Quiz 4 Rubrics

Syllabus: SVM, K-means, K- median
No. of Questions - 10 MCQs
Date: 30th November 2021

Q1.

You are given a labeled binary classification data set with N data points and D features. Suppose that N < D. In training an SVM on this data set, which of the following kernels is likely to be most appropriate?
(a) Linear kernel (b) Quadratic kernel (c) Higher-order polynomial kernel (d) RBF kernel

Solution: a) linear kernel

Q2.

Which of the following statements are true? Check all that apply.

A. If you are training multi-class SVMs with the one-vs-all method, it is not possible to use a kernel.
B. Suppose you have 2D input examples (ie,$x^{(i)} \in R^2$). The decision boundary of the SVM (with the linear kernel) is a straight line.
C. If the data are linearly separable, an SVM using a linear kernel will return the same parameters θ regardless of the chosen value of C (i.e., the resulting value of θ does not depend on C) .
D. The maximum value of the Gaussian kernel (i.e., $sim(x,l^{(1)})$) is 1.

Solution: B and D

Q3.

What is the value of the question mark in the below image? Show your calculation with cell wise operations and the final result.

INPUT

1	1	1	0	0
0	1	1	1	0
0	0	1	1	1
0	0	1	1	0
0	1	1	0	0

FILTER

1	0	1
0	1	0
1	0	1

CONVOLVED FEATURE

?		

Solution: 4
The convolution operation has an output which is a linear combination of the filter and the image. Therefore, output value is 1*1+1*0+1*1+0*0+1*1+1*0+0*1+0*0+1*1 = 4.

Q4.

True / False

a) The support vectors are the subset of data points that determines the max-margin separator?
b) The max-margin separator is a non-linear combination of the support vectors?

Solution: a) True b) False

Q5.

Which of the following are correct statements with respect to SVMs and neural networks?
A. Unlike a neural network, SVM has the capability to transform the input space to an infinite-dimensional space.
B. An SVM may learn to apply non-linear transformations on the data which is not possible with a neural network.
C. An SVM should not get stuck in local minima, while a neural network might.
D. None of the above

Solution: A, C

Using appropriate kernel functions, SVM may map the data to an infinite-dimensional space which is not possible in the case of a neural network. The objective function of SVM is convex and therefore, the local minimum is the same as the global minimum.

Q6.

When the C parameter is set to infinite, which of the following holds true?
A) The optimal hyperplane if exists, will be the one that completely separates the data
B) The soft-margin classifier will separate the data
C) None of the above
Solution: A

Q7.

A CNN model has the following operations:
Input (size) = [216X216X3]
Filter(size) = [8X8X3]
Stride = 4
Max pooling 2X2 with size 2

What will be the output size after the given operations?

a. [208 X 208]
b. [53 X 53]
c. [27 X 27]
d. [52 X 52]

Solution. [27 x 27]
- Convolution : Dimensions = (216 - 8)/4 + 1 = 53 (stride 4)
- Max Pooling (2x2) with stride 2: Dimensions = (53 - 2)/2 + 1 = 27

Q8.

Which all can be said to be true when convolution of 1x1 is used for a CNN?

a. To increase the feature maps number, a projection made by 1x1 can be used.
b. To create a linear projection of a stack of feature maps, a 1x1 filter can be used.
c. 1x1 can act as channel-wise pooling and act as dimensionality reduction.
d. All of the above.

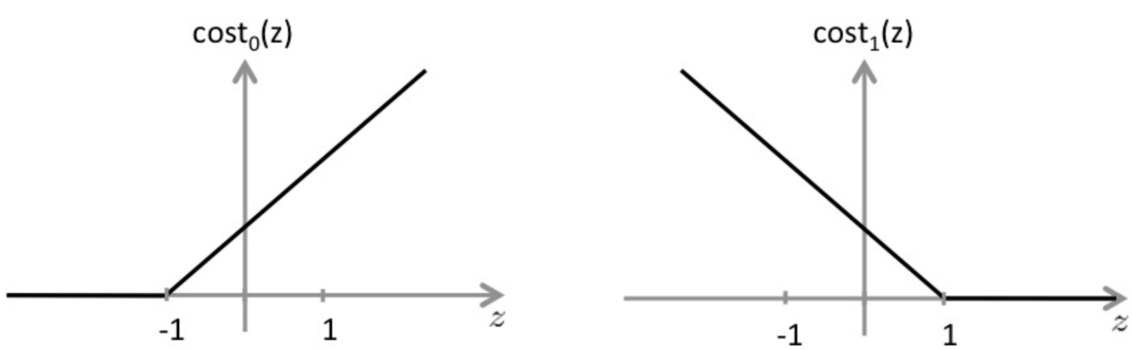
Ans. All of the above.

Q9.

The SVM solves

$$\min_\theta \; C\sum_{i=1}^{m} y^{(i)}\mathrm{cost}_1(\theta^T x^{(i)}) + (1-y^{(i)})\mathrm{cost}_0(\theta^T x^{(i)}) + \sum_{j=1}^{n}\theta_j^2$$

where the functions $\mathrm{cost}_0(z)$ and $\mathrm{cost}_1(z)$ look like this:



The first term in the objective is:

$$C\sum_{i=1}^{m} y^{(i)}\mathrm{cost}_1(\theta^T x^{(i)}) + (1-y^{(i)})\mathrm{cost}_0(\theta^T x^{(i)}).$$

This first term will be zero if two of the following four conditions hold true. Which are the two conditions that would guarantee that this term equals zero?

a. For every example with y^(i) = 1 we have a $\theta^T x^{\wedge(i)} >= 1$.
b. For every example with y^(i) = 0 we have a $\theta^T x^{\wedge(i)} <= 0$.
c. For every example with y^(i) = 1 we have a $\theta^T x^{\wedge(i)} >= 0$.
d. For every example with y^(i) = 0 we have a $\theta^T x^{\wedge(i)} <= -1$.

Solution: A and D

Q10.

Suppose we have a dataset with two classes in 2-D space with equal no. of samples from both classes. The positive samples of the dataset are taken from the points on the curve x_1^2 + x_2^2 = 5 and negative samples are taken from the curve x_1^2 + x_2^2 = 10. What will be the decision boundary that would be obtained by training SVM (with suitable hyperparameters chosen) when we use
I. Linear kernel
Ii. Polynomial kernel of order 2
Iii. RBF kernels

a) Cannot be determined, x_2 = sqrt(5) , x_1^2 + x_2^2 = (sqrt (5) + sqrt(10))/2
b) x_2 = sqrt(5) , x_1^2 + x_2^2 = ((sqrt (5) + sqrt(10))/2)^2, x_1^2 + x_2^2 =((sqrt (5) + sqrt(10))/2)^2
c) X_1+x_2 = sqrt (5) , x_1+x_2 = sqrt(10), x=15
d) X_1+x_2 = sqrt (5) , x_1+x_2 = sqrt(10),x_1+x_2 = sqrt(10)

Solution: B
i) one of the possible answers: x_2 = sqrt(5) line parallel to x axis passing through the centroid of the circles.
ii) x_1^2 + x_2^2 = (sqrt (5) + sqrt(10))/2
iii) same as ii