

25 Questions to test a Data Scientist on Support Vector Machines

ANKIT GUPTA, OCTOBER 5, 2017



**The Text Analytics Demand is Expected to Grow
More Than 200% By 2022**

Introduction

You can think of machine learning algorithms as an armory packed with axes, sword and blades. You have various tools, but you ought to learn to use them at the right time. As an analogy, think of 'Regression' as a sword capable of slicing and dicing data efficiently, but incapable of dealing with highly complex data. On the contrary, 'Support Vector Machines' is like a sharp knife – it works on smaller datasets, but on them, it can be much more stronger and powerful in building models.

This skilltest was specially designed for you to test your knowledge on SVM techniques and its applications. More than 550 people registered for the test. If you are one of those who missed out on this skill test, here are the questions and solutions.

Here is the [leaderboard](#) for the participants who took the test.

Helpful Resources

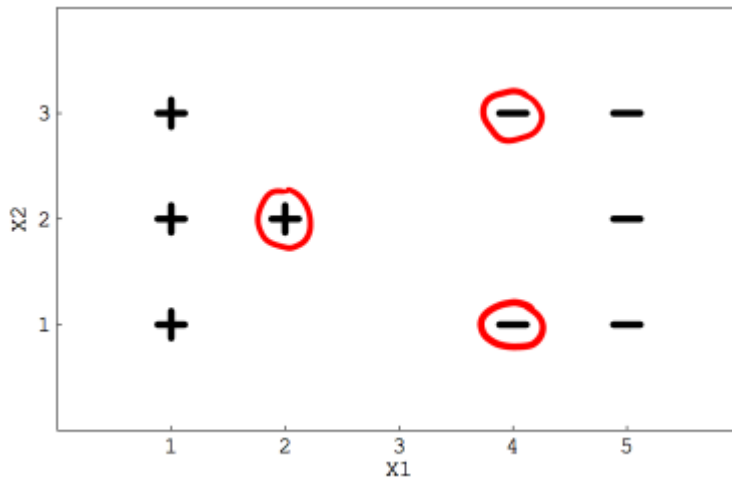
Here are some resources to get in depth knowledge in the subject.

- [Essentials of Machine Learning Algorithms \(with Python and R Codes\)](#).
- [Understanding Support Vector Machine algorithm from examples \(along with code\)](#).

Skill test Questions and Answers

Question Context: 1 – 2

Suppose you are using a Linear SVM classifier with 2 class classification problem. Now you have been given the following data in which some points are circled red that are representing support vectors.



1) If you remove the following any one red points from the data. Does the decision boundary will change?

- A) Yes
- B) No

Solution: A

These three examples are positioned such that removing any one of them introduces slack in the constraints. So the decision boundary would completely change.

2) [True or False] If you remove the non-red circled points from the data, the decision boundary will change?

- A) True
- B) False

Solution: B

On the other hand, rest of the points in the data won't affect the decision boundary much.

3) What do you mean by generalization error in terms of the SVM?

- A) How far the hyperplane is from the support vectors
- B) How accurately the SVM can predict outcomes for unseen data
- C) The threshold amount of error in an SVM

Solution: B

Generalisation error in statistics is generally the out-of-sample error which is the measure of how accurately a model can predict values for previously unseen data.

4) 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

At such a high level of misclassification penalty, soft margin will not hold existence as there will be no room for error.

5) What do you mean by a hard margin?

- A) The SVM allows very low error in classification
- B) The SVM allows high amount of error in classification
- C) None of the above

Solution: A

A hard margin means that an SVM is very rigid in classification and tries to work extremely well in the training set, causing overfitting.

6) The minimum time complexity for training an SVM is $O(n^2)$. According to this fact, what sizes of datasets are not best suited for SVM's?

- A) Large datasets
- B) Small datasets
- C) Medium sized datasets
- D) Size does not matter

Solution: A

Datasets which have a clear classification boundary will function best with SVM's.

7) The effectiveness of an SVM depends upon:

- A) Selection of Kernel
- B) Kernel Parameters
- C) Soft Margin Parameter C
- D) All of the above

Solution: D

The SVM effectiveness depends upon how you choose the basic 3 requirements mentioned above in such a way that it maximises your efficiency, reduces error and overfitting.

8) Support vectors are the data points that lie closest to the decision surface.

- A) TRUE
- B) FALSE

Solution: A

They are the points closest to the hyperplane and the hardest ones to classify. They also have a direct bearing on the location of the decision surface.

9) The SVM's are less effective when:

- A) The data is linearly separable
- B) The data is clean and ready to use
- C) The data is noisy and contains overlapping points

Solution: C

When the data has noise and overlapping points, there is a problem in drawing a clear hyperplane without misclassifying.

10) Suppose you are using RBF kernel in SVM with high Gamma value. What does this signify?

- A) The model would consider even far away points from hyperplane for modeling
- B) The model would consider only the points close to the hyperplane for modeling
- C) The model would not be affected by distance of points from hyperplane for modeling
- D) None of the above

Solution: B

The gamma parameter in SVM tuning signifies the influence of points either near or far away from the hyperplane.

For a low gamma, the model will be too constrained and include all points of the training dataset, without really capturing the shape.

For a higher gamma, the model will capture the shape of the dataset well.

11) The cost parameter in the SVM means:

- A) The number of cross-validations to be made
- B) The kernel to be used
- C) The tradeoff between misclassification and simplicity of the model
- D) None of the above

Solution: C

The cost parameter decides how much an SVM should be allowed to “bend” with the data. For a low cost, you aim for a smooth decision surface and for a higher cost, you aim to classify more points correctly. It is also simply referred to as the cost of misclassification.

12)

Suppose you are building a SVM model on data X. The data X can be error prone which means that you should not trust any specific data point too much. Now think that you want to build a SVM model which has quadratic kernel function of polynomial degree 2 that uses Slack variable C as one of it's hyper parameter. Based upon that give the answer for following question.

What would happen when you use very large value of C(C->infinity)?

Note: For small C was also classifying all data points correctly

- A) We can still classify data correctly for given setting of hyper parameter C
- B) We can not classify data correctly for given setting of hyper parameter C
- C) Can't Say
- D) None of these

Solution: A

For large values of C, the penalty for misclassifying points is very high, so the decision boundary will perfectly separate the data if possible.

13) What would happen when you use very small C (C~0)?

- A) Misclassification would happen
- B) Data will be correctly classified
- C) Can't say
- D) None of these

Solution: A

The classifier can maximize the margin between most of the points, while misclassifying a few points, because the penalty is so low.

14) If I am using all features of my dataset and I achieve 100% accuracy on my training set, but ~70% on validation set, what should I look out for?

- A) Underfitting
- B) Nothing, the model is perfect
- C) Overfitting

Solution: C

If we're achieving 100% training accuracy very easily, we need to check to verify if we're overfitting our data.

15) Which of the following are real world applications of the SVM?

- A) Text and Hypertext Categorization
- B) Image Classification
- C) Clustering of News Articles
- D) All of the above

Solution: D

SVM's are highly versatile models that can be used for practically all real world problems ranging from regression to clustering and handwriting recognitions.

Question Context: 16 – 18

Suppose you have trained an SVM with linear decision boundary after training SVM, you correctly infer that your SVM model is under fitting.

16) Which of the following option would you more likely to consider iterating SVM next time?

- A) You want to increase your data points
- B) You want to decrease your data points
- C) You will try to calculate more variables
- D) You will try to reduce the features

Solution: C

The best option here would be to create more features for the model.

17) Suppose you gave the correct answer in previous question. What do you think that is actually happening?

1. We are lowering the bias
2. We are lowering the variance

3. We are increasing the bias
4. We are increasing the variance

- A) 1 and 2
- B) 2 and 3
- C) 1 and 4
- D) 2 and 4

Solution: C

Better model will lower the bias and increase the variance

18) In above question suppose you want to change one of it's(SVM) hyperparameter so that effect would be same as previous questions i.e model will not under fit?

- A) We will increase the parameter C
- B) We will decrease the parameter C
- C) Changing in C don't effect
- D) None of these

Solution: A

Increasing C parameter would be the right thing to do here, as it will ensure regularized model

19) We usually use feature normalization before using the Gaussian kernel in SVM. What is true about feature normalization?

1. We do feature normalization so that new feature will dominate other
2. Some times, feature normalization is not feasible in case of categorical variables
3. Feature normalization always helps when we use Gaussian kernel in SVM

- A) 1
- B) 1 and 2
- C) 1 and 3
- D) 2 and 3

Solution: B

Statements one and two are correct.

Question Context: 20-22

Suppose you are dealing with 4 class classification problem and you want to train a SVM model on the data for that you are using One-vs-all method. Now answer the below questions?

20) How many times we need to train our SVM model in such case?

- A) 1
- B) 2
- C) 3
- D) 4

Solution: D

For a 4 class problem, you would have to train the SVM at least 4 times if you are using a one-vs-all method.

21) Suppose you have same distribution of classes in the data. Now, say for training 1 time in one vs all setting the SVM is taking 10 second. How many seconds would it require to train one-vs-all method end to end?

- A) 20
- B) 40
- C) 60
- D) 80

Solution: B

It would take $10 \times 4 = 40$ seconds

22) Suppose your problem has changed now. Now, data has only 2 classes. What would you think how many times we need to train SVM in such case?

- A) 1
- B) 2
- C) 3
- D) 4

Solution: A

Training the SVM only one time would give you appropriate results

Question context: 23 – 24

Suppose you are using SVM with linear kernel of polynomial degree 2, Now think that you have applied this on data and found that it perfectly fit the data that means, Training and testing accuracy is 100%.

23) Now, think that you increase the complexity(or degree of polynomial of this kernel). What would you think will happen?

- A) Increasing the complexity will overfit the data
- B) Increasing the complexity will underfit the data
- C) Nothing will happen since your model was already 100% accurate
- D) None of these

Solution: A

Increasing the complexity of the data would make the algorithm overfit the data.

24) In the previous question after increasing the complexity you found that training accuracy was still 100%. According to you what is the reason behind that?

1. Since data is fixed and we are fitting more polynomial term or parameters so the algorithm starts memorizing everything in the data
2. Since data is fixed and SVM doesn't need to search in big hypothesis space

- A) 1
- B) 2
- C) 1 and 2
- D) None of these

Solution: C

Both the given statements are correct.

25) What is/are true about kernel in SVM?

1. Kernel function map low dimensional data to high dimensional space
2. It's a similarity function

- A) 1
- B) 2
- C) 1 and 2
- D) None of these

Solution: C

Both the given statements are correct.

Overall Distribution

Below is the distribution of the scores of the participants: