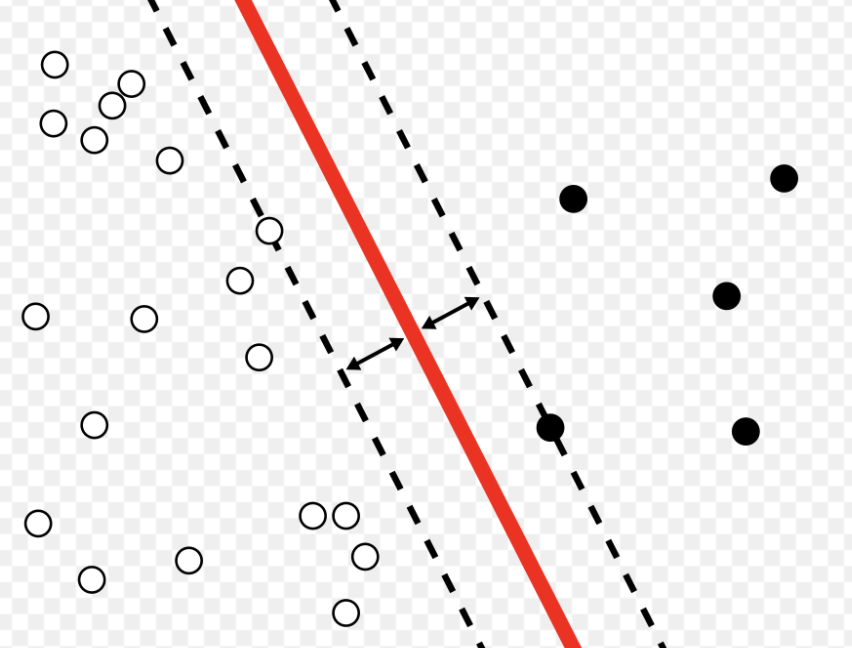
1. What is the underlying concept of Support Vector Machines?

Ans: [The fundamental idea behind Support Vector Machines (SVMs) is to find a **hyperplane that maximally separates the different classes in the training data**](https://www.bing.com/ck/a?!&&p=a4a87c458a5cd725JmltdHM9MTY4NjYxNDQwMCZpZ3VpZD0yMjc0ZWY0My05NzBjLTZkMzAtMTg0My1mZmU5OTY5NzZjZTAmaW5zaWQ9NTczOA&ptn=3&hsh=3&fclid=2274ef43-970c-6d30-1843-ffe996976ce0&psq=What+is+the+underlying+concept+of+Support+Vector+Machines+%3f&u=a1aHR0cHM6Ly93d3cuZ2Vla3Nmb3JnZWVrcy5vcmcvaW50cm9kdWN0aW9uLXRvLXN1cHBvcnQtdmVjdG9yLW1hY2hpbmVzLXN2bS8&ntb=1). [SVM is a linear model for classification and regression problems that can solve linear and non-linear problems](https://www.bing.com/ck/a?!&&p=af5a37abc90dfabcJmltdHM9MTY4NjYxNDQwMCZpZ3VpZD0yMjc0ZWY0My05NzBjLTZkMzAtMTg0My1mZmU5OTY5NzZjZTAmaW5zaWQ9NTc0MQ&ptn=3&hsh=3&fclid=2274ef43-970c-6d30-1843-ffe996976ce0&psq=What+is+the+underlying+concept+of+Support+Vector+Machines+%3f&u=a1aHR0cHM6Ly9tdWxsLW92ZXJ0aGluZy5jb20vd2hhdC1pcy10aGUtZnVuZGFtZW50YWwtaWRlYS1iZWhpbmQtc3VwcG9ydC12ZWN0b3ItbWFjaGluZXMv&ntb=1). [SVM can also be used to classify text into text that belongs to a particular topic and text that does not belong to the topic](https://www.bing.com/ck/a?!&&p=9057a06abeb2a63eJmltdHM9MTY4NjYxNDQwMCZpZ3VpZD0yMjc0ZWY0My05NzBjLTZkMzAtMTg0My1mZmU5OTY5NzZjZTAmaW5zaWQ9NTc0Mw&ptn=3&hsh=3&fclid=2274ef43-970c-6d30-1843-ffe996976ce0&psq=What+is+the+underlying+concept+of+Support+Vector+Machines+%3f&u=a1aHR0cHM6Ly9nZWVrZmxhcmUuY29tL3N1cHBvcnQtdmVjdG9yLW1hY2hpbmUv&ntb=1)[3](https://www.bing.com/ck/a?!&&p=f37497555a31a39bJmltdHM9MTY4NjYxNDQwMCZpZ3VpZD0yMjc0ZWY0My05NzBjLTZkMzAtMTg0My1mZmU5OTY5NzZjZTAmaW5zaWQ9NTc0NA&ptn=3&hsh=3&fclid=2274ef43-970c-6d30-1843-ffe996976ce0&psq=What+is+the+underlying+concept+of+Support+Vector+Machines+%3f&u=a1aHR0cHM6Ly9nZWVrZmxhcmUuY29tL3N1cHBvcnQtdmVjdG9yLW1hY2hpbmUv&ntb=1). [This is achieved by first converting and representing the text data into a dataset with several features](https://www.bing.com/ck/a?!&&p=6c1538d277c3db7fJmltdHM9MTY4NjYxNDQwMCZpZ3VpZD0yMjc0ZWY0My05NzBjLTZkMzAtMTg0My1mZmU5OTY5NzZjZTAmaW5zaWQ9NTc0NQ&ptn=3&hsh=3&fclid=2274ef43-970c-6d30-1843-ffe996976ce0&psq=What+is+the+underlying+concept+of+Support+Vector+Machines+%3f&u=a1aHR0cHM6Ly9nZWVrZmxhcmUuY29tL3N1cHBvcnQtdmVjdG9yLW1hY2hpbmUv&ntb=1)[**3**](https://www.bing.com/ck/a?!&&p=36575af5c631fee3JmltdHM9MTY4NjYxNDQwMCZpZ3VpZD0yMjc0ZWY0My05NzBjLTZkMzAtMTg0My1mZmU5OTY5NzZjZTAmaW5zaWQ9NTc0Ng&ptn=3&hsh=3&fclid=2274ef43-970c-6d30-1843-ffe996976ce0&psq=What+is+the+underlying+concept+of+Support+Vector+Machines+%3f&u=a1aHR0cHM6Ly9nZWVrZmxhcmUuY29tL3N1cHBvcnQtdmVjdG9yLW1hY2hpbmUv&ntb=1).

1. What is the concept of a support vector?

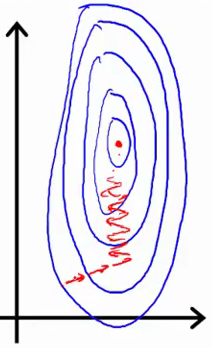
Ans: Support vectors are observations that lie on the margin surrounding the data-separating hyperplane. Since the margin defines the minimum distance observations should have from the plane, the observations that lie on the margin impact the orientation and position of the hyperplane.

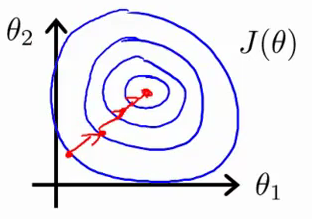


3. When using SVMs, why is it necessary to scale the inputs?

Ans:

Feature scaling is a general trick applied to optimization problems (not just SVM). The underline algorithm to solve the optimization problem of SVM is gradient descend. SVMs try to fit the largest possible “street” between the classes (see the first answer), so if the training set is not scaled, the SVM will tend to neglect small features.

Suppose you have only two parameters and one of the parameters can take a relatively large range of values. Then the contour of the cost function can look like very tall and skinny ovals (see blue ovals below). Your gradients (the path of gradient is drawn in red) could take a long time and go back and forth to find the optimal solution.  


Instead if your scaled your feature, the contour of the cost function might look like circles; then the gradient can take a much more straight path and achieve the optimal point much faster. 

1. When an SVM classifier classifies a case, can it output a confidence score? What about a percentage chance?

Ans: [An SVM classifier can output the distance between the test instance and the decision boundary as a confidence score, but it cannot output probabilities for each class](https://www.bing.com/ck/a?!&&p=49780ebdd03e8987JmltdHM9MTY4NzEzMjgwMCZpZ3VpZD0yMjc0ZWY0My05NzBjLTZkMzAtMTg0My1mZmU5OTY5NzZjZTAmaW5zaWQ9NTY1Ng&ptn=3&hsh=3&fclid=2274ef43-970c-6d30-1843-ffe996976ce0&psq=4.%09When+an+SVM+classifier+classifies+a+case%2c+can+it+output+a+confidence+score%3f+What+about+a+percentage+chance%3f&u=a1aHR0cHM6Ly93d3cuYW5hbHl0aWNzdmlkaHlhLmNvbS9ibG9nLzIwMjEvMDUvdG9wLTE1LXF1ZXN0aW9ucy10by10ZXN0LXlvdXItZGF0YS1zY2llbmNlLXNraWxscy1vbi1zdm0v&ntb=1)[1](https://www.bing.com/ck/a?!&&p=ddcb1c52f47a8a1eJmltdHM9MTY4NzEzMjgwMCZpZ3VpZD0yMjc0ZWY0My05NzBjLTZkMzAtMTg0My1mZmU5OTY5NzZjZTAmaW5zaWQ9NTY1Nw&ptn=3&hsh=3&fclid=2274ef43-970c-6d30-1843-ffe996976ce0&psq=4.%09When+an+SVM+classifier+classifies+a+case%2c+can+it+output+a+confidence+score%3f+What+about+a+percentage+chance%3f&u=a1aHR0cHM6Ly93d3cuYW5hbHl0aWNzdmlkaHlhLmNvbS9ibG9nLzIwMjEvMDUvdG9wLTE1LXF1ZXN0aW9ucy10by10ZXN0LXlvdXItZGF0YS1zY2llbmNlLXNraWxscy1vbi1zdm0v&ntb=1)[2](https://www.bing.com/ck/a?!&&p=fdcf6faa20a051bdJmltdHM9MTY4NzEzMjgwMCZpZ3VpZD0yMjc0ZWY0My05NzBjLTZkMzAtMTg0My1mZmU5OTY5NzZjZTAmaW5zaWQ9NTY1OA&ptn=3&hsh=3&fclid=2274ef43-970c-6d30-1843-ffe996976ce0&psq=4.%09When+an+SVM+classifier+classifies+a+case%2c+can+it+output+a+confidence+score%3f+What+about+a+percentage+chance%3f&u=a1aHR0cHM6Ly9hbmphbmNzZTA3Lm1lZGl1bS5jb20vc29sdXRpb24tb2YtdGhlLWV4ZXJjaXNlLWNoYXB0ZXItNS1zdXBwb3J0LXZlY3Rvci1tYWNoaW5lLTZkMzBkMzExMWI0Zg&ntb=1)[3](https://www.bing.com/ck/a?!&&p=648d5595960690b5JmltdHM9MTY4NzEzMjgwMCZpZ3VpZD0yMjc0ZWY0My05NzBjLTZkMzAtMTg0My1mZmU5OTY5NzZjZTAmaW5zaWQ9NTY1OQ&ptn=3&hsh=3&fclid=2274ef43-970c-6d30-1843-ffe996976ce0&psq=4.%09When+an+SVM+classifier+classifies+a+case%2c+can+it+output+a+confidence+score%3f+What+about+a+percentage+chance%3f&u=a1aHR0cHM6Ly9zdGF0cy5zdGFja2V4Y2hhbmdlLmNvbS9xdWVzdGlvbnMvMTYzMjk3L2hvdy10by1nZXQtcHJvYmFiaWxpdHktZnJvbS10aGUtY29uZmlkZW5jZS1zY29yZS1pbi1zdm0&ntb=1). [One standard way to obtain a “probability” out of an SVM is to use Platt scaling, which is available in many decent SVM implementations. In the binary case, the probabilities are calibrated using Platt scaling: logistic regression on the SVM’s scores, fi](https://www.bing.com/ck/a?!&&p=a7d714c6b4f2993bJmltdHM9MTY4NzEzMjgwMCZpZ3VpZD0yMjc0ZWY0My05NzBjLTZkMzAtMTg0My1mZmU5OTY5NzZjZTAmaW5zaWQ9NTY2MA&ptn=3&hsh=3&fclid=2274ef43-970c-6d30-1843-ffe996976ce0&psq=4.%09When+an+SVM+classifier+classifies+a+case%2c+can+it+output+a+confidence+score%3f+What+about+a+percentage+chance%3f&u=a1aHR0cHM6Ly9tbXVyYXRhcmF0LmdpdGh1Yi5pby8yMDE5LTEwLTEyL3Byb2JhYmlsaXN0aWMtb3V0cHV0LW9mLXN2bQ&ntb=1)t by additional cross validation on training data

5. Should you train a model on a training set with millions of instances and hundreds of features using the primal or dual form of the SVM problem?

Ans: This question applies only to linear SVMs since kernelized can only use the dual form. The computational complexity of the primal form of the SVM problem is proportional to the number of training instances m, while the computational complexity of the dual form is proportional to a number between m2 and m3. So if there are millions of instances, you should definitely use the primal form, because the dual form will be much too slow.

6. Let's say you've used an RBF kernel to train an SVM classifier, but it appears to underfit the training collection. Is it better to raise or lower (gamma)? What about the letter C?

Ans: If an SVM classifier trained with an RBF kernel underfits the training set, there might be too much regularization. To decrease it, you need to increase gamma or C (or both).

7. To solve the soft margin linear SVM classifier problem with an off-the-shelf QP solver, how should the QP parameters (H, f, A, and b) be set?

Ans: Let’s call the QP parameters for the hard-margin problem H′, f′, A′ and b′. The QP parameters for the soft-margin problem have m additional parameters (np = n + 1 + m) and m additional constraints (nc = 2m). They can be defined like so:

* H is equal to H′, plus m columns of 0s on the right and m rows of 0s at the bottom
* f is equal to f′ with m additional elements, all equal to the value of the hyperparameter C.
* b is equal to b′ with m additional elements, all equal to 0.
* A is equal to A′, with an extra m × m identity matrix Im appended to the right

8. On a linearly separable dataset, train a Linear SVC. Then, using the same dataset, train an SVC and an SGD Classifier. See if you can get them to make a model that is similar to yours.

Ans: [ML-Assignment/Assign\_20\_Q8.ipynb at main · tanujadhope/ML-Assignment (github.com)](https://github.com/tanujadhope/ML-Assignment/blob/main/Assign_20_Q8.ipynb)

9. On the MNIST dataset, train an SVM classifier. You'll need to use one-versus-the-rest to assign all 10 digits because SVM classifiers are binary classifiers. To accelerate up the process, you might want to tune the hyperparameters using small validation sets. What level of precision can you achieve?

Ans: [ML-Assignment/Assign\_20\_Q9.ipynb at main · tanujadhope/ML-Assignment (github.com)](https://github.com/tanujadhope/ML-Assignment/blob/main/Assign_20_Q9.ipynb)

10. On the California housing dataset, train an SVM regressor.

Ans: [ML-Assignment/Assign\_20\_Q10.ipynb at main · tanujadhope/ML-Assignment (github.com)](https://github.com/tanujadhope/ML-Assignment/blob/main/Assign_20_Q10.ipynb)