Q1
5 Points

Write a convolution filter that will calculate the image gradients in either horizontal or vertical direction.

There are two convolution filters used for this:
1. [-1, 1] => This calculates the image gradients in the horizontal direction. (vertical edges detected)
2. [-1, 1]^T (Becomes a column vector)=> This calculates the image gradients in the vertical direction. (horizontal edges detected).

Q2 Max Pooling
5 Points

What is the purpose of max-pooling in a neural network?

Max Pooling helps in :
1. It provides translational/spatial invariance.
2. As it performs extraction of the most important features from the features extracted by the convolutional layers, it provides a way to downselect the most important features in the image, which helps learn the model efficiently.
3. As there are no learnable parameters involved, it does the task of finding the most significant features without requiring any calculation of gradients.

Q3 Course Evaluation
5 Points

Did you fill out the course evaluation? Did you tell us a) what you liked and

b) what could be improved?

- [ ] No
- [x] Yes, I will/did

Q4 Loss Functions

10 Points

A very common loss function in machine learning is the Euclidean loss or mean squared error: $\mathcal{L}(x, y) = ||x - y||_2^2$. However, when there is uncertainty in the output, this loss function is not a good choice. Why?

When there is an uncertainty in the output, this can introduce some kind of noise in the data. This could result in the predictions being a little off or could learn to model even those uncertainties very well. As this loss function penalizes a lot, even a little difference between the target and the predicted labels, this could be problematic where there are uncertainties in the output.
Furthermore, a neural network trained using the euclidean loss actually predicts the mean of the modes of the output. So, for possible multimodal distributions, this would regressively predict the mean of those modes. Due to the uncertainties involved, this could be very problematic to aggressively train based on those uncertainties as those shouldn't be learned very well but due to the heavy penalty involved, it will do so!

Q5 Interpretability

10 Points

Neuroscience has established evidence for "grandmother neurons" or "Halle Berry neurons" where individual neurons in the brain respond to just one input. There are also results in deep learning that shows similar hidden units are learned by convolutional networks.

However, these results remain controversial. Why? (Hint: Don't damage your brain thinking too hard about this question.)

The controversy is as follows:
If there is an accident and some part of the brain is damaged, then some specific neurons will become inactive as those are damaged. Following this logic, the person won't be able to recognize Halle Berry. This was stated as a factual statement, but since this was not found to be true in general, that's the reason that the results remain controversial since, technically, this could happen.

Q6 Convolution

10 Points

What is the answer to the following convolution? Be careful: we are asking for convolution, not cross-correlation. You can handle border effects by padding with zero. Your answer should be a 4×4 matrix.

$$\begin{pmatrix} 5 & 0 & 8 & 3 \\ 4 & 1 & 2 & 4 \\ 6 & 8 & 7 & 9 \\ 1 & 5 & 6 & 0 \end{pmatrix} * \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix} =$$

Write out your answer in plain text. Put a space between columns, and a newline between rows.

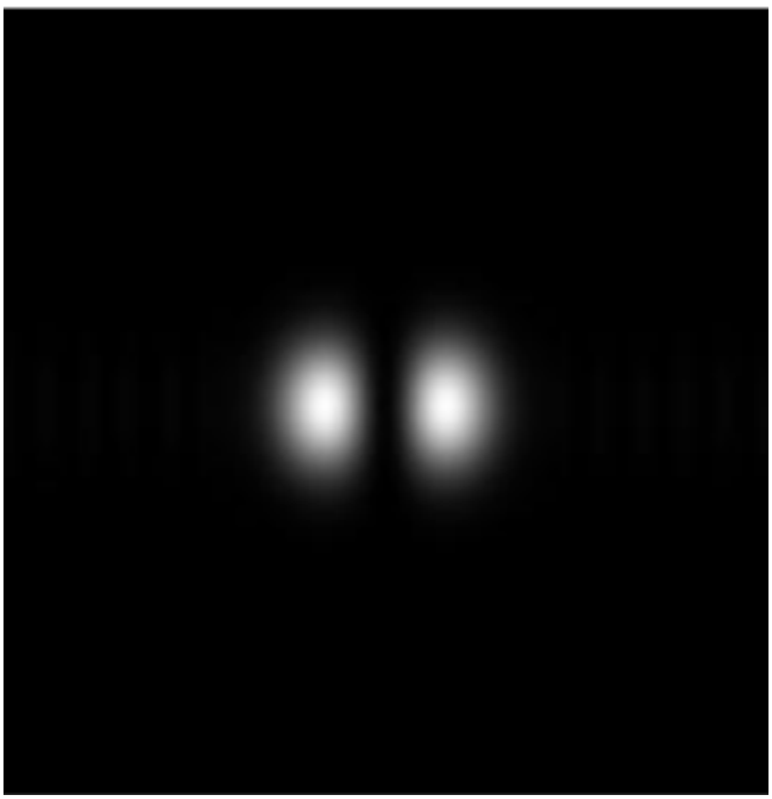
0 -5 0 -8

0 -4 -1 -2
0 -6 -8 -7
0 -1 -5 -6

Q7 Fourier Transform

10 Points

The convolution theorem relates the convolution operation with the Fourier transform. Below is the magnitude in Fourier space of one convolution filter.



What does this filter do? Be as specific as possible.

This is for the Horizontal Laplacian filter, which detects the change in the gradient across the horizontal direction, thus, detecting vertical edges. Also, as some of the fine information is also lost, we can say the image would also be blurred, which will make us lose some of the vertical edges in the final output.

Q8
5 Points

Edge detectors often blur the image as a first step. Why?

Edge detectors often blur the image in the first step. That's primarily because if there is noise in the image, then we need to smoothen the image first to remove the noise; otherwise, there are a lot of gradient changes suggesting edges/lines in the image which are not actually present. As edges are the points in the image where the gradient or derivative is the maximum, and at the points where noise is there, the gradients are also maximum or very high, which would make the edge detection task all the more difficult. So, we do the smoothening in the first step with a gaussian filter, which blurs the image.

Q9 Supervised Learning
10 Points

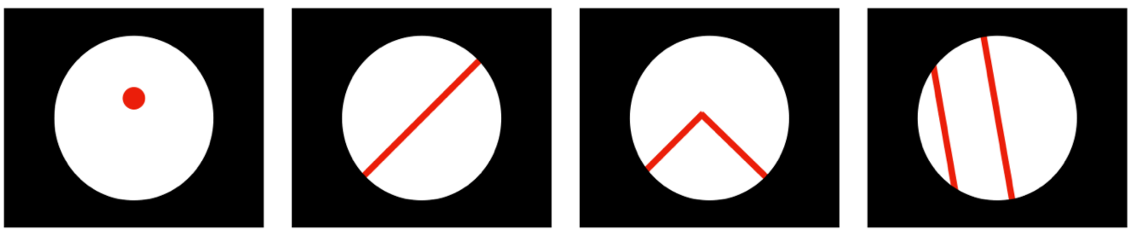
What is the difference between supervised and unsupervised learning?

1. Supervised learning involved learning from the labeled data where we have labels for each image/data row instance. On the other hand, unsupervised learning doesn't learn on the basis of labels, as no labels are provided in the dataset.
2. Basically, in supervised learning, the aim is to learn a mapping from input data to the labels, and the losses are calculated on the basis of the difference between the predicted labels and the ground truth labels. On the other hand, in unsupervised learning, we try to learn the structure of the data based on the features in the dataset involved and cluster the data according to the structure learned by the model.
3. Examples of supervised learning: Binary or Multi-class Classification/Regression. Examples of unsupervised learning:

Clustering.

Q10 Motion and Optical Flow
10 Points

Optical flow aims to estimate how each pixel has moved in video. However, this is not always possible due to the aperture problem. Below we show four apertures with a red shape inside it. Which red shapes CANNOT be tracked correctly from this aperture?



A B C D

- [x] D
- [ ] A
- [ ] C
- [ ] B

Q11 Recognition
10 Points

Before neural networks, computer vision researchers use handcrafted features. They never operated directly on the pixel space. Give two reasons why RGB is a poor choice of feature space for object recognition.

Reason 1:

In euclidean space, the RGB color space is not perceptually uniform as the brightness data is encoded inside the 3 RGB channels themselves, and we need to separate that for correctly tracking or recognizing an object. Furthermore, the "lightness" component isn't separated. Therefore, under different illuminations, the RGB values of an object can vary significantly.

Reason 2:

Among the RGB channels, there is a high correlation. This can be problematic for training a model as there are highly correlated features present in the dataset across the 3 channels. This might increase the training time and also might cause the model to perform poorly in the testing phase. Also, it might perform poorly for tasks like scene segmentation.

Q12 Sound

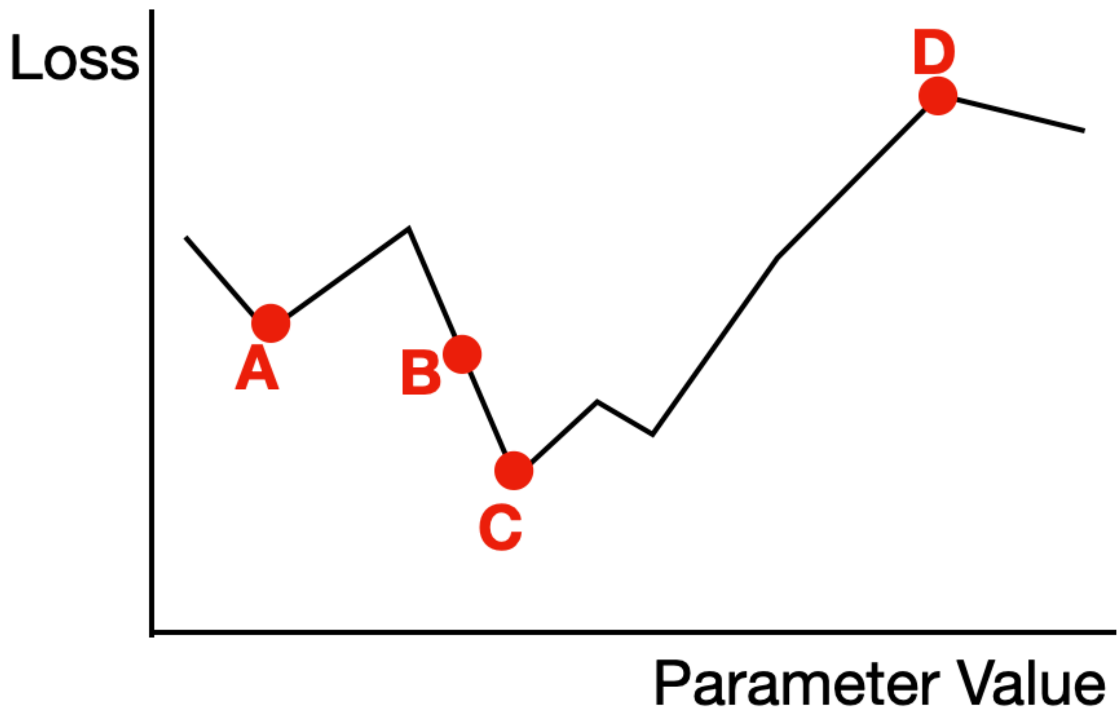
5 Points

In class, we saw examples of audio source separation, also called the cocktail party problem. Why is this a challenging problem?

Primarily, the cocktail part problem is a challenging one because it's an ill-posed problem. Furthermore, it's a permutation problem that suggests that there are a lot of permutations possible in which the audio can be decomposed.
Additionally, if there are multiple sources of sound present that emit the sound with the same frequency or amplitude, then the problem becomes even more challenging to solve.

Q13 Optimization
30 Points

Consider the following loss curve for this question:



Q13.1
5 Points

Select all of the global minimums.

- [ ] A
- [ ] D
- [x] C
- [ ] B

Q13.2
5 Points

Select all of the local minimums.

- [x] A
- [ ] D
- [ ] B
- [x] C

Q13.3
10 Points

Is this loss surface convex?

- ( ) yes
- (•) no

Q13.4
5 Points

If you start gradient descent at B, where will it converge to? Assume there is no momentum.

- ( ) B
- ( ) A
- ( ) D
- (•) C

Q13.5
5 Points

If you start gradient descent at a different position, will you always converge to the point indicated in the previous question?

- (●) no
- ( ) yes

Q14
5 Points

True or false: There exists a convolution filter that will shift an image by two pixels to the left.

- ( ) true
- (●) false

Q15 Non-euclidean Geometry
10 Points

Most neural networks operate in Euclidean space. However, recently there has been interest in training neural networks in hyperbolic space. What is the main advantage of hyperbolic space over Euclidean space?

Assume finite number of dimensions!

When considering a finite number of dimensions, we cannot embed a hierarchical tree structure in the euclidean space without any distortion. On the other hand, hyperbolic spaces can naturally embed hierarchical structures. Thus, for class predictions that have several levels of abstraction, neural networks in hyperbolic space tend to perform better and several levels of abstraction for different categories resemble a hierarchical tree structure.

Q16 3D

10 Points

Why are voxel representations not an ideal representation for 3D scenes?

Voxel representations are not an ideal representation for 3D scenes because:
1. They are very memory intensive as most of the space inside the 3D voxel is empty as the data is sparse.
2. It has another limitation related to the trade-off with voxel resolution. As depth can increase by a lot if we break the object into very small voxels, the trade-off is really hard to make.

Q17

5 Points

True or false: There exists a convolution filter that will rotate the image by 90 degrees.

- (•) false
- ( ) true

Q18 Camera Basics

5 Points

5 Points

Chromatic aberration is a lens defect that most camera lenses have. What causes it? One sentence is enough.

Chromatic Aberration is basically caused by the dispersion inherent in the optical lenses. This, in turn, is because different wavelengths of light can pass through the lens at different speeds and thus, focus differently.

Q19 Activation Functions

5 Points

Before the 2010s, most neural network researchers used the sigmoid or tanh activation function inside their neural networks. Why might this be a bad choice for gradient based optimization?

For gradient-based optimization, these activation functions might cause the problem of vanishing gradients. This might happen if, at a certain hidden layer, the gradient calculation is very close to zero, which might, in turn, make the gradients for the layer before that one (from the start of the network) even less and at some point become zero. This would make all the gradients before that zero by the chain rule. This would stop any updates for the neurons in the neural network and thus, hinder learning. This could also mean that the network could learn very slowly, or the learning might even stop. Thus, those might be bad choices for gradient-based optimization.

Q20

5 Points

True or false: There exists a convolution filter that will blur an image.

- (selected) true
- false

Q21 Regularization

10 Points

What is regularization and why would you use it when training a neural network?

Regularization helps us prevent overfitting. As the regularization term in the loss function would always add something to the gradient of the prediction so that the loss doesn't go to zero, this prevents the model from overfitting on the dataset. Additionally, L1 regularization even helps the model to sample from the features the most significant ones and tells us that it might depend on a limited number of features.

Q22 Film and Data

10 Points

Shirley cards are widely used in color film development. However, they had a major problem until 1995. What was the problem, and how was it fixed?

The major problem with the Shirley cards was that they only had fair-skinned women for a long time. Because of the nature of these cards, the color calibrations' procedure became biased toward fairer skin tones. Thus, this resulted in a poor calibration for dark-skinned tones (for people).
It was fixed by incorporating people with different skin tones in those cards so as to remove the bias from the procedure of color calibration.

Q23 Back-propagation

20 Points

Consider the neural network module that computes the following operation during the forward pass:

$$z_{t+1} = f(z_t; w) = w \cdot z_t^2$$

where z_t is the input layer and z_{t+1} is the output layer. The parameter w is learned. You can assume all variables are a scalar. In this problem, you will calculate the gradients for implementing back-propagation. Your answers can include the term $\frac{\delta \text{Loss}}{\delta z_{t+1}}$, but otherwise write out the full equation.

Q23.1

10 Points

What is the gradient $\frac{\delta \text{Loss}}{\delta w}$?

You can either type in your response, or upload a file with your answer.

If you type in your answer, you can use LaTeX by surrounding the equation with $$. For example: $$ x^2 $$ renders as x^2

$dLoss/dw = (dLoss/d(z_{t+1})) * z_t^2$
d suggests del operator (partial differentiation) here.

No files uploaded

Q23.2

10 Points

What is the gradient $\frac{\delta \text{Loss}}{\delta z_t}$?

You can either type in your response, or upload a file with your answer.

$dLoss/dz_t = 2 * (dLoss/d(z_{t+1})) * z_t * w$
d suggests del operator (partial differentiation) here.

No files uploaded

Final Quiz

UNGRADED

STUDENT

Chandan Suri

TOTAL POINTS

- / 210 pts

QUESTION 1

(no title) 5 pts

QUESTION 2

Max Pooling 5 pts

QUESTION 3

Course Evaluation 5 pts

QUESTION 4

Loss Functions 10 pts

QUESTION 5

Interpretability 10 pts

QUESTION 6

Convolution 10 pts

QUESTION 7

Fourier Transform 10 pts

QUESTION 8
(no title) 5 pts

QUESTION 9
Supervised Learning 10 pts

QUESTION 10
Motion and Optical Flow 10 pts

QUESTION 11
Recognition 10 pts

QUESTION 12
Sound 5 pts

QUESTION 13
Optimization 30 pts

13.1 (no title) 5 pts

13.2 (no title) 5 pts

13.3 (no title) 10 pts

13.4 (no title) 5 pts

13.5 (no title) 5 pts

QUESTION 14
(no title) 5 pts

QUESTION 15
Non-euclidean Geometry 10 pts

QUESTION 16
3D 10 pts

QUESTION 17
(no title) 5 pts

QUESTION 18
Camera Basics 5 pts

Camera Basics 5 pts

QUESTION 19

Activation Functions 5 pts

QUESTION 20

(no title) 5 pts

QUESTION 21

Regularization 10 pts

QUESTION 22

Film and Data 10 pts

QUESTION 23

Back-propagation 20 pts

23.1 (no title) 10 pts

23.2 (no title) 10 pts