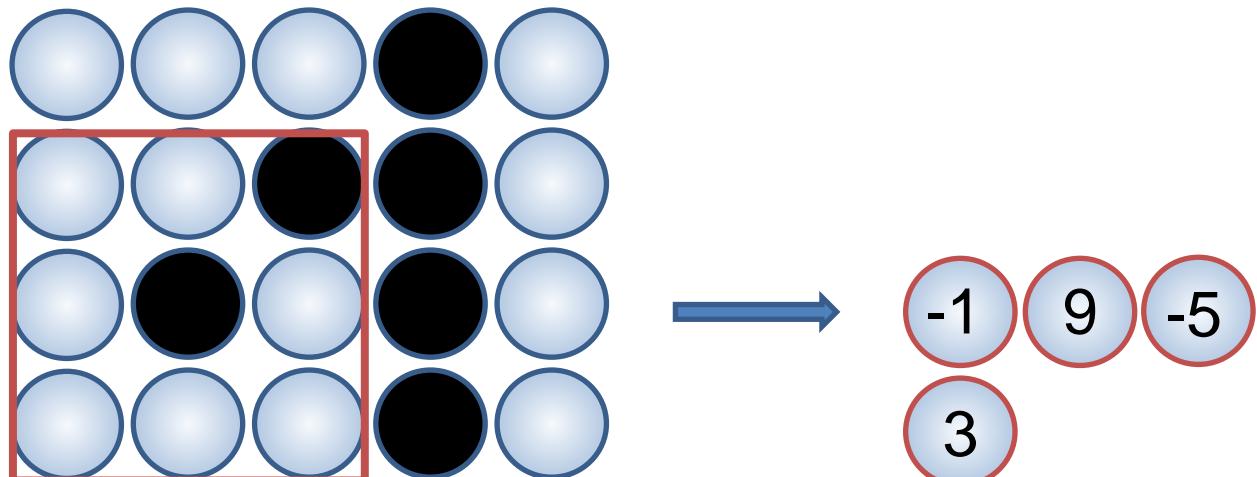
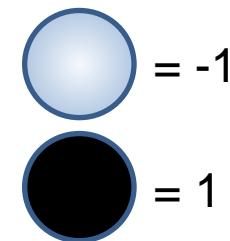
The output consists of three numerical values: -1, 9, and -5, each enclosed in a red-bordered circle. The value 9 is highlighted in blue, suggesting it is the most significant or correct response despite the overall low score.



A diagram illustrating a convolution step. On the left, a 3x3 input patch is shown with values: -1, 1, -1; 1, 1, -1; -1, 1, -1. This patch is multiplied (indicated by an 'x') with a 3x3 kernel on the right, which has values: -1, -1, 1; -1, 1, 1; 1, -1, 1. The result of the multiplication is shown as a sum of products: $= 1 + (-1) + (-1) + (-1) + 1 + (-1) + (-1) + (-1) = -5$.

Convolutional Neural Networks

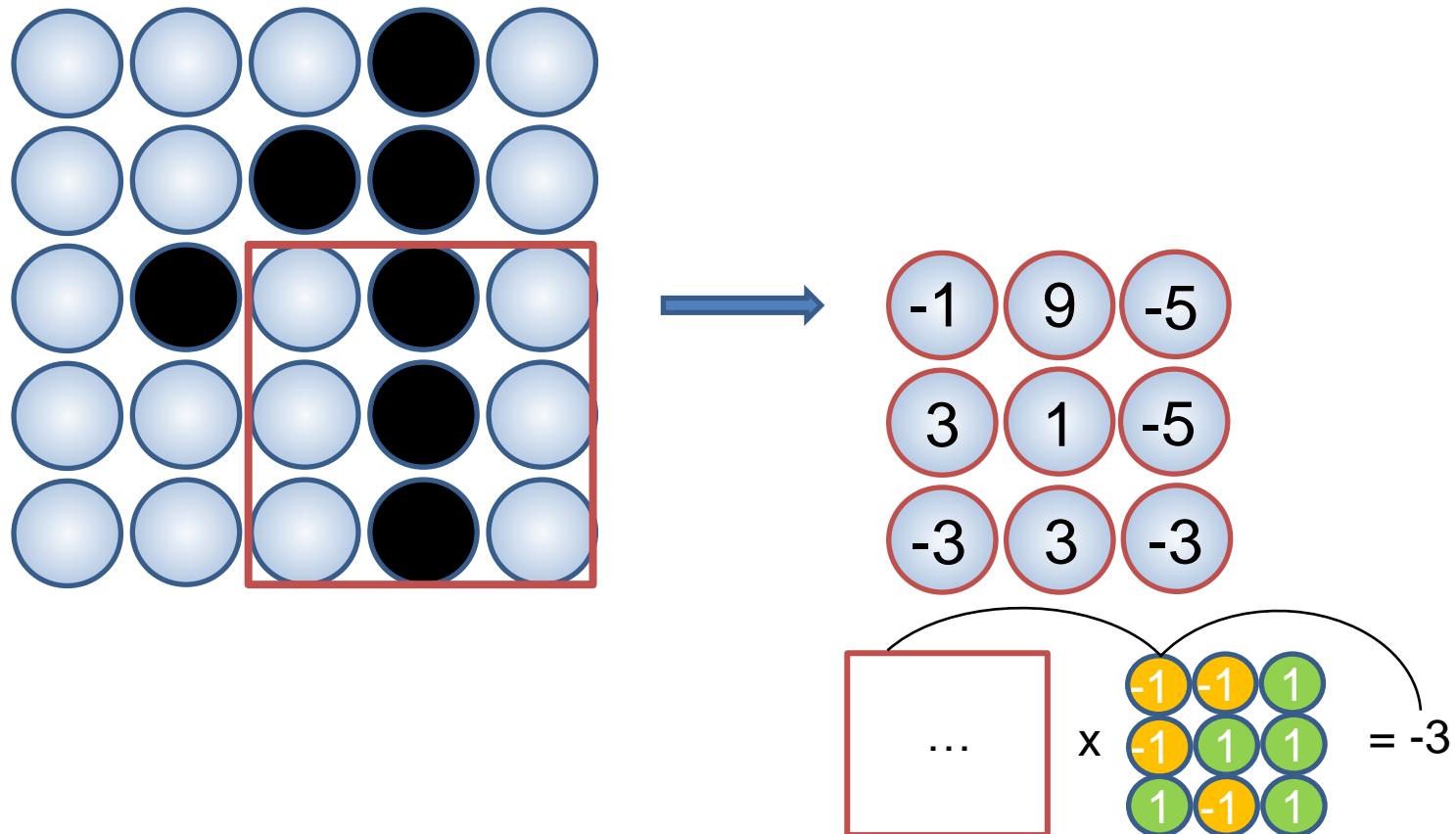
Learning – Feature Extraction



A diagram illustrating a convolution operation. A red-bordered input patch containing three dots (...), labeled '...', is multiplied by a 3x3 kernel matrix. The kernel matrix has alternating yellow and green values: -1, -1, 1; -1, 1, 1; 1, -1, 1. The result of the multiplication is 3.

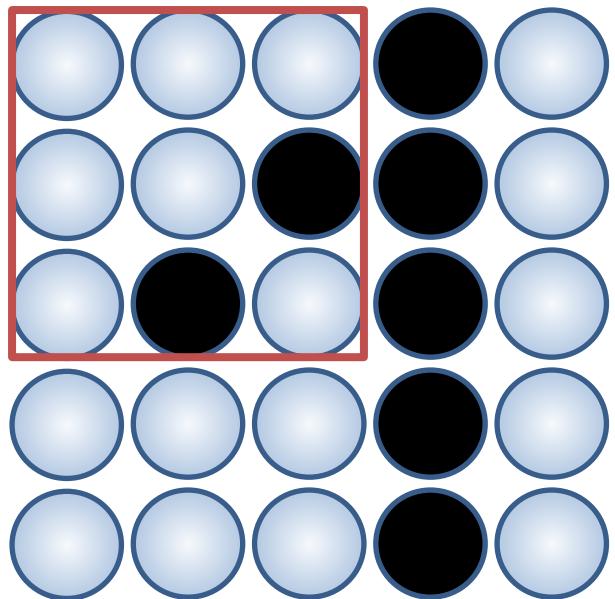
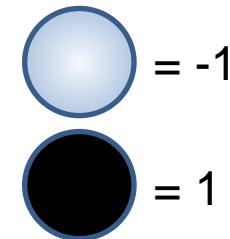
Convolutional Neural Networks

Learning – Feature Extraction



Convolutional Neural Networks

Learning – Feature Extraction

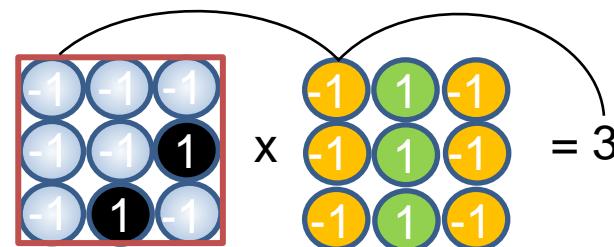


- Different feature will extract different features
- **Spatially share** parameters of each filter



Next filter $\mathbf{f}_i =$

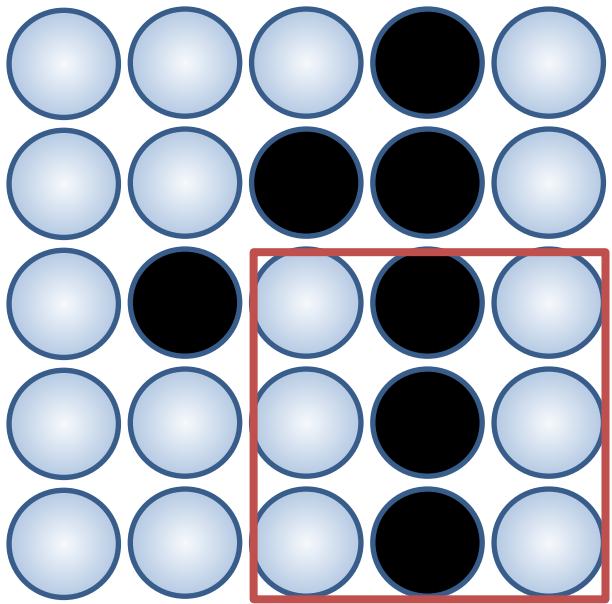
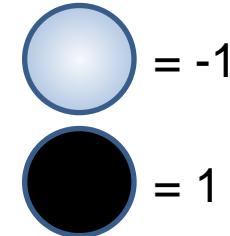




A diagram illustrating the convolution operation. It shows the 3x3 input subgrid with a red border and the 3x3 filter kernel with a green center. The equation $\begin{bmatrix} -1 & -1 & -1 \\ -1 & -1 & 1 \\ -1 & 1 & -1 \end{bmatrix} \times \begin{bmatrix} -1 & 1 & -1 \\ -1 & 1 & -1 \\ -1 & 1 & -1 \end{bmatrix} = 3$ represents the calculation of the output value 3. The filter is applied to the highlighted input subgrid, with the result being the sum of the products of corresponding elements from the input and filter matrices.

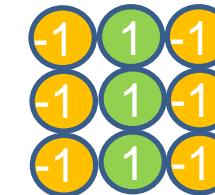
Convolutional Neural Networks

Learning – Feature Extraction



- Different feature will extract different features
- **Spatially share** parameters of each filter

Next filter $\mathbf{f}_2 =$



A diagram illustrating the convolution operation. An input layer (3x3 grid) is multiplied by a filter \mathbf{f}_2 (3x3 kernel) to produce an output layer (3x3 grid). The result is 9.

Input layer (3x3 grid):

-1	1	-1
-1	1	-1
-1	1	-1

Filter \mathbf{f}_2 (3x3 kernel):

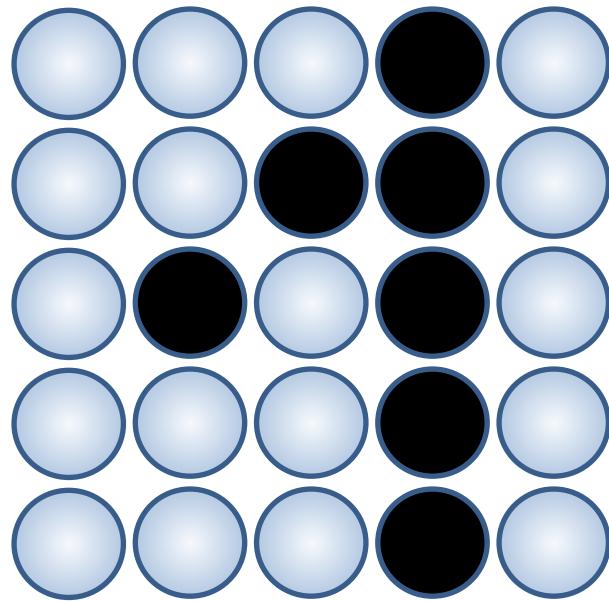
-1	1	-1
-1	1	-1
-1	1	-1

Output layer (3x3 grid):

3	-3	7
3	-3	7
5	-5	9

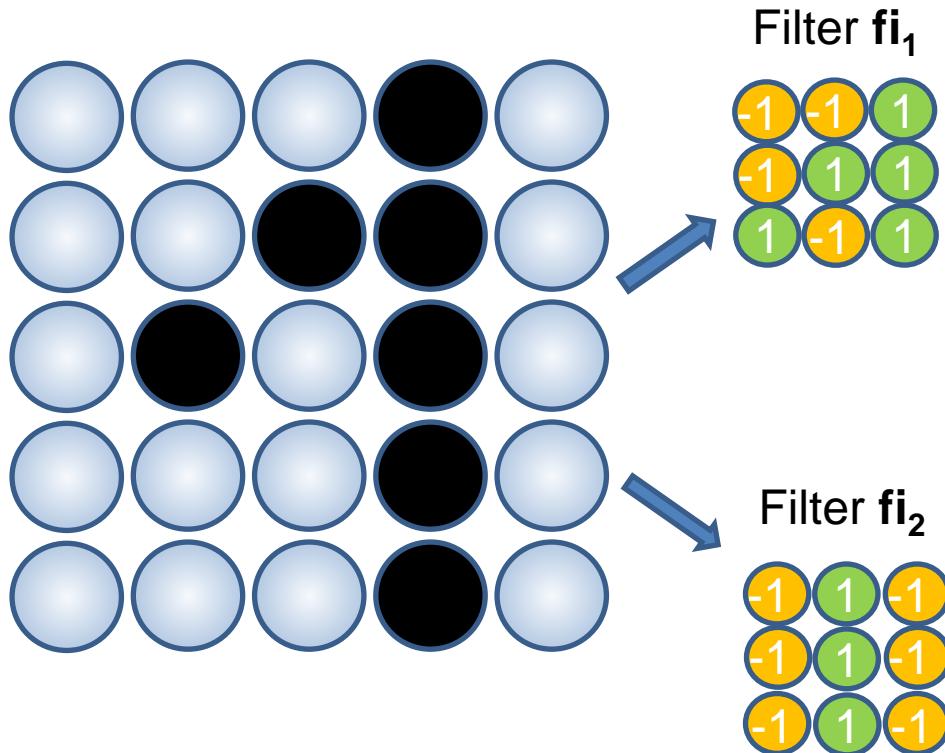
Convolutional Neural Networks

Learning – Feature Extraction: filter comparison



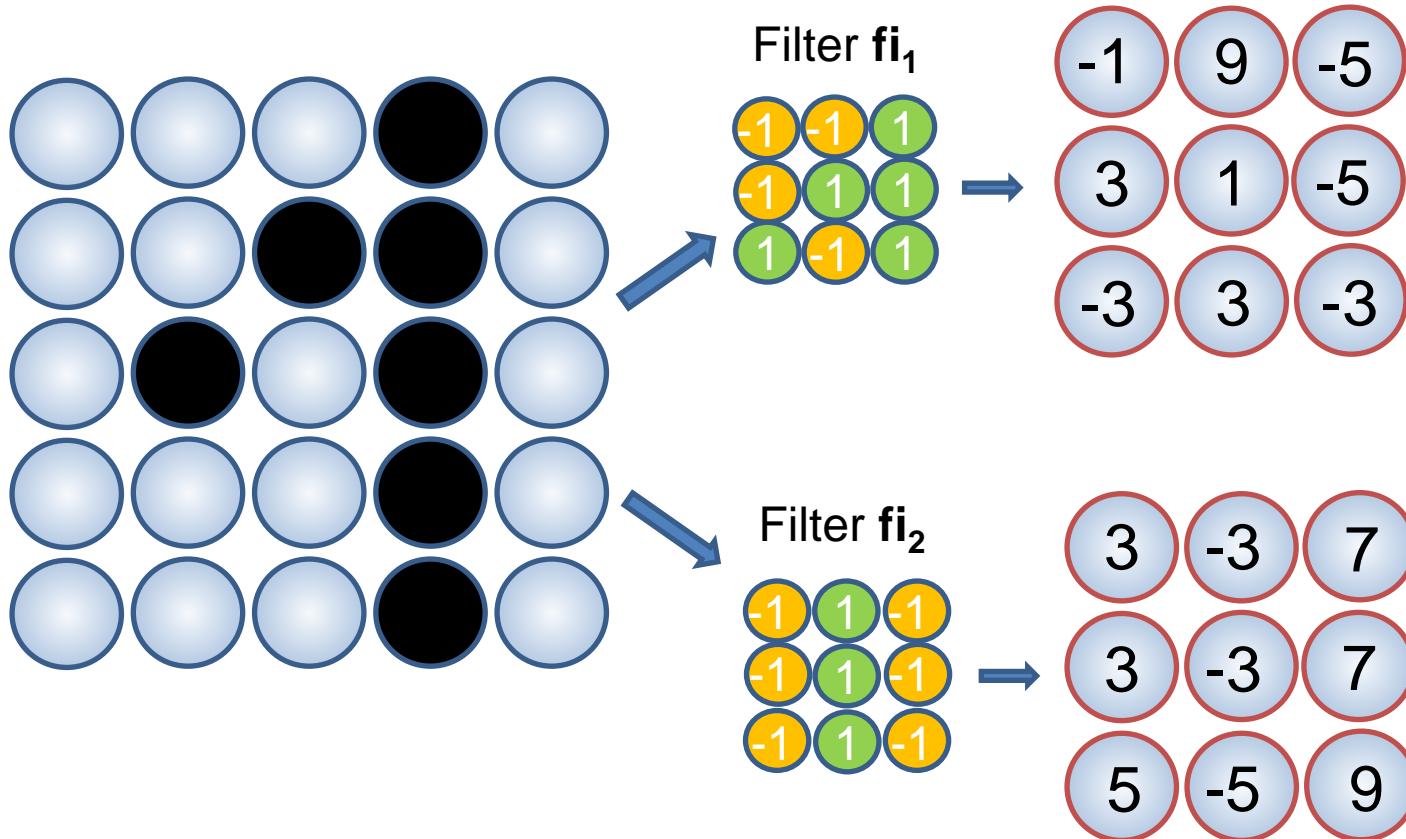
Convolutional Neural Networks

Learning – Feature Extraction: filter comparison



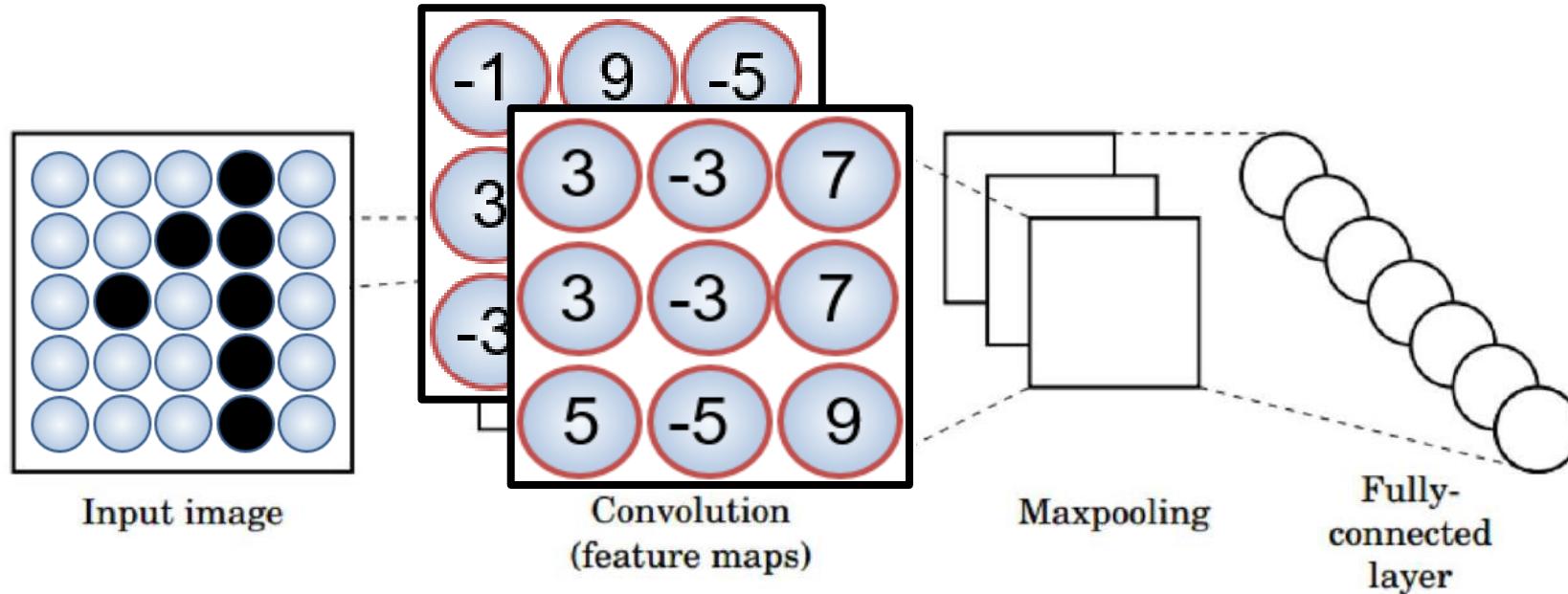
Convolutional Neural Networks

Learning – Feature Extraction: filter comparison



Convolutional Neural Networks

Building a CNN - Example

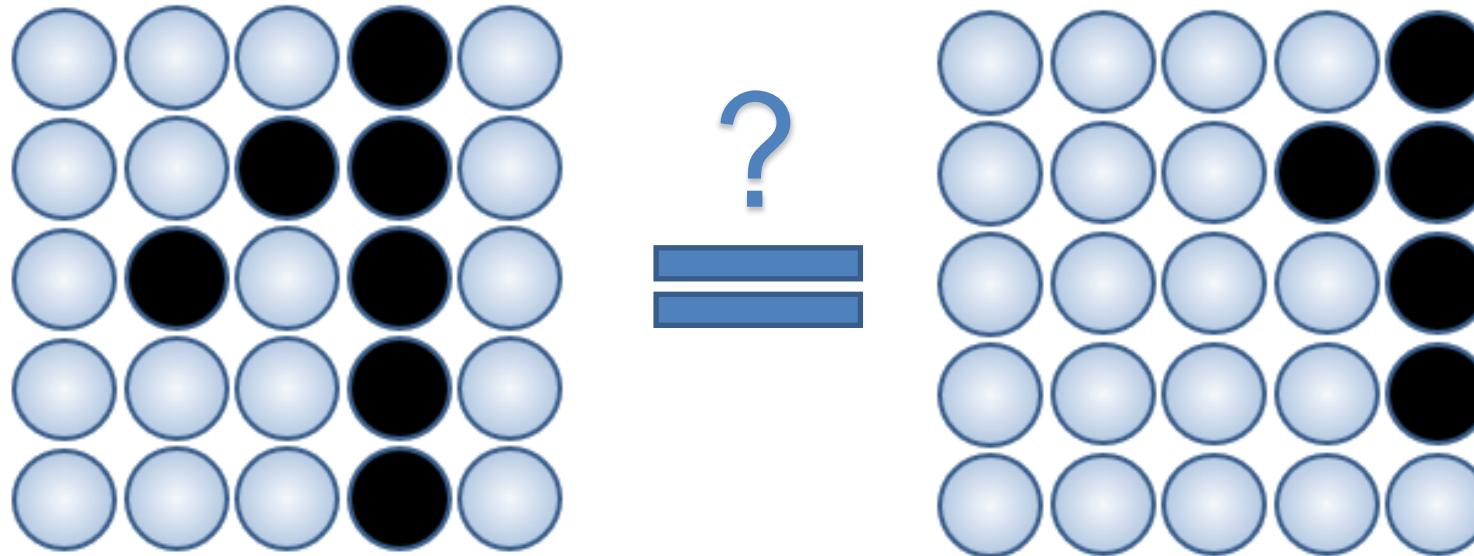


- 1. Convolution:** Apply filters with learned weights to generate feature maps.
- 2. Non-linearity:** Often ReLU.
- 3. Pooling:** Downsampling operation on each feature map.

Train model with image data.
Learn weights of filters in convolutional layers.

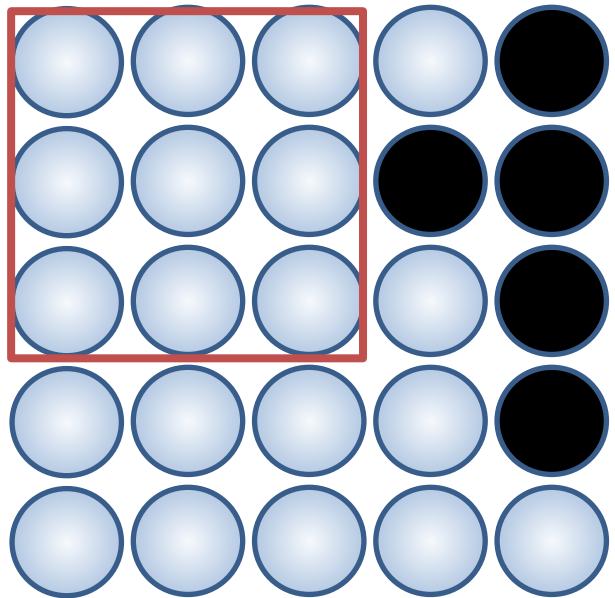
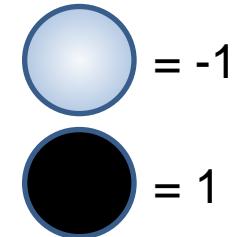
Convolutional Neural Networks

Filters to detect features



Convolutional Neural Networks

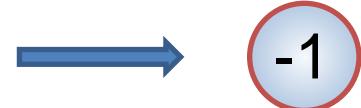
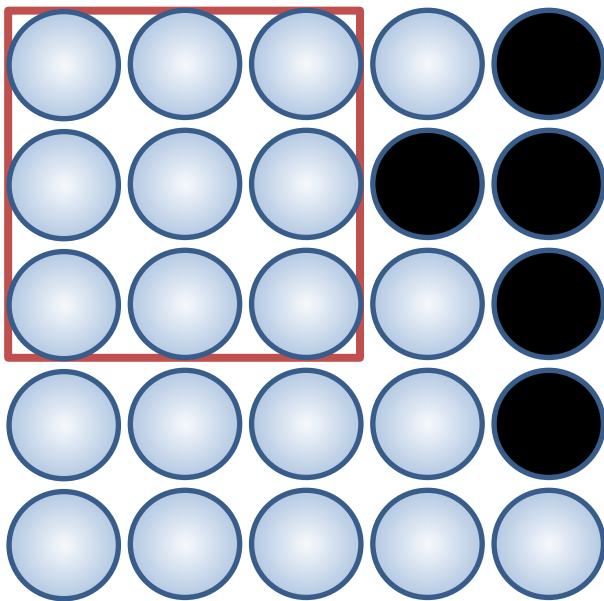
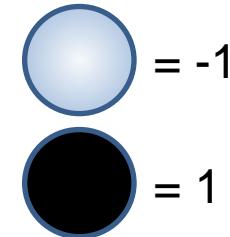
Filters to detect features



The usage of filters in detecting features allow good identification even in the event of **displacement, shrinkage, rotation or deformation**.

Convolutional Neural Networks

Filters to detect features



The usage of filters in detecting features allow good identification even in the event of **displacement, shrinkage, rotation or deformation**.

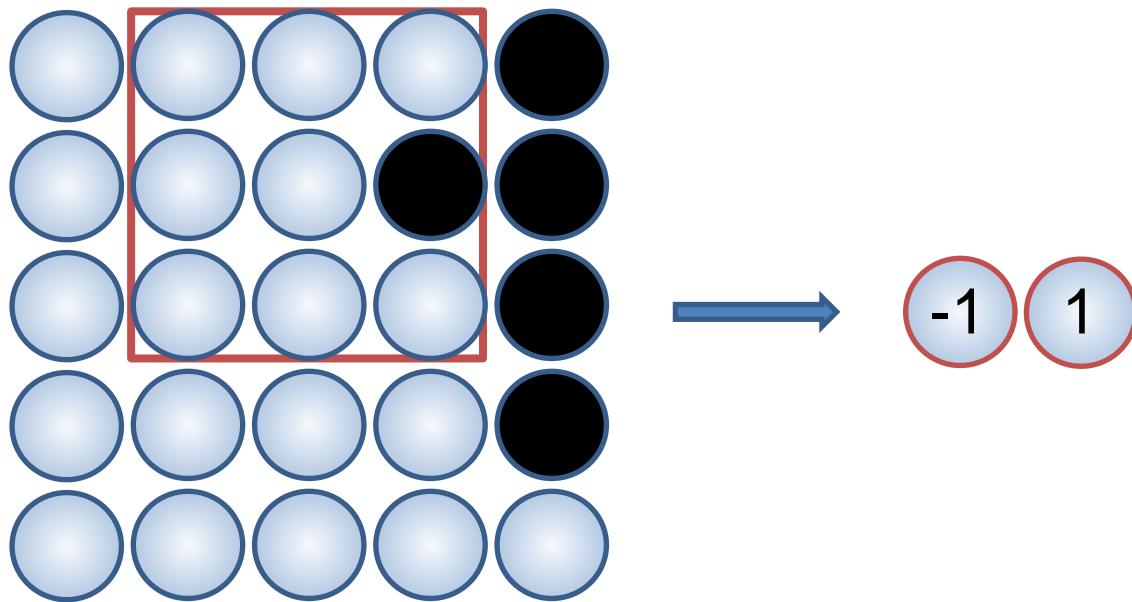
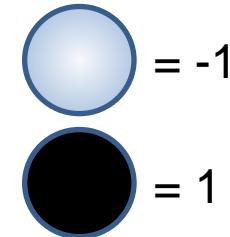
e.g. Filter $f_{i_1} =$

$$\begin{matrix} -1 & -1 & 1 \\ -1 & 1 & 1 \\ 1 & -1 & 1 \end{matrix}$$

$$\begin{matrix} -1 & -1 & -1 \\ -1 & -1 & -1 \\ -1 & -1 & -1 \end{matrix} \times \begin{matrix} -1 & -1 & 1 \\ -1 & 1 & 1 \\ 1 & -1 & 1 \end{matrix} = \dots$$

Convolutional Neural Networks

Filters to detect features



e.g. Filter $\mathbf{f}_1 =$

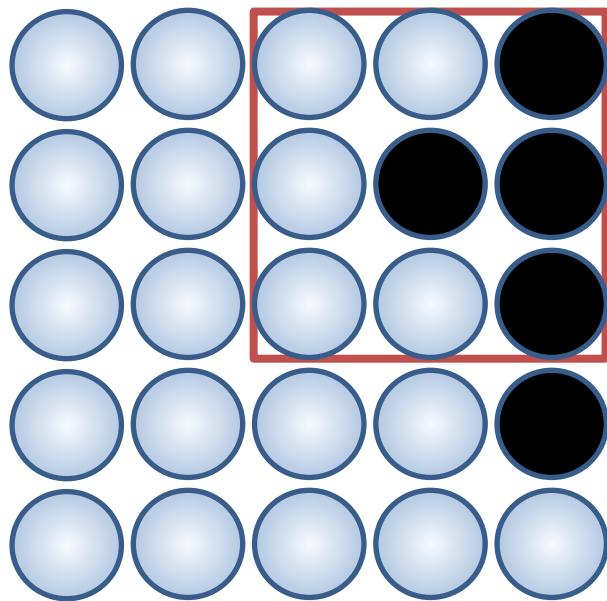


Feature/filter \mathbf{f}_1 still matches the deformed image well

$$\begin{bmatrix} -1 & -1 & -1 \\ -1 & \mathbf{-1} & -1 \\ -1 & -1 & -1 \end{bmatrix} \times \begin{bmatrix} -1 & -1 & 1 \\ -1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix} = \dots$$

Convolutional Neural Networks

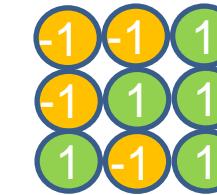
Filters to detect features



Not a full match
but still a high score



e.g. Filter $f_{i_1} =$



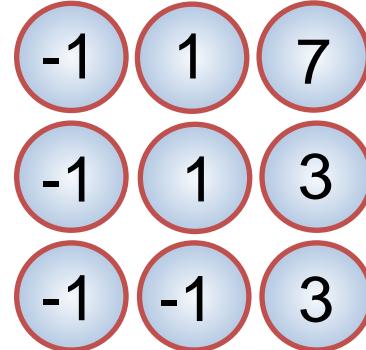
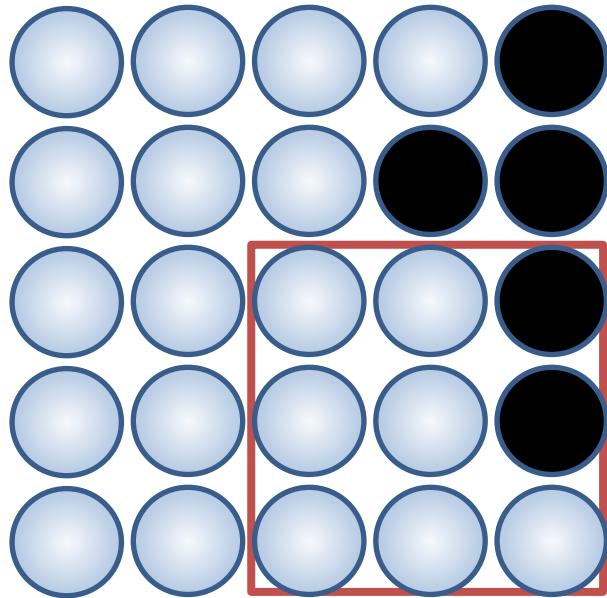
Feature/filter f_{i_1} still matches the
deformed image well
Remember! **A full match would be 9**

The diagram illustrates the convolution operation. On the left, a 3x3 subgrid of the input image is shown with a red border, containing the values: -1, -1, 1; -1, 1, 1; -1, -1, 1. To its right is a multiplication sign (\times). To the right of the multiplication sign is a 3x3 filter matrix with colored circles: top row is orange, middle row is green, bottom row is green. Below the multiplication sign is an equals sign (=). To the right of the equals sign is the result of the dot product: $1+1+1+1+1+(-1)+1+1 = 7$.

Convolutional Neural Networks

Filters to detect features

$$\begin{array}{c} \text{Light Blue Circle} \\ = -1 \\ \text{Black Circle} \\ = 1 \end{array}$$



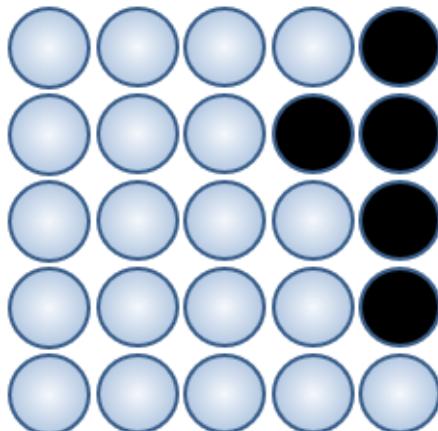
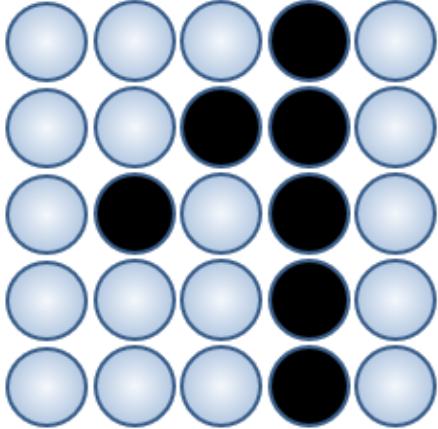
e.g. Filter $f_{i_1} =$



Finished feature map
using filter f_{i_1}

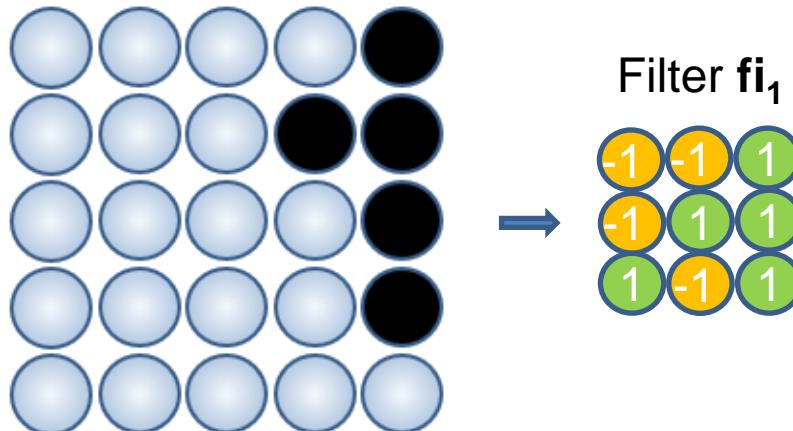
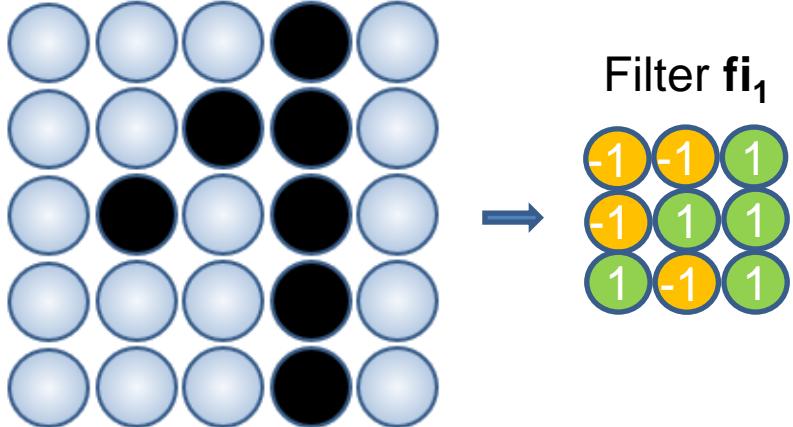
Convolutional Neural Networks

Filters to detect features: comparison



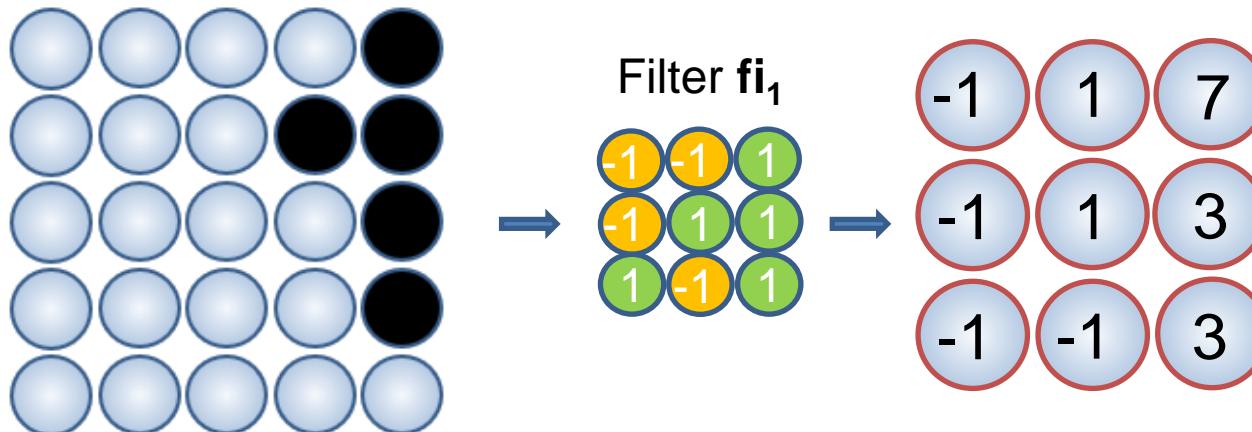
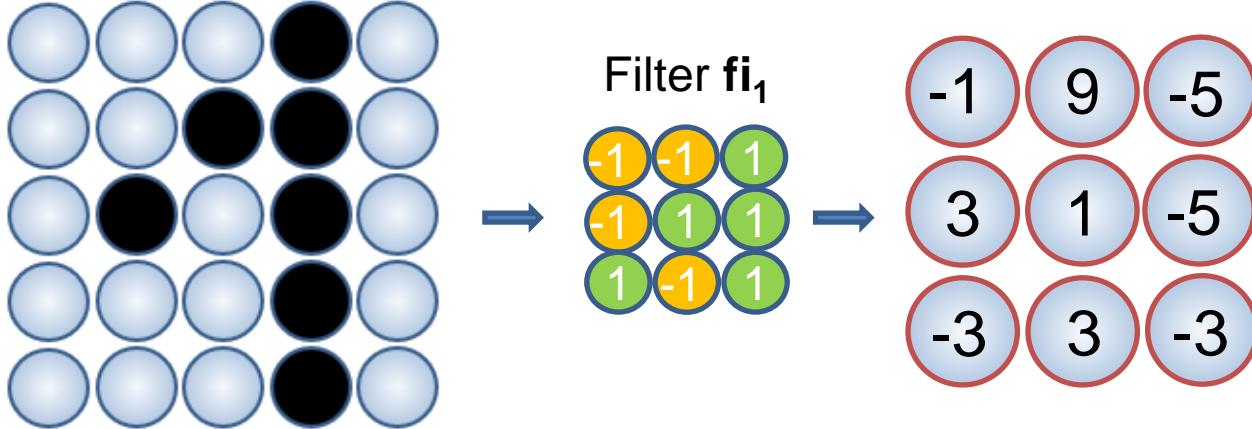
Convolutional Neural Networks

Filters to detect features: comparison



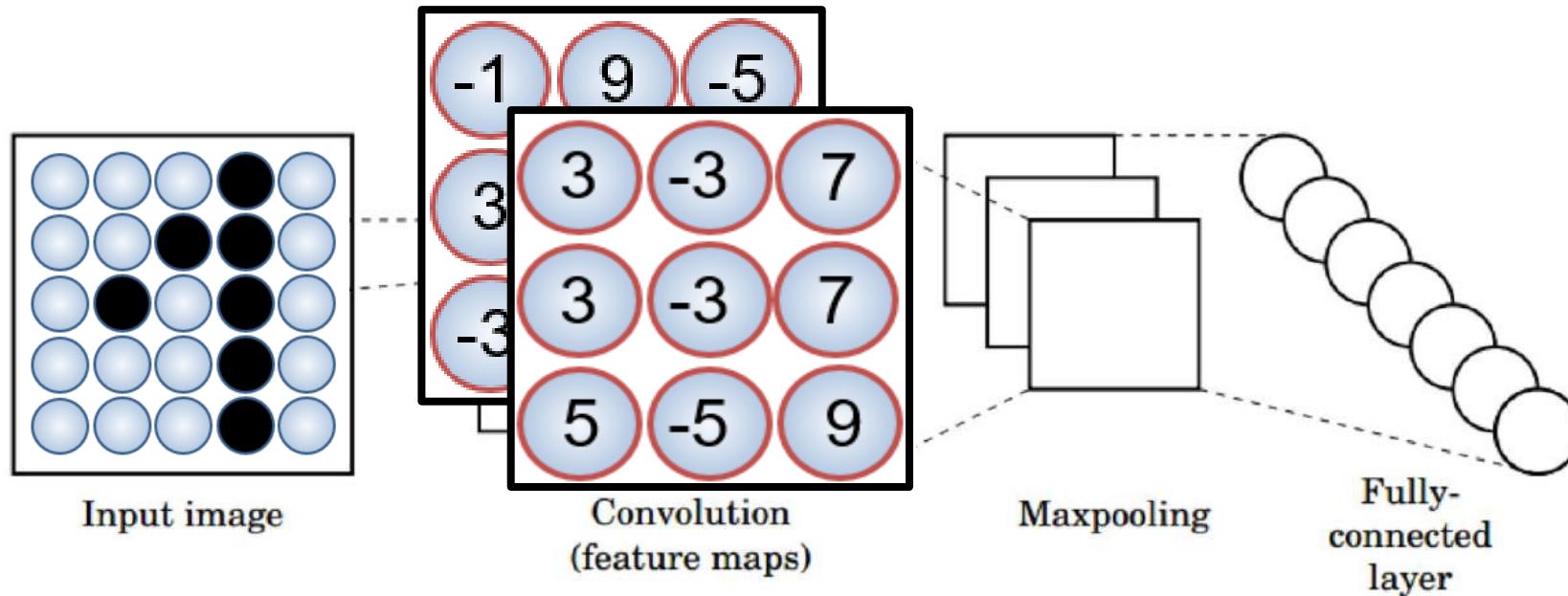
Convolutional Neural Networks

Filters to detect features: comparison



Convolutional Neural Networks

Pooling



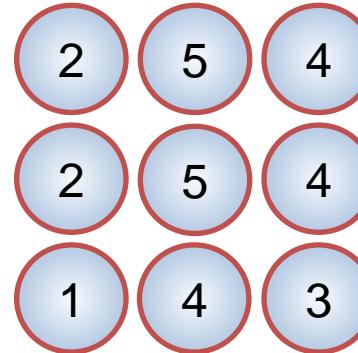
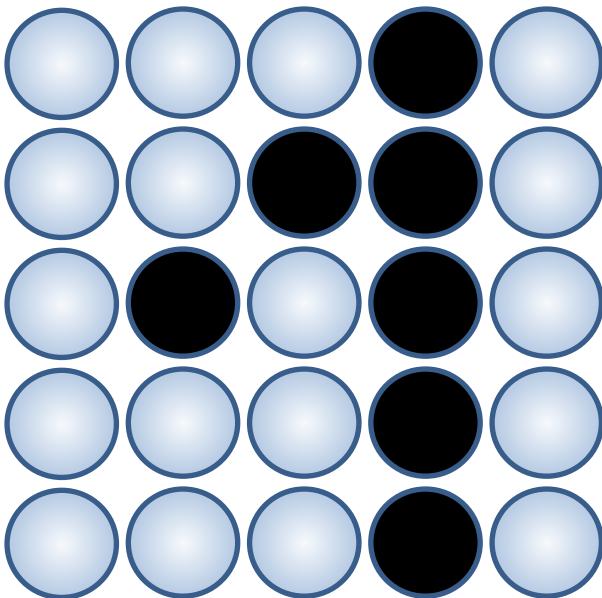
- 1. Convolution:** Apply filters with learned weights to generate feature maps.
- 2. Non-linearity:** Often ReLU.
- 3. Pooling:** Downsampling operation on each feature map.

**Train model with image data.
Learn weights of filters in convolutional layers.**

Convolutional Neural Networks

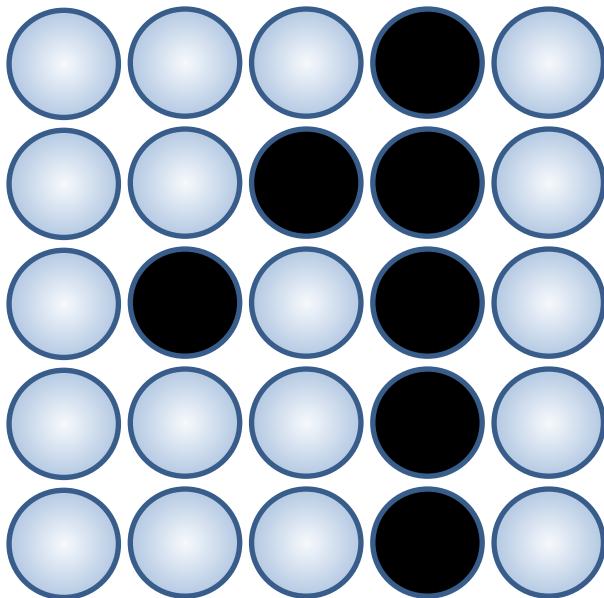
Max Pool – Example

- We already have the layer with the 9 neurons for the 5×5 image

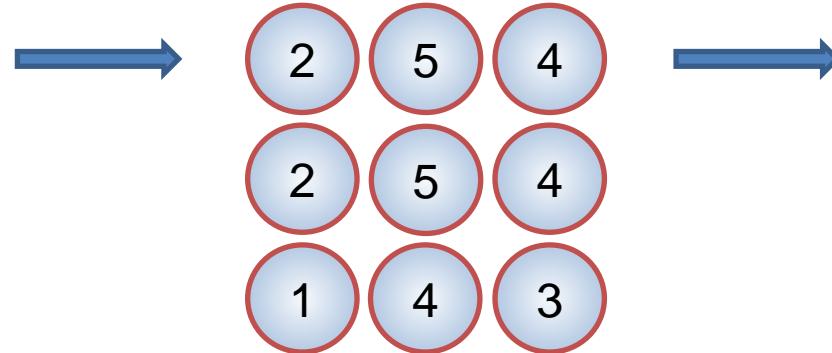


Convolutional Neural Networks

Max Pool – Example

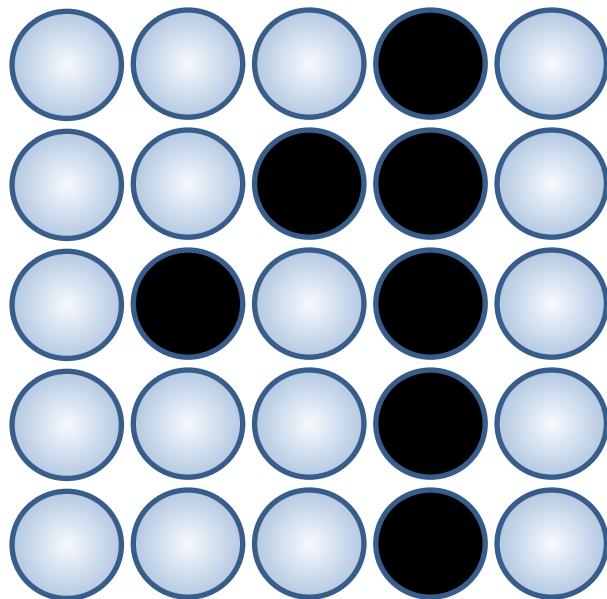


- We already have the layer with the 9 neurons for the 5x5 image
- Now we want to reduce dimensionality and spatial invariance

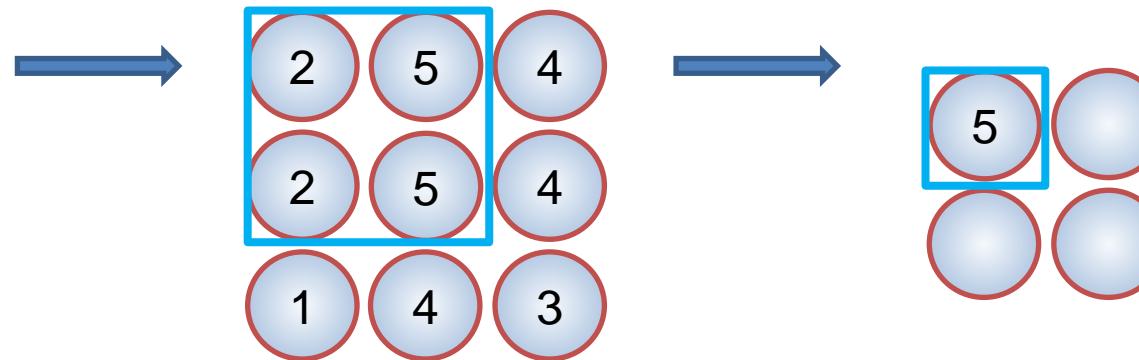


Convolutional Neural Networks

Max Pool – Example

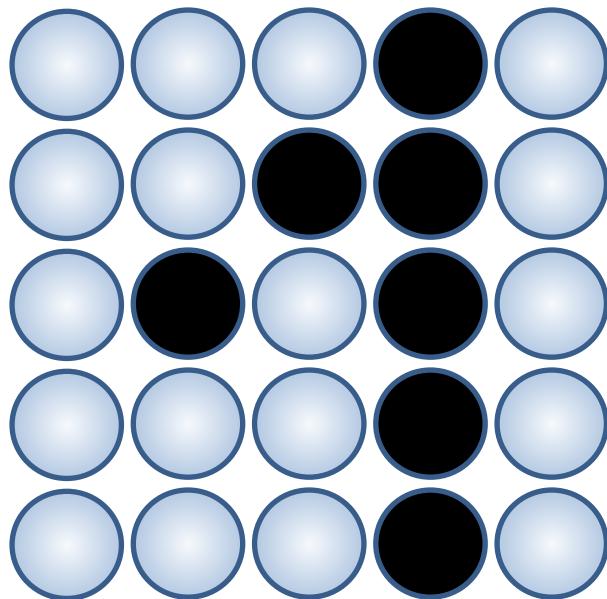


- We already have the layer with the 9 neurons for the 5x5 image
- Now we want to reduce dimensionality and spatial invariance
- Using max pool (find max with 2x2 filters and stride 1)

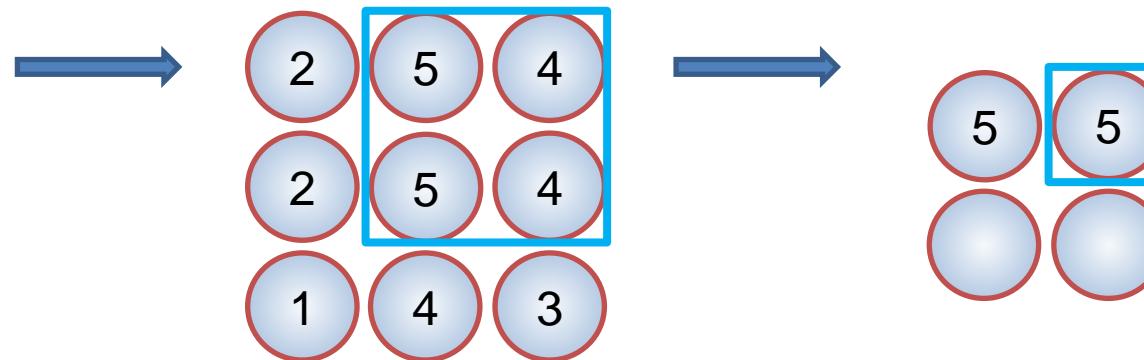


Convolutional Neural Networks

Max Pool – Example

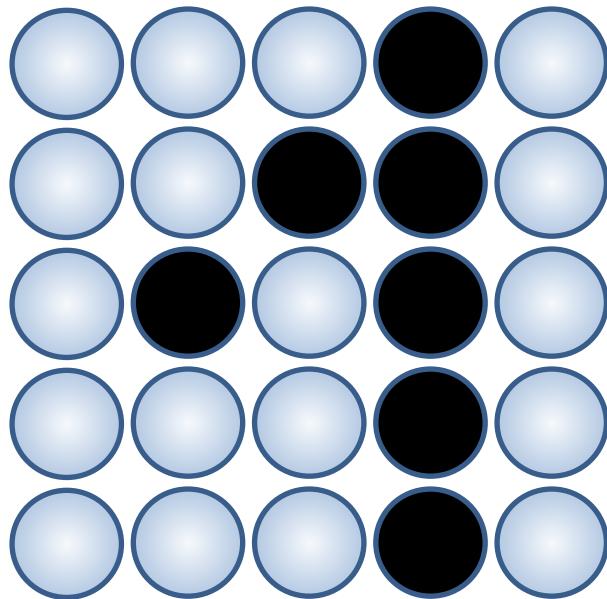


- We already have the layer with the 9 neurons for the 5x5 image
- Now we want to reduce dimensionality and spatial invariance
- Using max pool (find max with 2x2 filters and stride 1)

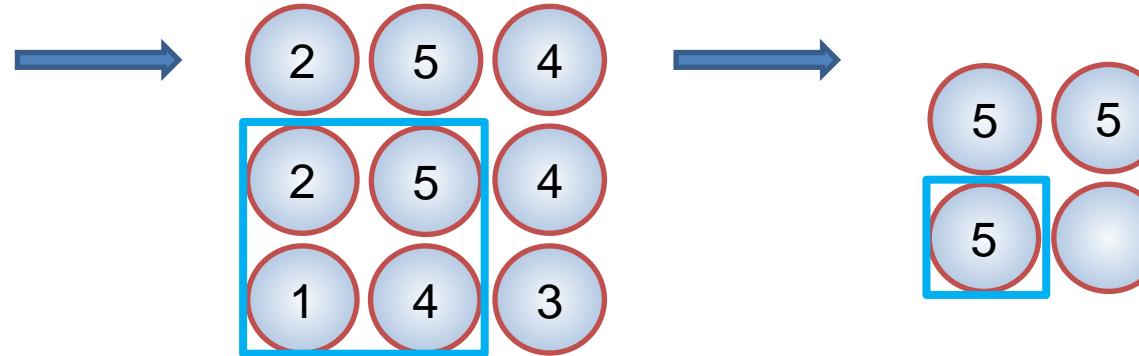


Convolutional Neural Networks

Max Pool – Example

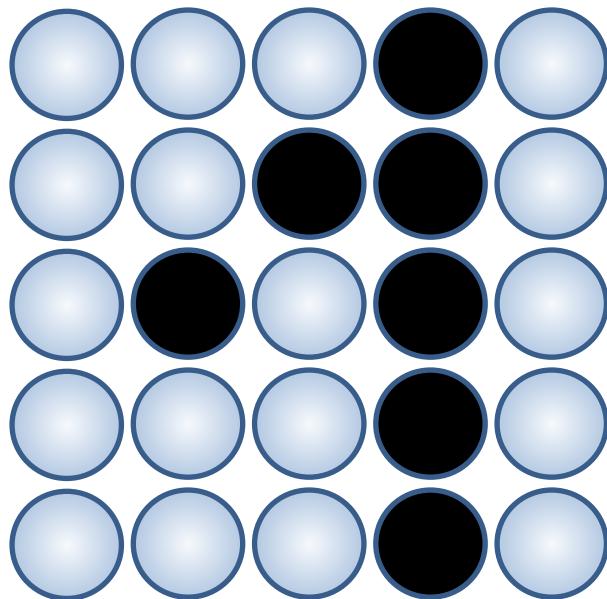


- We already have the layer with the 9 neurons for the 5x5 image
- Now we want to reduce dimensionality and spatial invariance
- Using max pool (find max with 2x2 filters and stride 1)

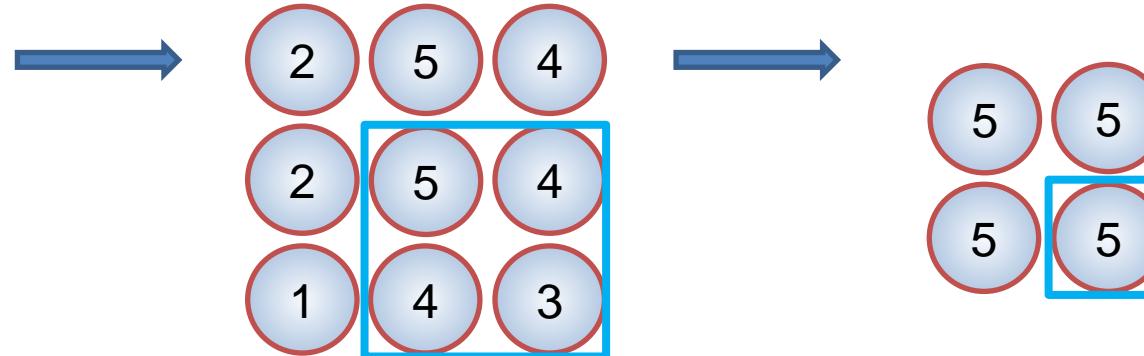


Convolutional Neural Networks

Max Pool – Example

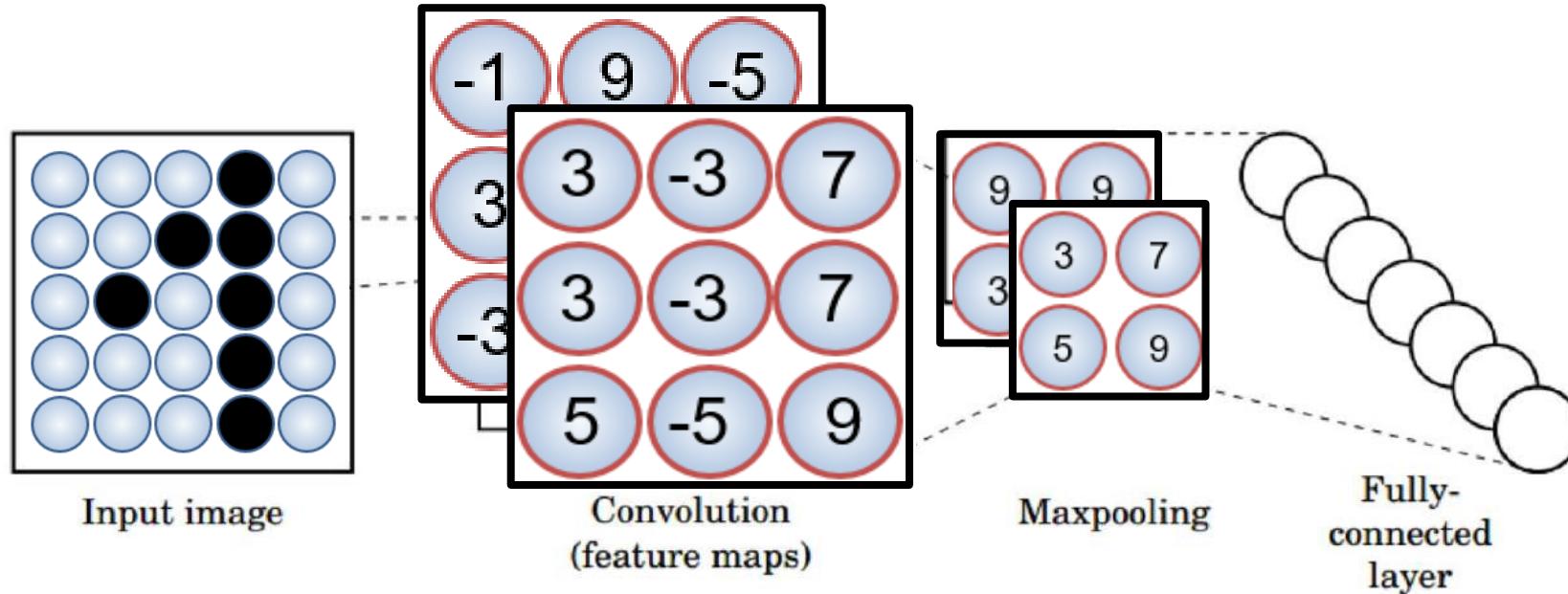


- We already have the layer with the 9 neurons for the 5×5 image
- Now we want to reduce dimensionality and spatial invariance
- Using max pool (find max with 2×2 filters and stride 1)



Convolutional Neural Networks

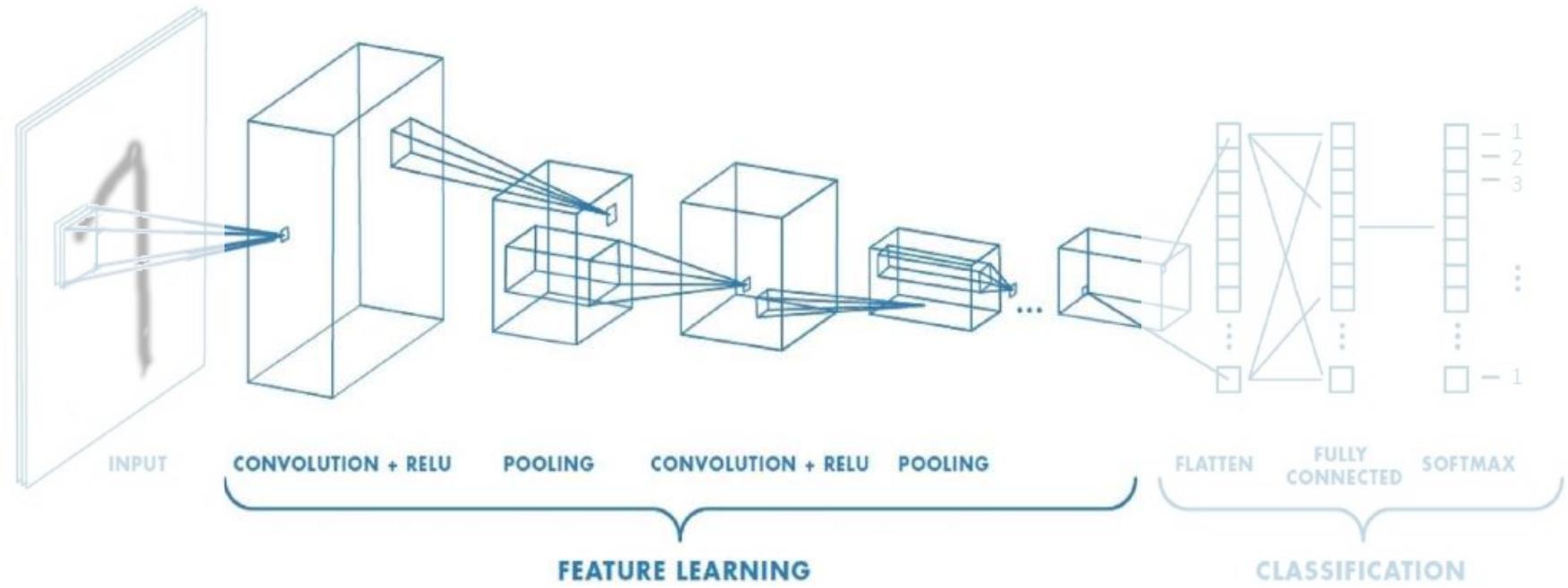
Building a CNN – Example



- 1. Convolution:** Apply filters with learned weights to generate feature maps.
- 2. Non-linearity:** Often ReLU.
- 3. Pooling:** Downsampling operation on each feature map.

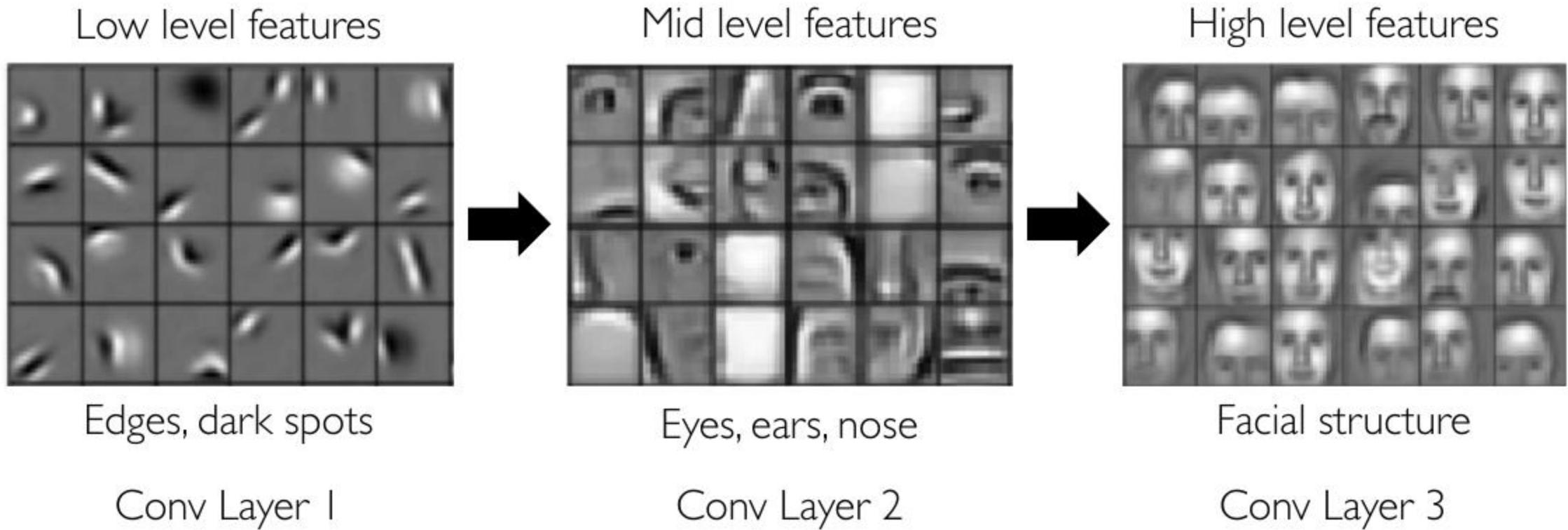
Train model with image data.
Learn weights of filters in convolutional layers.

Representation Learning

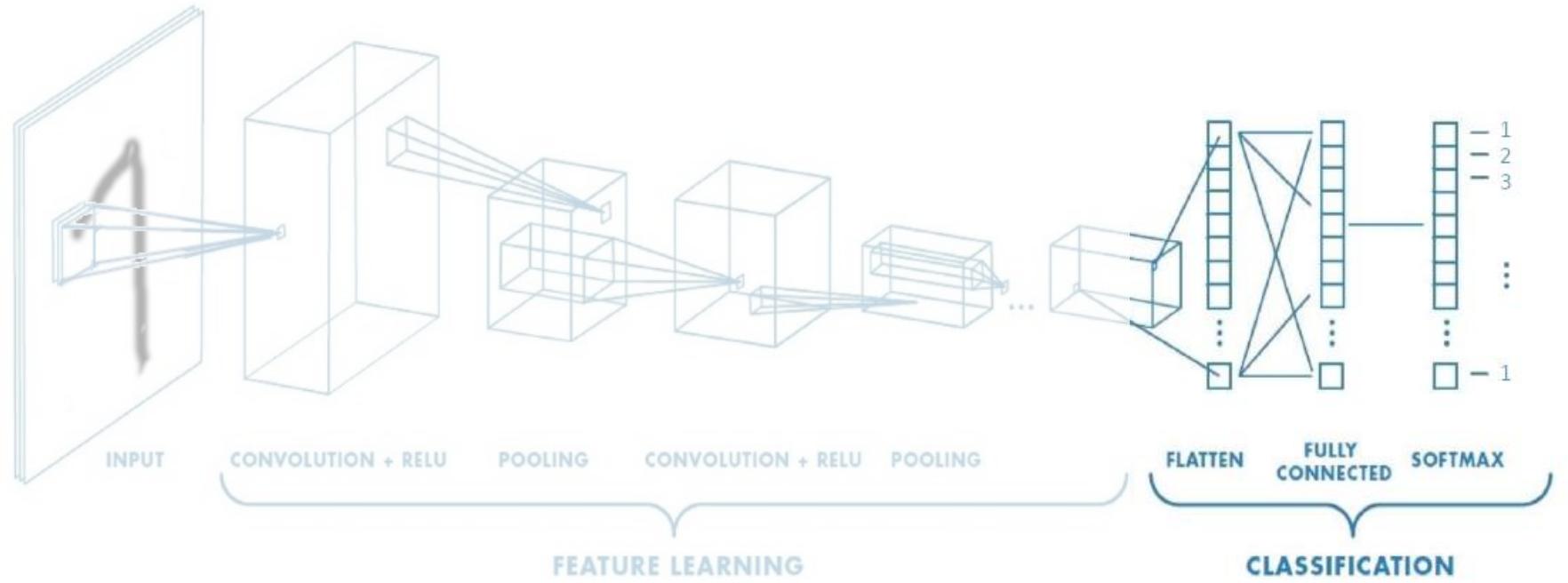


Convolutional Neural Networks

Representation Learning



Classification

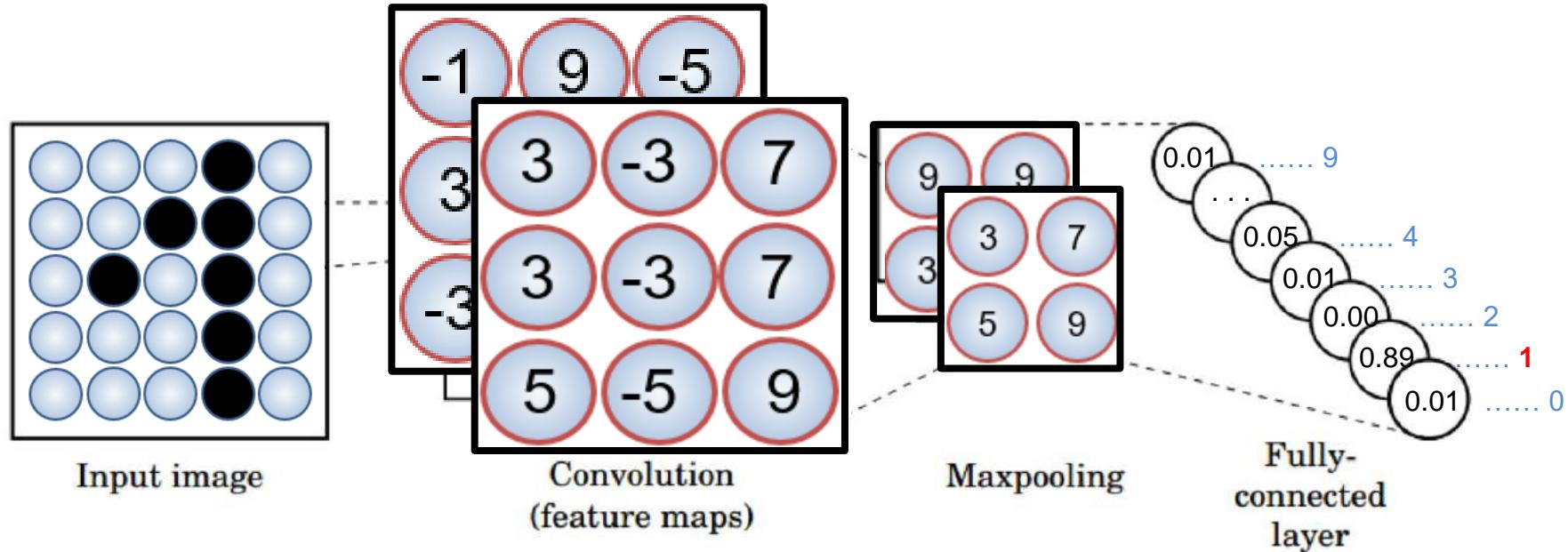


- CONV and POOL layers output high-level features of input
- Fully connected layer uses these features for classifying input image
- Express output as **probability** of image belonging to a particular class

$$\text{softmax}(y_i) = \frac{e^{y_i}}{\sum_j e^{y_j}}$$

Convolutional Neural Networks

Building a CNN – Example



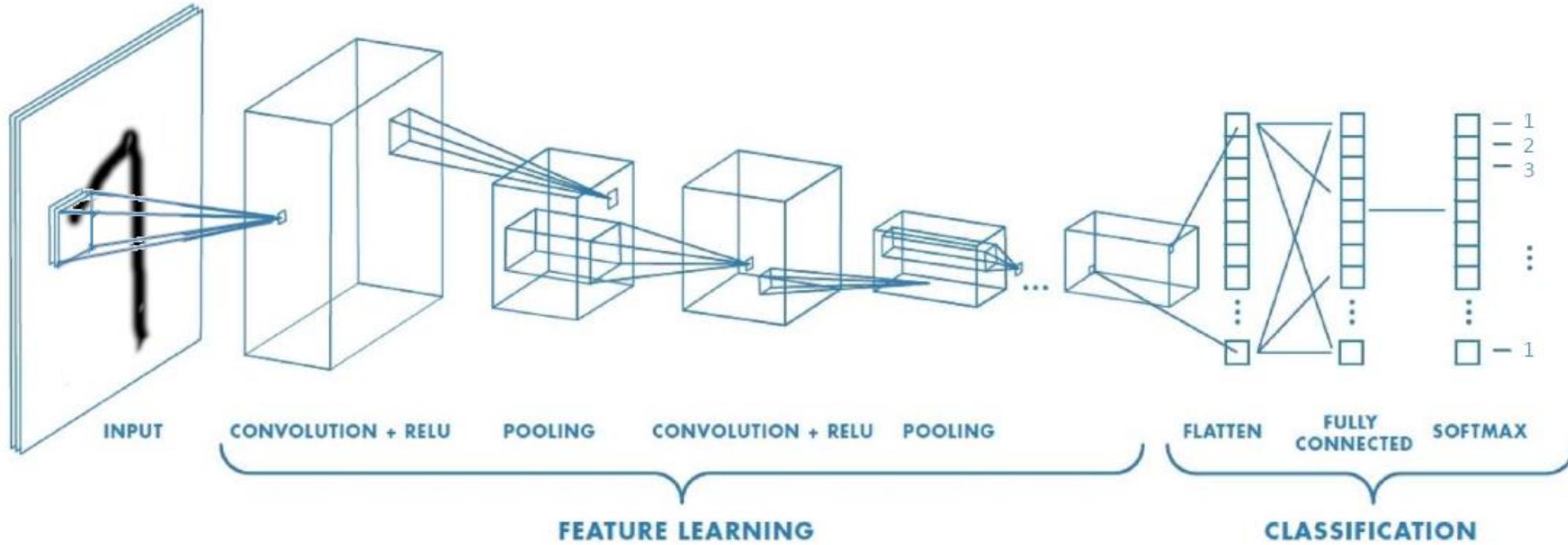
- 1. Convolution:** Apply filters with learned weights to generate feature maps.
- 2. Non-linearity:** Often ReLU.
- 3. Pooling:** Downsampling operation on each feature map.

Train model with image data.

Learn weights of filters in convolutional layers.

Convolutional Neural Networks

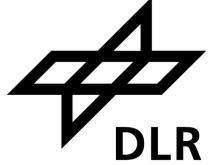
Training with backpropagation



Learn weights for convolutional filters and fully connected layers
Backpropagation: cross-entropy loss

$$J(\theta) = \sum_i y^{(i)} \log(\hat{y}^{(i)})$$

Curriculum



Next:

- Hands-on II

Imprint



Topic: **Introduction to Deep Learning**
Part II – Advanced Concept and Convolutional Neural Network

Date: 2025-11-13

Author: Auliya Fitri, Sai Vemuri, Sreerag Naveenachandran

Institute: Data Science

Image sources: All images “DLR (CC BY-NC-ND 3.0)” unless otherwise stated