

Midterm

Student name: *Zahin Mohammad*

z5mohamm

20669584

Course: *ECE 495* – Professor: *Krzysztof Czarnecki*

Due date: *March 3rd, 2021*

Q1. ADS Fundamentals [1 mark]

1a. A vehicle is equipped with a stop-and-go pilot, which can fully operate a vehicle on a highway in traffic jams, but with a fallback-ready user. Which level(s) of driving automation is this driving automation system operating at? [1 mark]

Answer. The driving automation system is operating at level 3.

Q2. Computer Vision Fundamentals [11 marks]

2a. Compute 1-D cross-correlation by applying the following filter $[0 \ 2 \ 1]$ to the following signal $[0 \ 1 \ 3 \ 0]$ (assume enough zero padding to show all non-zero output). [2 marks]

Answer.

$$\begin{aligned} &= [0 * 0 + 2 * 0 + 1 * 1 \quad 0 * 0 + 2 * 1 + 3 * 1 \quad 0 * 1 + 2 * 3 + 1 * 0 \quad 0 * 3 + 2 * 0 + 1 * 0] \\ &= [1 \quad 5 \quad 6 \quad 0] \end{aligned}$$

2b. Assume that the output of cross-correlating the 1-D filter $[2 \ 3 \ 2]$ with some input signal resulted in the following output signal $[7 \ 10 \ 7]$. What would be the output signal have we used convolution instead of cross-correlation and why? [1 mark]

Answer. There the resulting output signal is

$$= [7 \quad 10 \quad 7]$$

The result of cross-correlation and convulsion in this instance are the same with this specific filter. This is because during convolution the filter is "flipped", but in this instance the filter is symmetric.

2c. What would be the result of applying this filter to an image (Hint: add the Gaussian kernel to this filter from question 2e to recognize it)? [1 mark] Explain.

$$\frac{1}{16} \begin{bmatrix} -1 & -2 & -1 \\ -2 & 28 & -2 \\ -1 & -2 & -1 \end{bmatrix}$$

Answer. This is a high pass filter (laplacian filter). The resulting image would only have high frequency components such as edges (the filter de-emphasizes slow varying gradients; regions without edges).

2d. What is the name of the following filter and what is it computing? [1 mark]

$$\frac{1}{16} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 1 & 0 \\ 0 & 2 & 4 & 2 & 0 \\ 0 & 1 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} - \frac{1}{273} \begin{bmatrix} 1 & 4 & 7 & 4 & 1 \\ 4 & 16 & 26 & 16 & 4 \\ 7 & 26 & 41 & 26 & 7 \\ 4 & 16 & 26 & 16 & 4 \\ 1 & 4 & 7 & 4 & 1 \end{bmatrix}$$

Answer. This is a DoG filter, which calculates a difference in gaussian's. The first gaussian has a max magnitude of 0.25, and the second gaussian has a max magnitude of 0.15. This shows that the first gaussian filter has a smaller σ then the second ($\sigma_1 < \sigma_2$), therefore this is a band pass filter.

2e. Complete the Gaussian filter shown below by specifying the cells marked by "?". How did you determine the missing values? [2 marks]

$$\frac{1}{16} \begin{bmatrix} 1 & 2 & ? \\ ? & ? & ? \\ ? & ? & ? \end{bmatrix}$$

Answer.

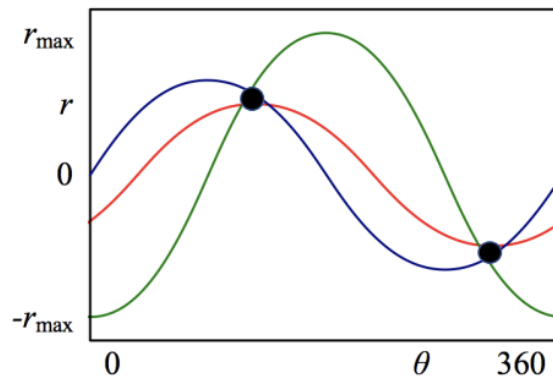
$$= \frac{1}{16} * \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$$

All values in the grid need to add to 16 ($\frac{1}{16} * 16 = 1$), and the matrix must be symmetric.

2f. Why is the Canny filter using double thresholding? [1 mark]

Answer. The Canny filter uses double thresholding to remove unwanted edges. It does this by allowing all strong edges (above top threshold) and all weak edges (above second threshold) that also connect to a strong edge. This results in relevant continuous lines (discontinuous strong edges connected with weak edges). Anything below the lower threshold is discarded.

2g. Consider the following representation in Hough space (polar coordinates).



2g.1 What is each of the individual sinusoids corresponding to in the input image? [1 mark]

Answer. Each sinusoid in the Hough space correspond to a point in the input image.

2g.2 What is each of the black dots (sinusoid intersections) corresponding to in the input image (note that the horizontal axis sweeps 360 degrees)? [1 mark]

Answer. The first black dot refers to a line in the input image connecting three points (red, green and blue point). The second black dot refers to the same line but rotated 180 degrees (i.e. the same line).

2h. Which of the following value profiles represent black in HSV? (select all that apply) [1 mark]

1. 0,high,high
2. any,low,low
3. any,low,high
4. 60,high,high
5. any,high,low
6. any,any,low

Answer. Black can be represented with options 2, 5, and 6. The key is any HSV that has a low Value.

Q3. Machine Learning Fundamentals [21 marks]

3a. Starting with the probabilistic model for linear regression (assume single input x and single out y), show that maximizing the likelihood for a dataset (x, y) (with i.i.d. datapoints) is equivalent to minimizing the sum of squared errors. Hint: go via negative log likelihood [5 marks]

Answer. Maximizing the likelihood is the same as maximizing the log likelihood as \ln is monotonic.

$$P(y|x; w_0, w_1) = \prod P(y = y^i | x^i; w_0, w_1) \quad (1)$$

$$LL(y|x; W) = \ln P(y|x; w_0, w_1) = \sum_{i=1}^n \ln(P(y = y^i | x^i; w_0, w_1)) \quad (2)$$

Maximizing log likelihood is equivalent to minimizing negative log likelihood.

$$(w_0^*, w_1^*) = \operatorname{argmin} \left(\sum_{i=1}^n -\ln(P(y = y^i | x^i; w_0, w_1)) \right) \quad (3)$$

3b. What is the regression loss (as used in class) for a data point with label 0.4 and predicted output 0.7? [1 mark]

Answer.

$$\text{loss} = (0.4 - 0.7)^2 = 0.09 \quad (4)$$

3c. (Apply what you've learned in lecture and Assignment 2) Given the following set of input vector X , ground truth vector Y , and weight matrices W_1 , W_2 , B_1 , and B_2 of a 2 layer fully connected neural network, what is the inference probability of the correct class? What is the cross-entropy loss value? Assume ReLU activation on the first hidden layer and softmax activation on the output layer. Show each step of the computation. Hint: Use numerically stable softmax and assume $e^{-1} \approx 0.37$ and $e^{-7} \approx 0.00$ [5 marks]

X	W1		B1	W2			B2	Y
2	10	8	-25	3	1	1	-3	0 (Man)
0	-5	3	11	1	2	3	3	1 (Moose)
	4	2	-10	5	4	2	0	0 (Deer)

Answer.

$$\begin{aligned}
 Y &= \text{softmax}\left(\begin{bmatrix} -3 \\ 3 \\ 0 \end{bmatrix} + \begin{bmatrix} 3 & 1 & 1 \\ 1 & 2 & 3 \\ 5 & 4 & 2 \end{bmatrix} * \text{relu}\left(\begin{bmatrix} -25 \\ 11 \\ -10 \end{bmatrix} + \begin{bmatrix} 10 & 8 \\ -5 & 3 \\ 4 & 2 \end{bmatrix} * \begin{bmatrix} 2 \\ 0 \end{bmatrix}\right)\right) \\
 &= \text{softmax}\left(\begin{bmatrix} -3 \\ 3 \\ 0 \end{bmatrix} + \begin{bmatrix} 3 & 1 & 1 \\ 1 & 2 & 3 \\ 5 & 4 & 2 \end{bmatrix} * \text{relu}\left(\begin{bmatrix} -5 \\ 1 \\ -2 \end{bmatrix}\right)\right) \\
 &= \text{softmax}\left(\begin{bmatrix} -3 \\ 3 \\ 0 \end{bmatrix} + \begin{bmatrix} 3 & 1 & 1 \\ 1 & 2 & 3 \\ 5 & 4 & 2 \end{bmatrix} * \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right) \\
 &= \text{softmax}\left(\begin{bmatrix} -2 \\ 5 \\ 4 \end{bmatrix}\right)
 \end{aligned} \tag{5}$$

Let

$$A = \begin{bmatrix} -2 \\ 5 \\ 4 \end{bmatrix} \tag{6}$$

Then

$$\begin{aligned}
 Y &= \text{softmax}(A) \\
 &= \frac{e^{A - \max(A)}}{\sum_{j=1}^k e^{A_j - \max(A)}} \\
 &= \frac{1}{e^{-7} + e^0 + e^{-1}} * \begin{bmatrix} e^{-7} \\ e^0 \\ e^{-1} \end{bmatrix} \\
 &= \frac{1}{1.37} * \begin{bmatrix} 0.00 \\ 1 \\ 0.37 \end{bmatrix} \\
 &= \begin{bmatrix} 0.00 \\ 0.73 \\ 0.27 \end{bmatrix}
 \end{aligned} \tag{7}$$

The inference probability of the correct class is 73%. The cross entropy loss is defined as $-\ln(p)$ therefore it is equal to 0.31.

3d. Consider the computational graph below for the following function

$$f(x_1, x_2) = \ln(3x_1 + e^{2x_2}) \quad (8)$$

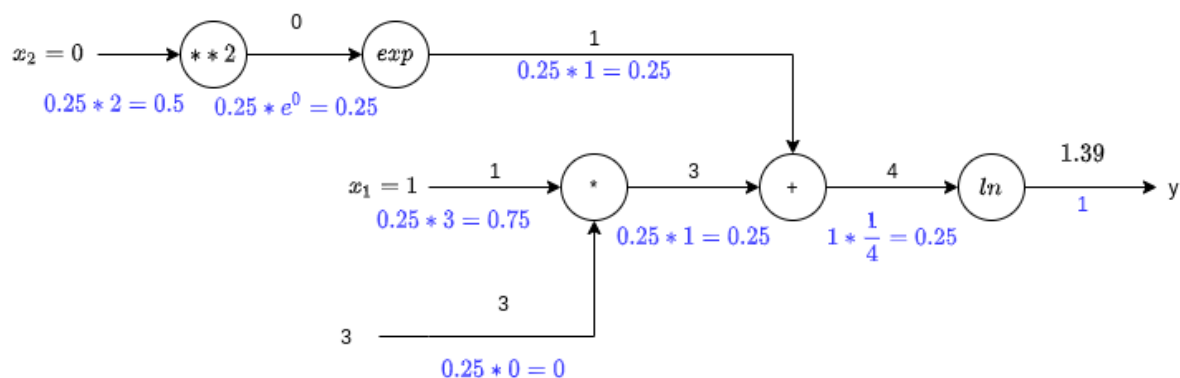
Draw the computational graph and annotate it with the forward pass (above the arrows) and backward pass (below the arrows) for $x_1 = 1$ and $x_2 = 0$ (propagate the gradient back to each function input). Recall

$$\frac{de^x}{dx} = e^x \quad (9)$$

$$\frac{d\ln(x)}{dx} = \frac{1}{x} \quad (10)$$

Assume $\ln(4) \approx 1.39$ [5 marks]

Answer.



3e. What is the difference between Stochastic Gradient Descent and ordinary (Batch) Gradient Descent? [1 mark]

Answer. Batch gradient descent computes the gradient using the entire dataset whereas Stochastic gradient descent computes the gradient using either one or a subset of the dataset.

3f. How is the condition of overfitting defined? [1 mark]

Answer. Overfitting is a case where the network achieves a low loss on training data, but has significantly larger loss when presented with input from the validation or test dataset. Overfitting can be seen as a failure to generalize the target function, and instead only fit to data it has already seen.

3g. Consider a convolutional layer with an input volume of depth 4 and output volume of depth 128. How many convolutional filters does the layer contain? What is the depth of each filter? [2 marks]

Answer.

$$D_{out} = k, \quad (11)$$

where k is the number of filters, and D_{out} is the output depth. Therefore there are $k = 128$ convolutional filters. The depth of each convolutional filter is 4 to match the input depth.

Q4. Semantic Segmentation [2 marks]

4a. (Apply what you've learned in lecture and Assignment 3) Semantic segmentation architectures sometimes use skip connections from early feature maps of the feature extractor to the corresponding-size upsampled maps in the decoder. What is the role of these connections? [1 mark]

Answer. Skip connections in semantic segmentation provides higher (deeper) layers in the neural network access to lower level features. Additionally, skip connections improve gradient flow in deep neural networks (mitigate the vanishing gradient problem) which leads to improved convergence.

4b. (Assignment 3) You were recommended to use batch norm as part of your network. Given a layer with the linear transformation and an activation function, where is the batch norm operation normally applied? [1 mark]

Answer. The batch normalization happens after the linear transformation and before the activation function.

Q5. Object Detection [9 marks]

5a. Assume that an object detector uses a 5-by-5 grid and 3 anchor boxes at each cell. What is the maximum number of objects that the detector can detect? [1 mark]

Answer. The detector can detect $5 * 5 * 3 = 75$ objects.

5b. On a test set with a total of 4 of cars in the ground truth, a detector produced 3 bounding boxes with the following (score, IoU): (0.5, 0.7), (0.7, 0.8), (0.9, 0.2) (assume that each returned bounding box overlaps with a different ground truth). Assuming a score threshold of 0.6 and IoU threshold of 0.6, specify the number of TP, FP, and FN. [3 marks]

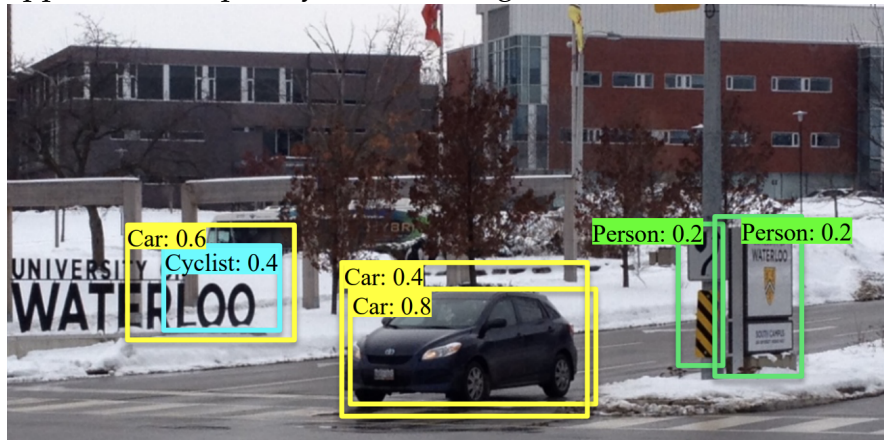
Answer.

$$TP = \text{count}[(0.7, 0.8)] = 1 \quad (12)$$

$$FP = \text{count}[(0.9, 0.2)] = 1 \quad (13)$$

$$FN = \#ground_truths - TP = 4 - 1 = 3 \quad (14)$$

5c. (Assignment 4) Consider the figure below, which shows the bounding boxes predicted by a network (before non-maximum suppression). With an IoU threshold of 0.5 and detection threshold of 0.2, how many bounding boxes are left after non-maximal suppression? Explain your reasoning. [3 marks]



Answer. In this example, no bounding boxes are removed for their score as they are all ≥ 0.2 . NMS is applied per class. Looking at the cyclist bounding boxes there are no overlaps with other cyclist bounding boxes, therefore it is not removed. The person classes visually do not have an $IoU \geq 0.5$ therefore they are not removed. For the car class, the car with a score of 0.4 and 0.8 overlap by more than 50%, therefore the larger score car will remain removing the lower score one. The car with a score of 0.6 is not overlapping with any other car class, therefore it is not removed. In total there are 5 bounding boxes left after non-maximal suppression.

5d. How is it possible for two different bounding boxes to be generated for the same object (before non-maximum suppression)? [1mark]

Answer. Each anchor box can have a detection, and as they overlap, it is possible for overlapping anchor boxes to detect the same object given that they have sufficient IoU and score. If an object fills up most of a grid-square, then likely most of the anchor boxes for that grid square will generate a bounding box for the same object.

5e. Which of the statements is correct? [1mark]

- a) An output neuron in an object detector is influenced only by input pixels within the anchor box assigned to it.
- b) An output neuron in an object detector is influenced only by input pixels within its positive anchor box.
- c) An output neuron in an object detector is influenced only by input pixels within its empirical receptive field.

Answer. The correct answer is c, an output neuron in an object detector is influenced only by input pixels within its empirical receptive field.