Exercises (Lectures 6)
Computer Vision 1, Master AI

Exercise 1. Deep Learning

Deep learning and ConvNets are very useful for object recognition and detection.

- (a) Consider the following linear perceptron classifier:

$$h(\vec{x}, \theta) = \begin{cases} \text{-1, if } 3x_1 + 4x_2 - 24 \\ \text{+1, if } 3x_1 + 4x_2 - 24 \end{cases} . \tag{1}$$

    Depict the neuron for this perceptron. What is the logit and which non-linear function f is used?

- (b) Consider an image with pixel values $\vec{x} = [10, 11, 1, 2]^T$, weight matrix
$$\vec{M} = \begin{bmatrix} 1 & 0.5 & \text{-0.2} & 0.3 \\ \text{-0.3} & \text{-0.1} & 0.7 & 0.3 \\ 0.5 & 0.2 & 1.0 & \text{-0.1} \\ \text{-0.1} & 0.3 & 0.2 & 0.3 \end{bmatrix}$$, and bias $\vec{b} = [1, 0.5, -0.1, 0.4]^T$. Compute the logit and output y_i using a softmax layer i.e. $y_i = \frac{e^{z_i}}{\sum_j e^{z_j}}$. What can you conclude about the prediction?

- (c) Consider a full-color 250x250 pixel image. What would be the number of weights for the input layer?

- (d) You have a 32x32x5 image and k filters with an extent of 5 and consider the way most convolutional neural networks are implemented. What is the depth of the k filters? If you use no padding and stride $s = 1$, what will be the output size of the activation map?

- (e) Consider a feature map with values:

20	23	12	1
12	10	2	5
200	190	6	7
10	20	1	2

    . Calculate the max pooling layer with an extent e and stride s of 2. When are max pooling layers useful?

- (f) What are the different layers in a standard CNN? What are the layers of VGG net?

- (g) What is transfer learning and when is it useful?

Exercise 2. Sliding Window Approach

Sliding window approach is a popular way to detect objects in images.

- (a) What is the basic pipeline for window-based object detection?

- (b) What are the advantages of a sliding-window approach?

- (c) Given an image of 256x256, how many windows are required to detect objects for 8 different orientations and 6 scales?

- (d) Assume that for a strong classifier (e.g. non-linear SVM), the time for window classification is about 0.01 seconds. How many hours does it take to detect objects in 10,000 images?

- (e) How can one reduce the number of bounding boxes for detection? Now, how long does it take to detect objects in 10,000 images?