Lecture 18. Elements of Image Analysis

COMP90051 Statistical Machine Learning

Semester 2, 2015
Lecturer: Andrey Kan

THE UNIVERSITY OF MELBOURNE

Applications of machine learning

- Machine learning methods (both supervised and unsupervised) have applications across a broad range of domains, including the following
- Bioinformatics and computational biology
  - Subject on Computational Genomics
- Image Analysis
  - Very brief overview in this lecture
- Social Networks
  - Introduction next week

Image segmentation using Otsu's method

A simple, but popular unsupervised method used in image analysis

Image segmentation

- Given an image $I(x, y)$, identify which pixels belong to objects, and which belong to the background
- Each pixel has a position and intensity value
  - Often we refer to pixel intensity value as pixel
- This problem can be viewed as a binary classification task
  - Pixels – data points; classes – {background, object}
- Additionally / alternatively this problem can be viewed as clustering into two clusters

Otsu's thresholding

- Consider a grayscale image $I(x, y)$, where pixels take values in a range from 0 (black) to I_{max} (white)
  - Intermediate values correspond to various levels of intensity (~~50~~ shades of gray)
- A thresholding method aims to identify a cutoff level θ, so that pixels $< \theta$ are considered background, and pixels $> \theta$ are considered a part of object
  - If background is considered darker than objects
- Note that choosing a threshold implies clustering in two groups (background pixel values, object pixel values)
- Otsu's thresholding is an <u>unsupervised</u> method that aims to find a cutoff θ^* that minimizes within-class variance, while maximizing between-class variance

Otsu's thresholding: problem statement

- Within-class variance as a function of θ
  - $\sigma_w^2(\theta) \stackrel{\text{def}}{=} n_1(\theta)\sigma_1^2(\theta) + n_2(\theta)\sigma_2^2(\theta)$
  - n_1, n_2 – proportions of pixels in each class
  - σ_1^2, σ_2^2 – variance of each class
- Otsu showed that minimizing σ_w^2 is the same as maximizing between-class variance that can be expressed as
  - $\sigma_b^2(\theta) = n_1(\theta)n_2(\theta)\left(\mu_1(\theta) - \mu_2(\theta)\right)^2$
  - μ_1, μ_2 – mean of each class
- An optimization problem with no closed-form solution
  - Iterative approach?

Otsu's thresholding: solution

- Iterative approach?
  - Sort of. We can simply try all possible thresholds!
  - Exhaustive search approach
  - E.g., for an 8-bit image, there are 256 levels → 256 thresholds

Image: www.freephotos.lu and Pikez33, Wikimedia Commons (CC1)

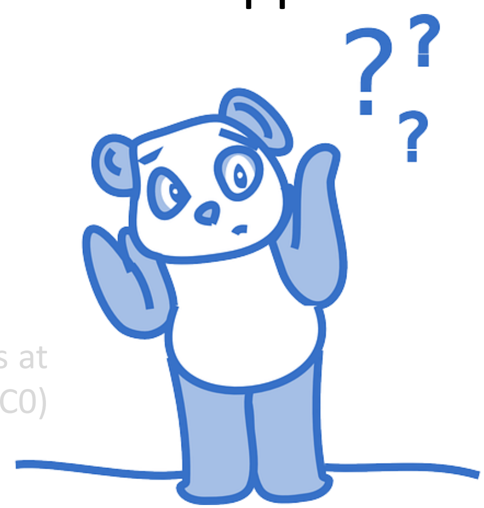
Checkpoint

- Which of the following statements is true?


For a given number of clusters K, EM algorithm will find the Gaussian mixture with highest possible likelihood


AIC enables one to identify a feasible probabilistic model for the data

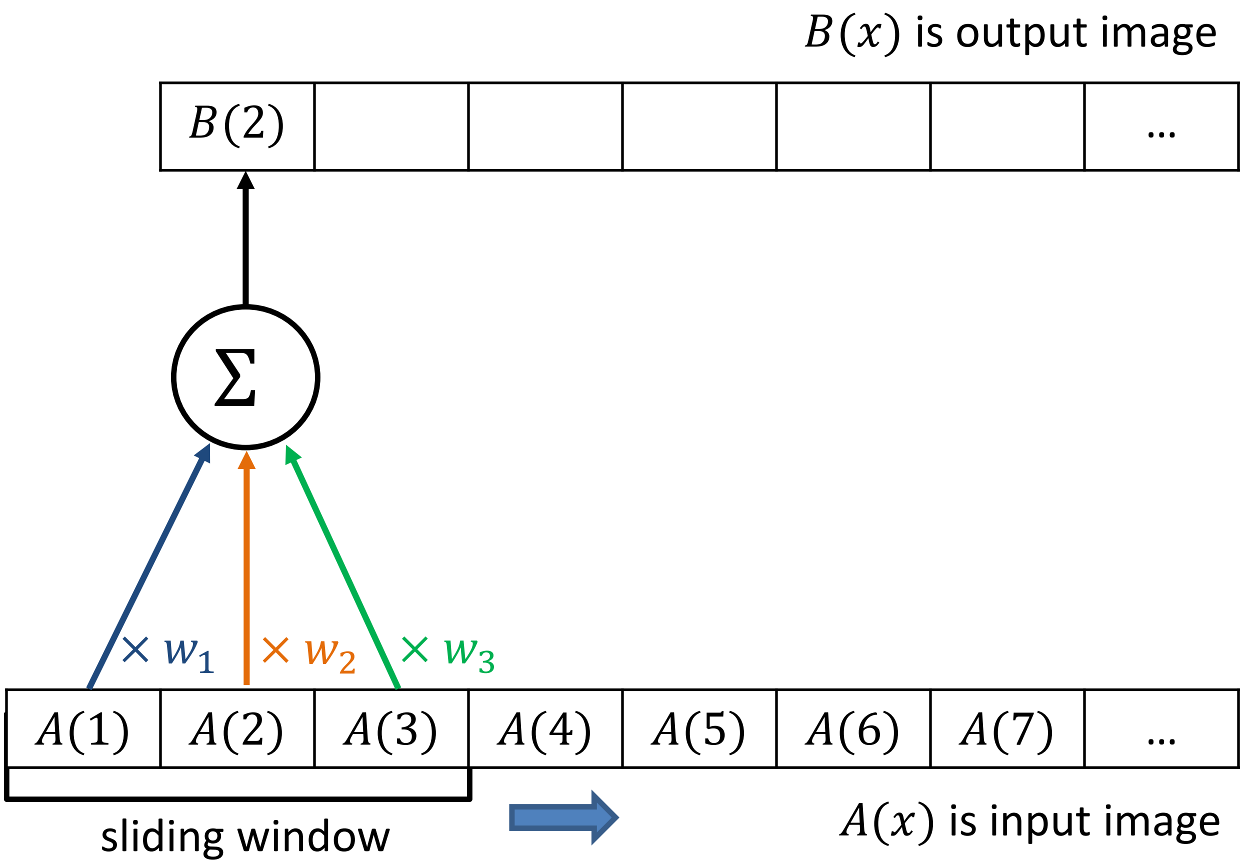

For Gaussian mixture models, EM algorithm is applied before AIC can be used



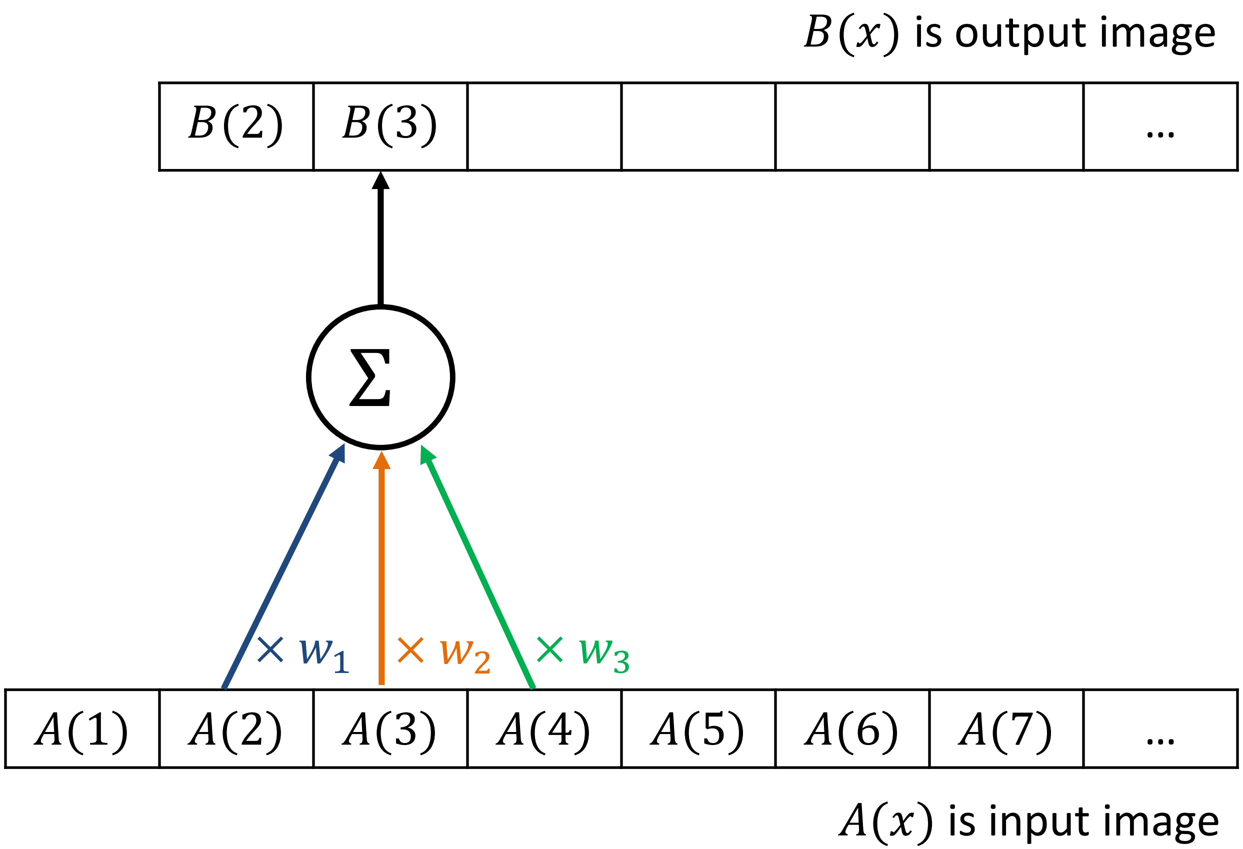
Convolution and its applications

A powerful signal processing method

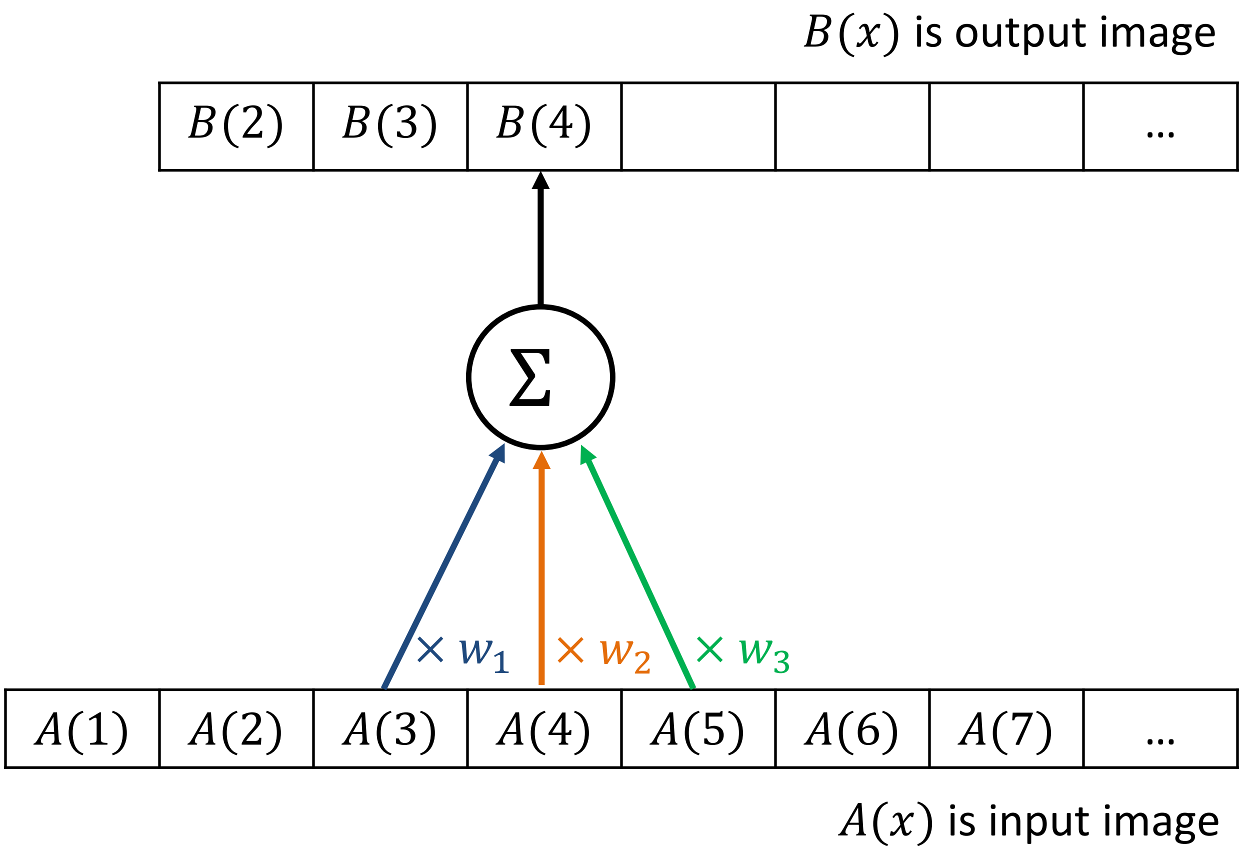
Convolution



Convolution



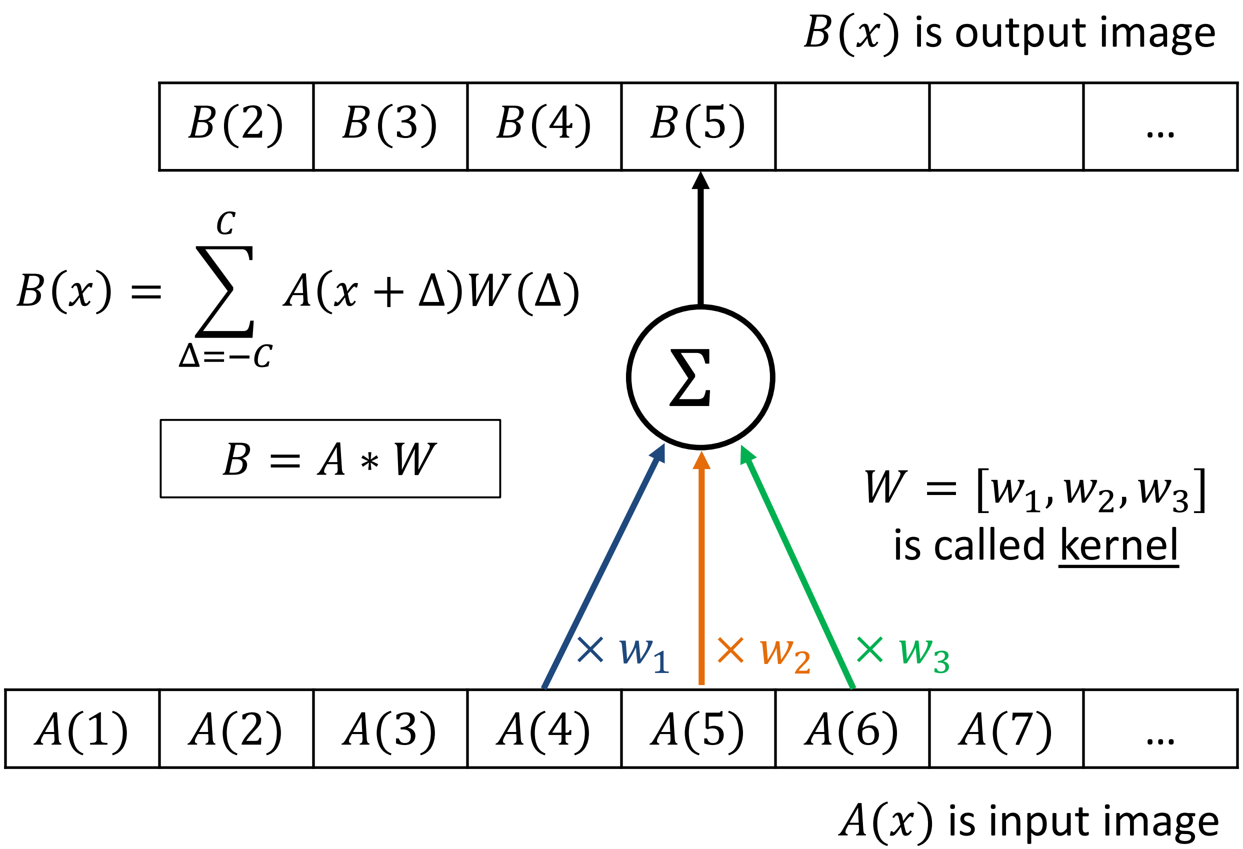
Convolution



Convolution

$$B(x) = \sum_{\Delta=-C}^{C} A(x+\Delta)W(\Delta)$$

$$\boxed{B = A * W}$$

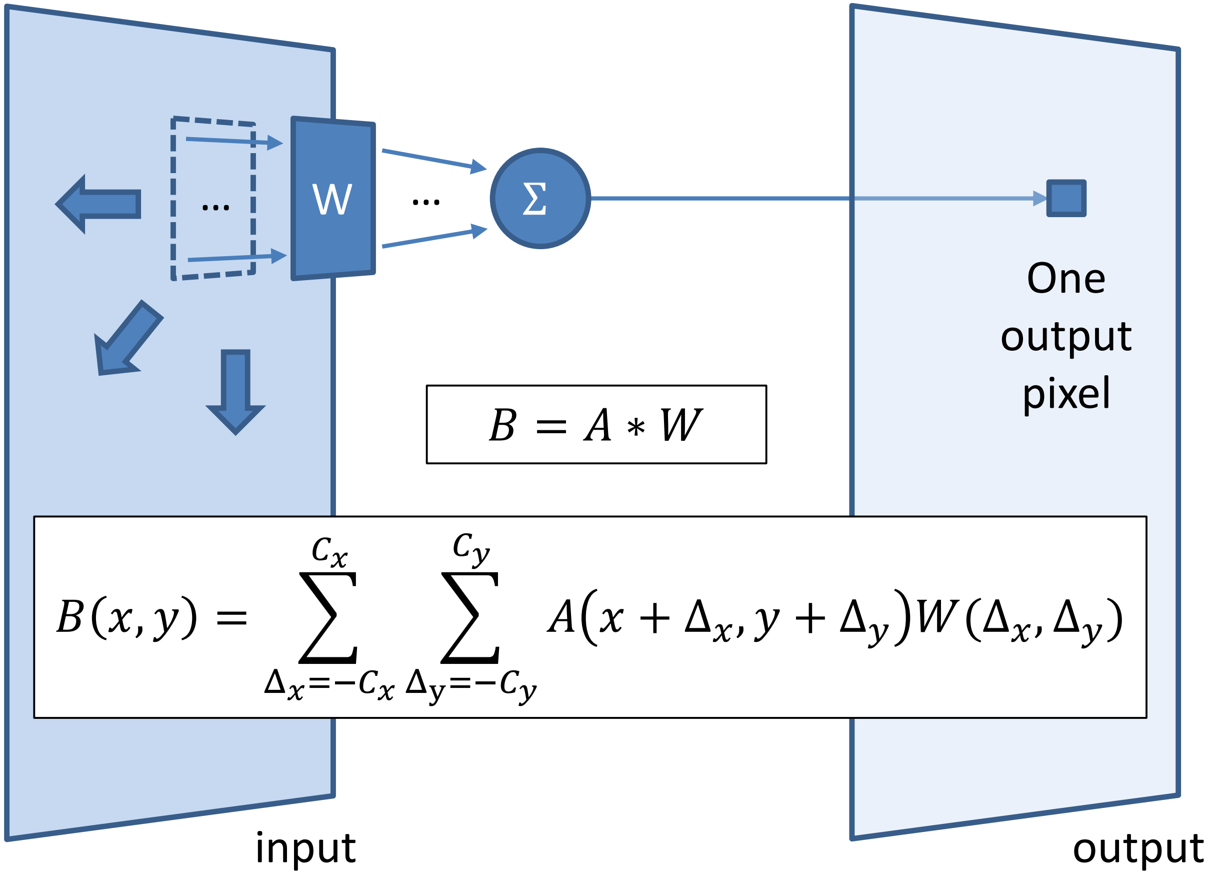


$W = [w_1, w_2, w_3]$ is called <u>kernel</u>

Heads up! Overridden definition

- In the context of kernel methods
  - Kernel is a function $K(\boldsymbol{u}, \boldsymbol{v}) = \varphi(\boldsymbol{u}) \cdot \varphi(\boldsymbol{v})$
- In the context of image analysis and convolutional neural networks
  - Kernel is a matrix (or vector) that slides along the input image

Convolution on 2D images



$$B = A * W$$

$$B(x, y) = \sum_{\Delta_x = -C_x}^{C_x} \sum_{\Delta_y = -C_y}^{C_y} A(x + \Delta_x, y + \Delta_y) W(\Delta_x, \Delta_y)$$

Gaussian blur

- Each pixel is replaced with a weighted sum of itself and neighboring pixels
  - Further away the neighbor is, smaller the weight
- Can be used for noise removal or smoothing

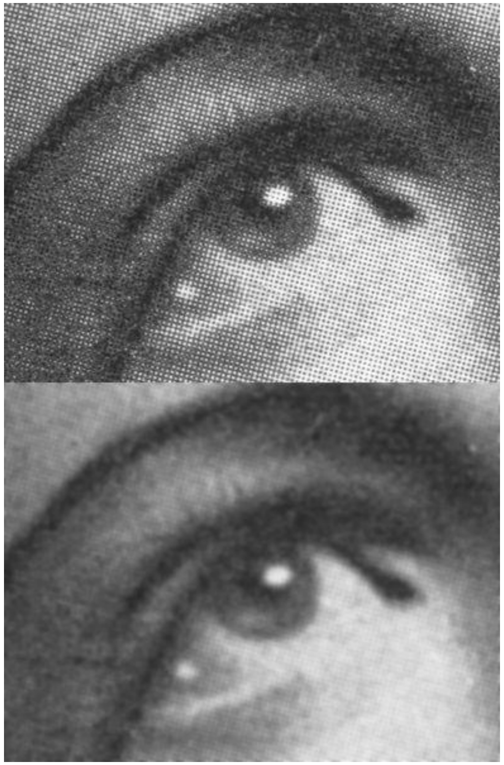


Image: Anonymous, Alex:D, Wikimedia Commons (public domain)

Gaussian blur

- Convolving image with a 2D Gaussian (symmetric uncorrelated)
  - $G(x,y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp - \frac{x^2+y^2}{2\sigma^2}$
  - Also known as 2D Weierstrass transform
  - One parameter σ
- In theory, the size of a Gaussian kernel is infinite
  - In practice, square kernel matrix with side of approximately 6σ

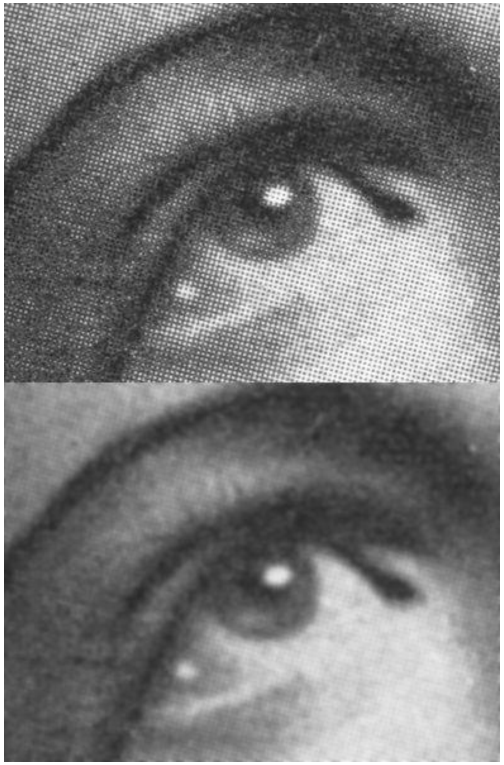


Image: Anonymous, Alex:D, Wikimedia Commons (public domain)

Edge detection using Sobel kernel

- Edges are sharp transitions in image intensity
- Gradient at each image location is a measure of a change at that point
- Sobel kernels (applied at every point) estimate the gradient in x or y direction

$$S_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} \qquad S_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix}$$

Edge detection using Sobel kernel

- Given input image A, the output edge image is

$$G_x = S_x * A \qquad G_y = S_y * A$$

$$G(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2}$$

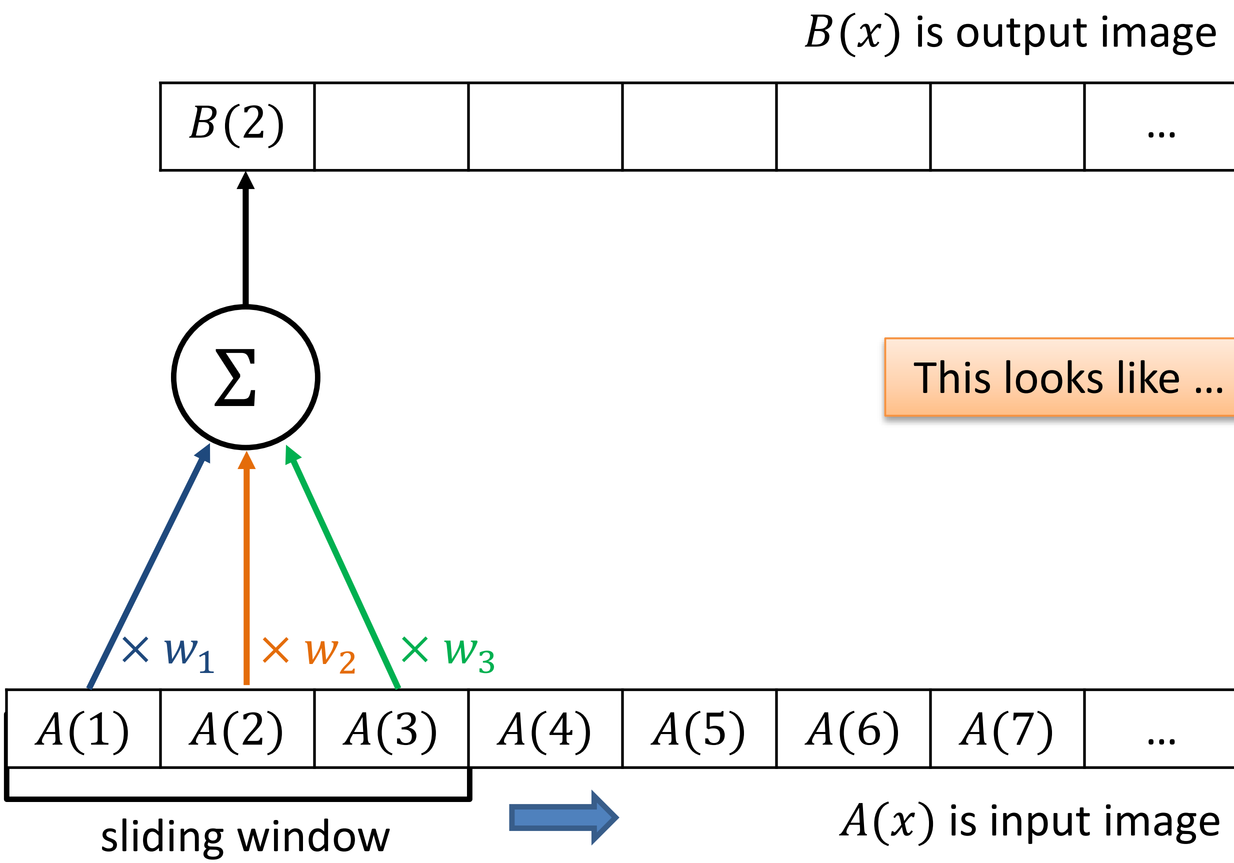
Image: Davidwkennedy, Wikimedia Commons (CC3)

Introduction to Convolutional Neural Networks

Something that can recognise cat images on the Internet

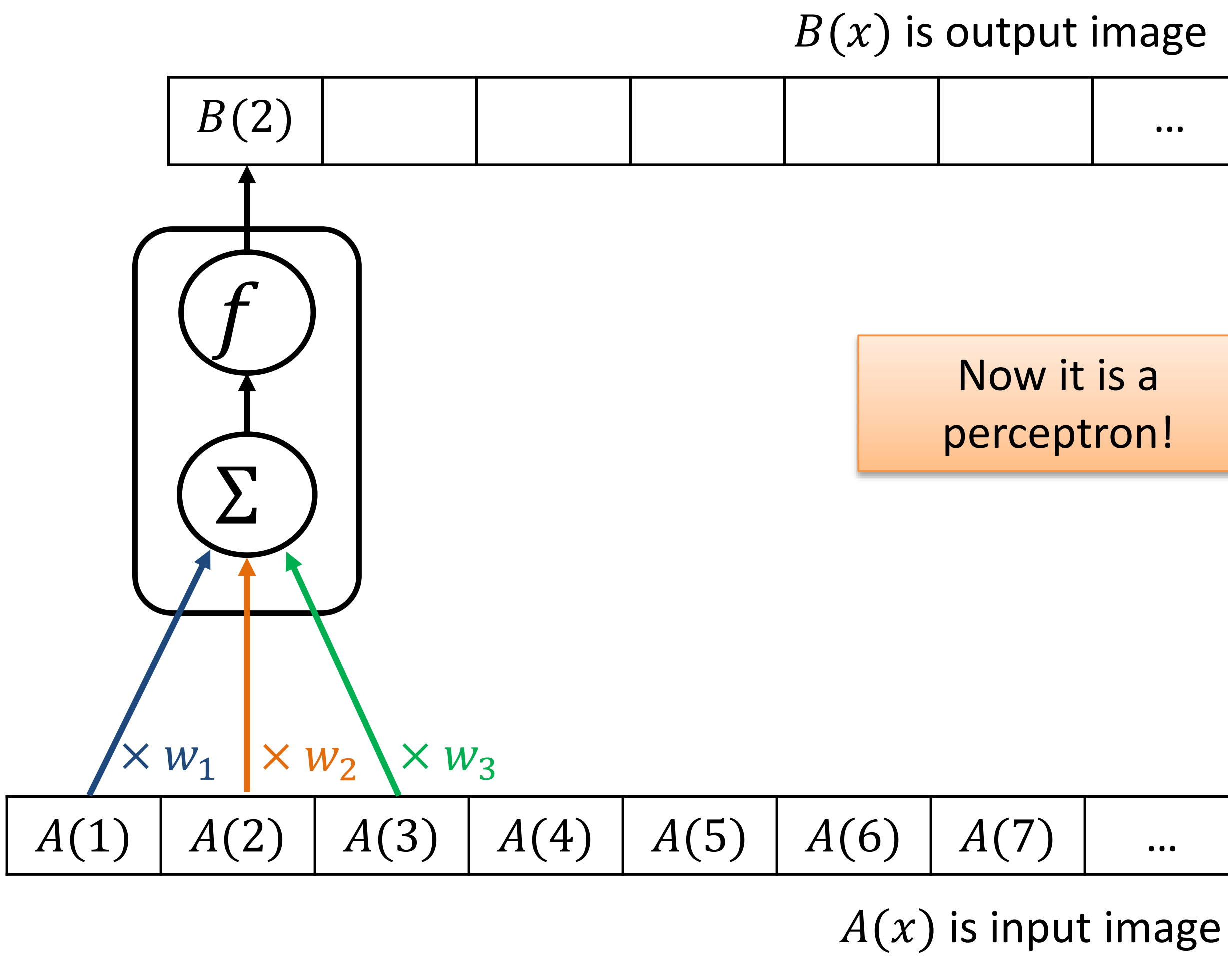
One of the best performing methods in image recognition

Convolution



This looks like …

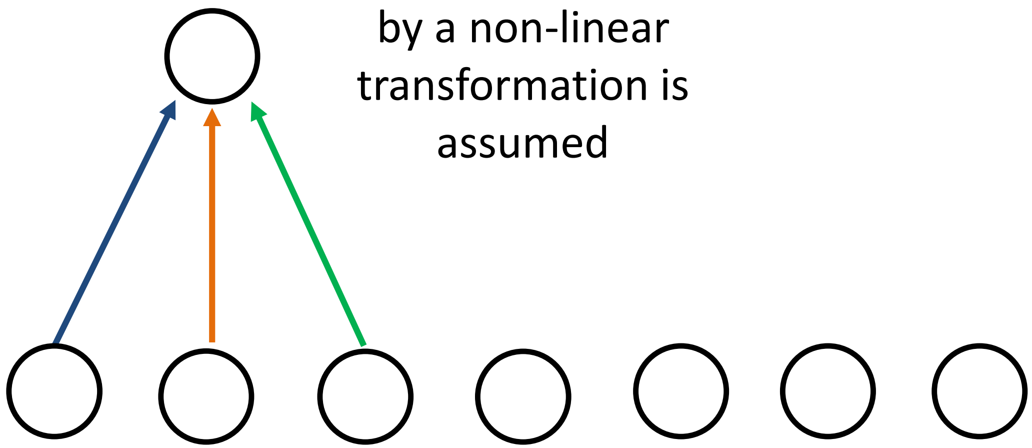
Convolutional layer



Now it is a perceptron!

Simplifying graphical notation

Summation followed by a non-linear transformation is assumed



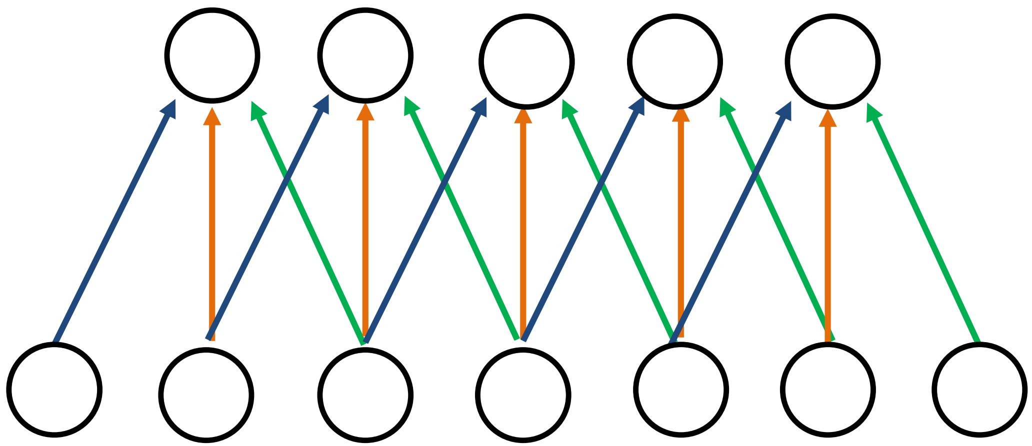
Each node in this layer is what was called $B(i)$

Each node in this layer is what was called $A(i)$

Each pixel form the input image is an input unit

Convolutional layer

Elements of a convolutional layer

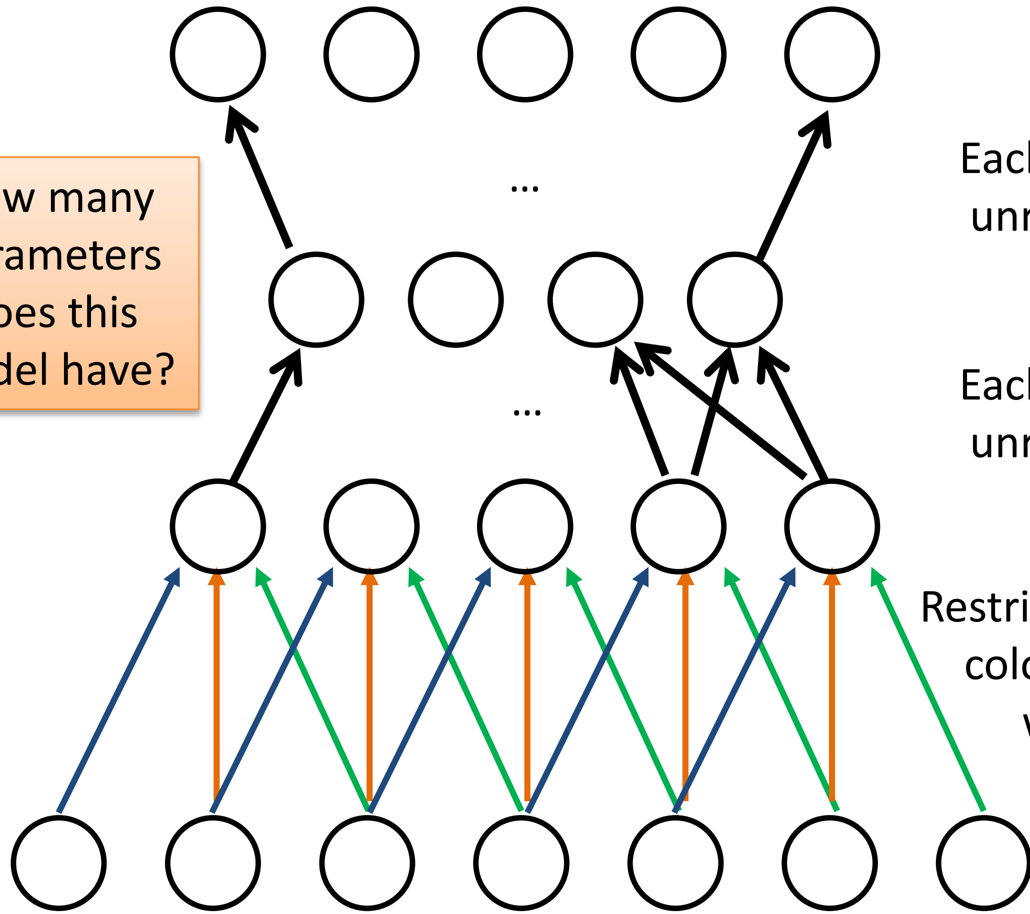


Each pixel form the input image is an input unit

Edges of the same colour have to have the same weights!

Convolutional neural network example

How many parameters does this model have?



Output layer (e.g., cats, dogs, birds, bats, men)

Each to each, unrestricted

Fully connected layer

Each to each, unrestricted

Convolutional layer

Restriction: same color – same weight

Input layer: each pixel form the input image is an input unit

Convolutional neural networks (CNN)

- CNN is a <u>supervised</u> learning method
- Rather than using a pre-defined kernel (e.g., Gaussian or Sobel), CNNs learn weights in the kernel matrix from training data
  - Other weights (in fully connected layers) are also learned from data
  - Training using backpropagation, same as ANN studied earlier
- There can be any number of convolutional layers, and any number of fully connected layers
  - Usually fully connected layers follow convolutional layers
- The same layer (e.g., input) can be convolved with different kernels, producing so called different feature maps
- Rectified linear units $f(s) = \max(0, s)$ are a popular choice for transfer function

Summary

- What are popular domains of application for machine learning?
- What is image segmentation? How is it related to classification and clustering
- What is convolution? What can it be used for?
- What is the basic principles of convolutional neural networks?