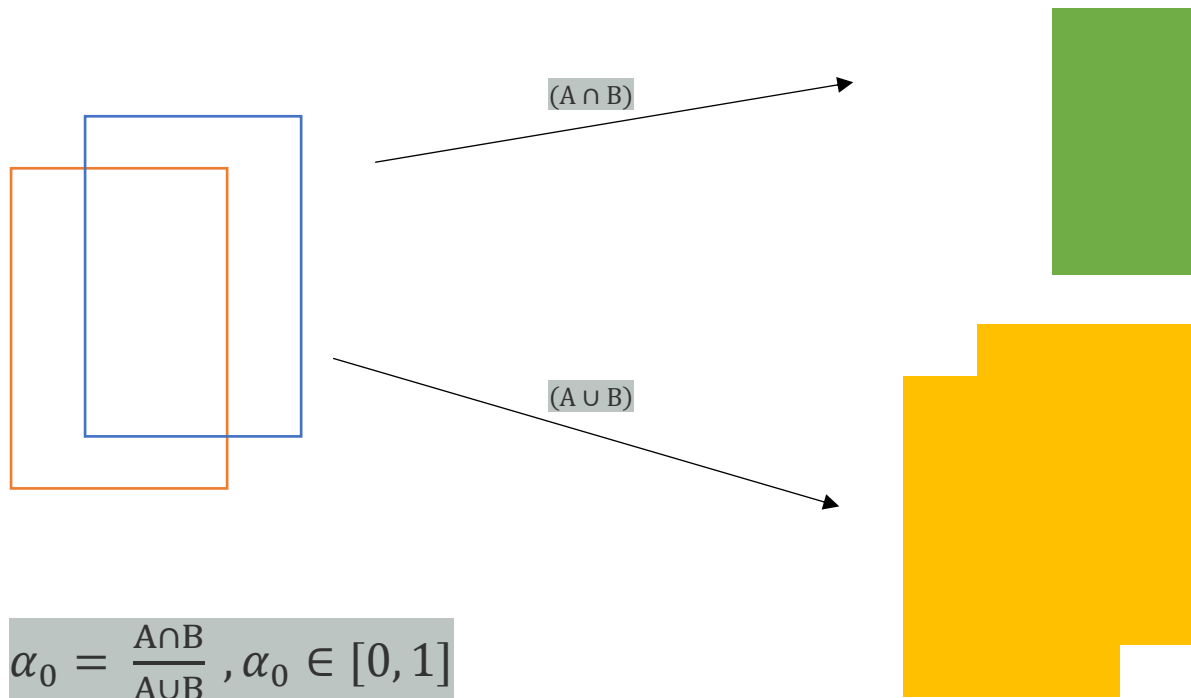


Task 1

a)

Intersection over union is a way to decide if a predicted bounding box with an associated confidence is a correct detection.



b)

$$precision = \frac{\text{correct predictions}}{\text{total predictions}} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$

$$recall = \frac{\text{correct predictions}}{\text{total wanted entries}} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

A true positive is a correct prediction. A false positive is a false prediction.

c)

Recall₁ = [0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0]

P_{interp1} = [1.0, 1.0, 1.0, 1.0, 1.0, 0.5, 0.5, 0.5, 0.2, 0.2, 0.2]

Recall₂ = [0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0]

P_{interp2} = [1.0, 1.0, 1.0, 1.0, 0.8, 0.6, 0.5, 0.5, 0.2, 0.2, 0.2]

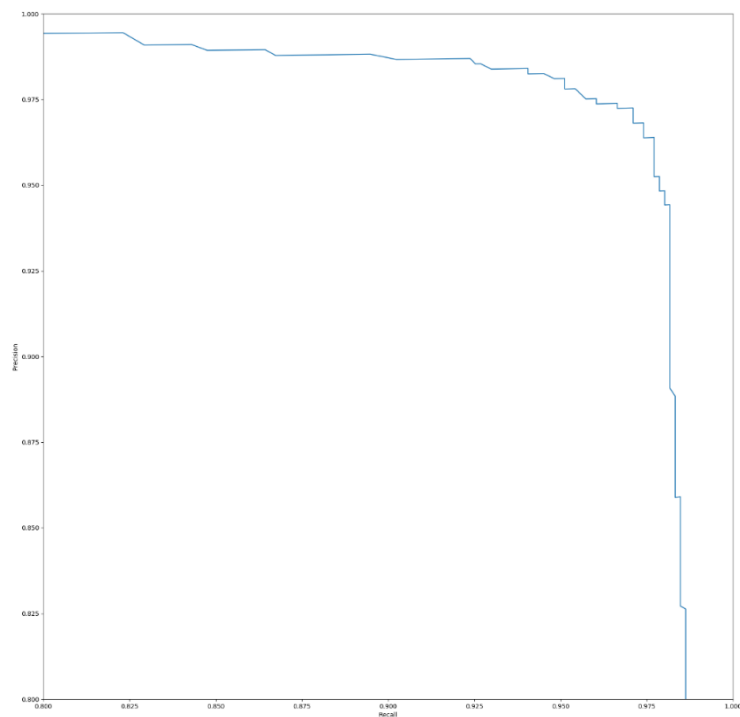
$$\text{average precision class 1} = \frac{5 \cdot 1 + 3 \cdot 0.5 + 3 \cdot 0.2}{11} \approx 0.65$$

$$\text{average precision class 2} = \frac{4 \cdot 1 + 0.8 + 0.6 + 2 \cdot 0.5 + 3 \cdot 0.2}{11} \approx 0.64$$

$$\text{mean average precision} = \frac{0.65 + 0.64}{2} \approx 0.641$$

Task 2

f)



Task 3

1. The technique of filtering out overlapping bounding boxes is called hard negative mining.
2. False. Higher resolution feature maps are responsible for detecting small objects.
3. Default-boxes are used to keep the complexity of the model low. Default-boxes help the model to make diverse initial predictions. If the shape of the boundary-box is determined during training, some classes will probably fit the shape (aspect ratio) of the boundary-box better than others.
4. SSD has additional convolutional feature layers at the end of the truncated base network which allow for predictions at multiple scales. YOLO use fully connected layers that work on a single scale feature map.