CSE343/CSE543/ECE363/ECE563: Machine Learning Sec A (Monsoon 2023)
Quiz - 4

Date of Examination: 21.11.2023 Duration: 30 mins Total Marks: 15 marks

---

Instructions –

- Attempt all questions.
- MCQs have a single correct option.
- State any assumptions you have made clearly.
- Standard institute plagiarism policy holds.
- No evaluation without suitable justification.

---

0 marks if the option or justification of MCQs is incorrect.

1. The advantages of CNNs over ANNs for image classification are - **[1 mark]**

   (A) Location Invariance of features.
   (B) Reusable weights.
   (C) Can process sequential data.
   (D) Accepts arbitrary input size.

   (i) B and C.
   (ii) B and D.
   (iii) A and B.
   (iv) A, B and D.

   Both (iii) and (iv) will be considered. Convolution filters are invariant to translation,rotation etc during feature extraction, also the same kernel is used over multiple patches of the image, thus reusing weights. Moreover if the network has only Convolution and pooling layers then it is also invariant to input size

2. Mahalnobis distance is better than euclidean distance for clustering as it takes into consideration: **[1 mark]**

   (A) Outliers in the sample data.
   (B) Non-spherical shape of the sample data.
   (C) Correlation among the features.
   (D) Unequal size of the clusters.

   (C) Mahalnobis distance takes into account the correlation matrix while calculating the distance.

3. What is the size of the output image if a 3x3 convolutional kernel is applied over an input image of shape 32x32 ? (assume stride=1) **[1 mark]**

   (A) 30x30.
   (B) 29x29.
   (C) 28x28.
   (D) 31x31.

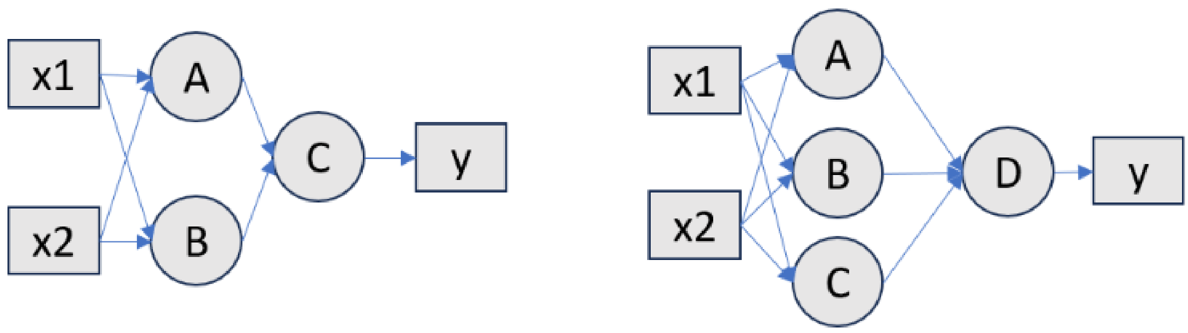


Figure 1: Diagram for Question 5

(A), If image is mxn and kernel is pxq then resultant size is (m-p+1)x(n-q+1) assuming that stride=1

4. In a CNN with two layers: 1 convolutional layer with kernel size 6 x 6 followed by a max pooling layer, the total number of trainable parameters is (the image size is 28×28): - [**1 mark**]

   1. 484
   2. 37
   3. 821
   4. 13

   B, The total number of parameters in the first layer is 6 x 6 = 36 and one bias which makes it 37. The pooling layer has no learnable parameters.

5. Consider the two MLPs as shown in the figure. They have identity activation (f(x) = x) at all the neurons. Which model has better expressive power? [**4 marks**]
   Rubric: 1.5 marks for each model output, 1 mark for final conclusion.

6. Consider the following data points: $A = (8, 8)$, $B = (14, 12)$, $C = (12, 14)$, and $D = (8, 10)$. We want to use the K-Means algorithm with Manhattan distance to cluster the data points into two clusters, C_1 and C_2. To initialize the algorithm, we consider $C_1 = \{A, C\}$ and $C_2 = \{B, D\}$. Perform the K-means clustering algorithm till it converges. What are the coordinates of the centers of clusters, and which points belong to which cluster? Hint: Manhattan distance is calculated as : [**3 marks**]

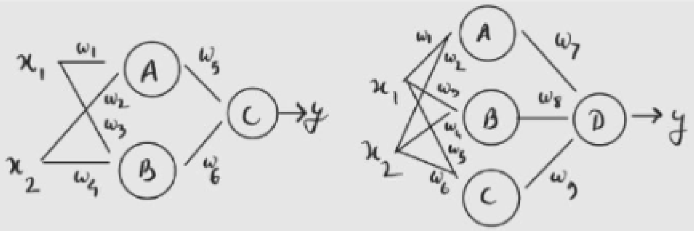
$$\text{distance}(x, y) = \sum_{i=1}^{n} |x_i - y_i|$$

   Rubric: 0.5 mark for initial clusters, 1 mark for assigning new clusters to each point, 1 mark for updating new cluster centers, 0.5 mark for next iteration to prove convergence.

7. Consider a hypothetical distance metric defined below. Does this qualify as a valid distance metric which we can use for clustering? If yes, state the reasons, else state how $d(X, Y)$ can be modified to make it a valid distance metric. [**4 marks**]

$$\text{d}(X, Y) = \log\left(\sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}\right)$$

Rubric: 1 mark for each condition and its rectification.



Output of 1st model:

$$y = w_5(w_1x_1 + w_2x_2) + w_6(w_3x_1 + w_4x_2)$$
$$= (w_1w_5 + w_3w_6)x_1 + (w_2w_5 + w_4w_6)x_2$$
$$= \alpha x_1 + \beta x_2$$

Output of 2nd model:

$$y = w_7(w_1x_1 + w_2x_2) + w_8(w_3x_1 + w_4x_2) + w_9(w_5x_1 + w_6x_2)$$
$$= (w_1w_7 + w_3w_8 + w_5w_9)x_1 + (w_2w_7 + w_4w_8 + w_6w_9)x_2$$
$$= \gamma x_1 + \delta x_2$$

Both are linear regression models with 2 features.

$\therefore$ They have same expressive power!!

Figure 2: Soln for Question 5

Answer: K-means clustering algorithm with Manhattan distance for the given data points and initial clusters:

Initialize the clusters:

C1={(8,8), (12,14)} and C2={(14,12),(8,10)}.

Calculating the centers of the initial clusters:

Center of C1: (8+12, 8+14)/2 = (10,11)

Center of C2: (14+8, 12+10)/2 = (11, 11)

Assign each point to the nearest cluster using Manhattan distance:

For point A, d(A, C1) = |8-10| +|8-11|= 5 and d(A,C2) = |8-11| +|8-11| = 6. Assign A to C1.

For point B, d(B, C1) = |14-10| +|12-11|= 5 and d(B,C2) = |14-11| +|12-11| = 4. Assign B to C2.

For point C, d(C, C1) = |12-10| +|14-11|= 5 and d(C,C2) = |12-11| +|14-11| = 4. Assign C to C2.

For point D, d(D, C1) = |8-10| +|10-11|= 3 and d(D,C2) = |8-11| +|10-11| = 4. Assign D to C1.

4. Update the clusters:

C1 = {(8,8), (8, 10)} and C2 = {(14, 12), (12, 14)}.

New centroids:

Center of C1: (8+8, 8+10)/2 = (8,9)

Center of C2: (14+12, 12+14)/2 = (13, 13)

Iteration 2: Similar to iteration 1, but cluster assignment will remain the same for all the points thus, we can conclude that the algo has converged.

Figure 3: Soln for Question 6

Soln -

It must satisfy the following four conditions -

i) Distance should be positive

$d(X,Y) <0$ if $(x_1-x_2)^2 + (y_1-y_2)^2 <1$, fail

ii) Distance can be zero only if X=Y

$d(X,Y)$ is not defined if X=Y, fail

iii) Distance must be symmetric i.e. $d(x,y) = d(y,x)$

iv) Triangle equality must hold

The first two conditions fail, trefore it doesn't qualify as a valid distance metric.

Modify $d(X,Y)$ as : $d(X,Y) = \log(1 + (x_1-x_2)^2 + (y_1-y_2)^2)$. Now each of the three criteria holds!!

i) Distance should be positive

$d(X,Y) >=0$ as $1+(x_1-x_2)^2 + (y_1-y_2)^2 >=1$ and log(something >=1)>0, pass

ii) Distance can be zero only if X=Y

$d(X,Y) = 0$ as $1+(x_1-x_2)^2 + (y_1-y_2)^2 = 1$ and $\log(1) = 0$, pass

iii) Distance must be symmetric i.e. $d(x,y) = d(y,x)$

True as $(x_1-x_2)^2 + (y_1-y_2)^2 = (x_2-x_1)^2 + (y_2-y_1)^2$, pass

iv) Triangle equality must hold

If ABC is a triangle, $d(A,B) + d(B,C) > d(A,C)$

$a = (x_1-x_2)^2 + (y_1-y_2)^2$ is just euclidean distance and it follows triangle inequality

Adding 1 to all the distances will still follow the triangle inequality

Finally taking log for each of the distances will still follow triangle inequality as log is a monotonic increasing function

Figure 4: Soln for Question 7