Machine Learning with R: Support Vector Machine Classification (SVM)

Shaila Jamal

Data Analysis Support Assistant, DASH, McMaster Library

Ph.D. Candidate, School of Earth, Environment and Society, McMaster University

November 07, 2022

McMaster University | DASH

McMaster University | Library

McMaster University sits on the traditional Territories of the Mississauga and Haudenosaunee Nations, and within the lands protected by the "Dish With One Spoon" wampum agreement.

Session Recording and Privacy

This session is being recorded with the intention of being shared publicly via the web for future audiences.

In respect of your privacy, participant lists will not be shared outside of this session, nor will question or chat transcripts.

Questions asked via the chat box will be read by the facilitator without identifying you. Note that you may be identifiable when asking a question during the session in an audio or visual format.

Code of Conduct

The DASH program and the McMaster University Library are committed to fostering a supportive and inclusive environment for its presenters and participants.

As a participant in this session, you agree to support and help cultivate an experience that is collaborative, respectful, and inclusive, as well as free of harassment, discrimination, and oppression. We reserve the right to remove participants who exhibit harassing, malicious, or persistently disruptive behaviour.

Please refer to our code of conduct webpage for more information:

scds.ca/events/code-of-conduct/

Certificate Program

The Sherman Centre offers a Certificate of Completion that rewards synchronous participation at 7 workshops. We also offer concentrations in Data Analysis and Visualization, Digital Scholarship, and Research Data Management.

Learn more about the Certificate Program: https://scds.ca/certificate-program

If you would like to be considered for the certificate, verify your participation in this form: https://u.mcmaster.ca/verification

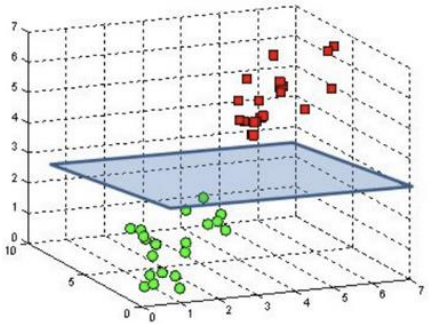
At an unspecified point during the workshop, a code will be read aloud. This is the answer to the third question of the form.

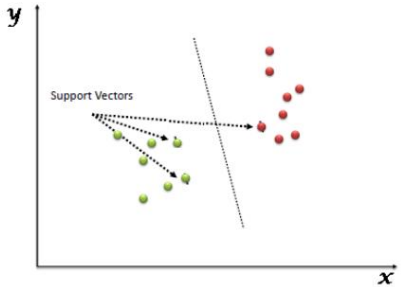
Recordings of the workshops

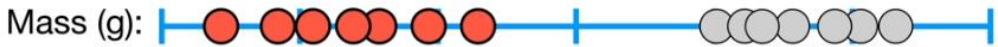
For workshop recordings, check here: Search the Online Learning Catalogue | Sherman Centre for Digital Scholarship (scds.ca)

What is SVM?

- “is a supervised machine learning algorithm which is mainly used to classify data into different classes.”
- “makes use of a hyperplane which acts like a decision boundary between the various classes.”
- “trains on a set of labeled data. SVM studies the labeled training data and then classifies any new input data depending on what it learned in the training phase.”
- widely used across disciplines such as healthcare, natural language processing, signal processing applications, and speech & image recognition fields.





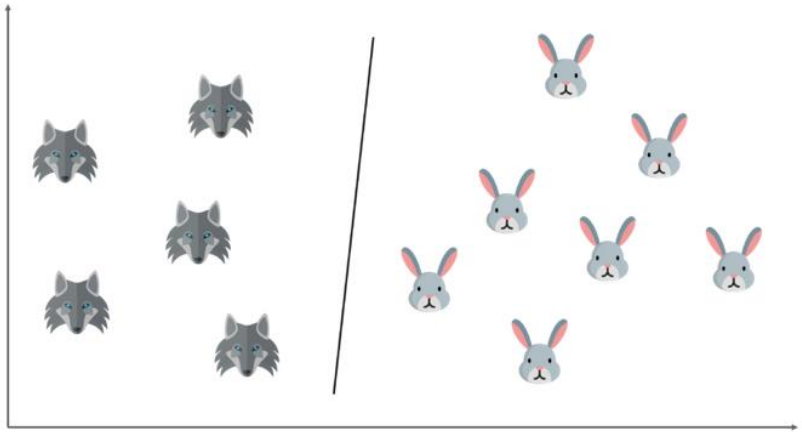


Source: Lateef, J. “Support Vector Machine In R: Using SVM To Predict Heart Diseases”. May 15, 2020, accessed on September 29, 2022. https://www.edureka.co/blog/support-vector-machine-in-r/

What is SVM?

A decision boundary, i.e. a hyperplane between the classes in order to separate them or classify them optimally

How does SVM work? – Support Vector Machine In R

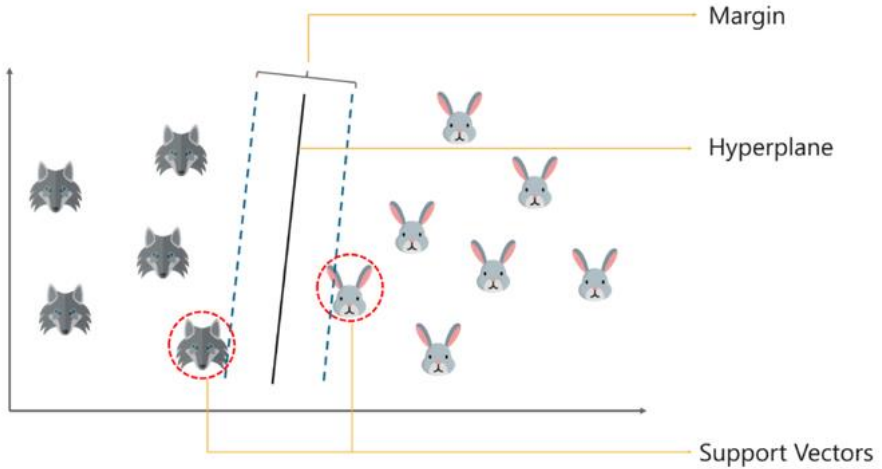


How does SVM work? – Support Vector Machine In R

Source: Lateef, J. "Support Vector Machine In R: Using SVM To Predict Heart Diseases". May 15, 2020, accessed on September 29, 2022. https://www.edureka.co/blog/support-vector-machine-in-r/

What is a Support Vector in SVM Classifier?

"These closest data points to the hyperplane are known as support vectors. And that's where the name comes from, support vector machine."

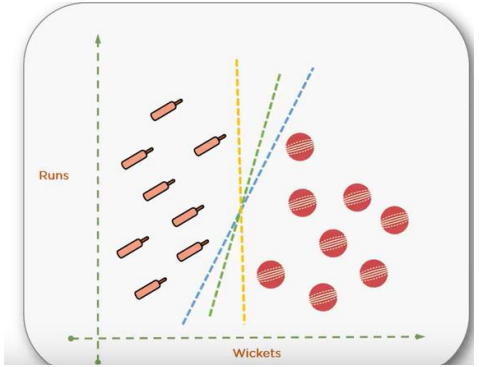


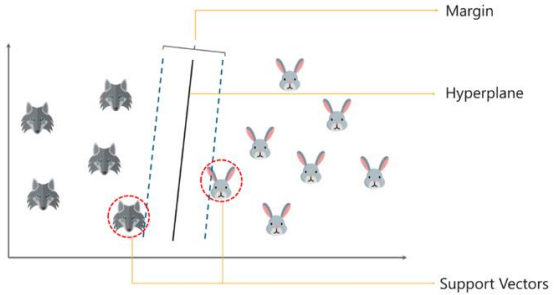
How does SVM work? – Support Vector Machine In R

Source: Lateef, J. "Support Vector Machine In R: Using SVM To Predict Heart Diseases". May 15, 2020, accessed on September 29, 2022. https://www.edureka.co/blog/support-vector-machine-in-r/

What are the other elements in SVM?

- “The hyperplane is drawn based on these support vectors and an optimum hyperplane will have a maximum distance from each of the support vectors. And this distance between the hyperplane and the support vectors is known as the margin.”
- “SVM is used to classify data by using a hyperplane, such that the distance between the hyperplane and the support vectors is maximum.”

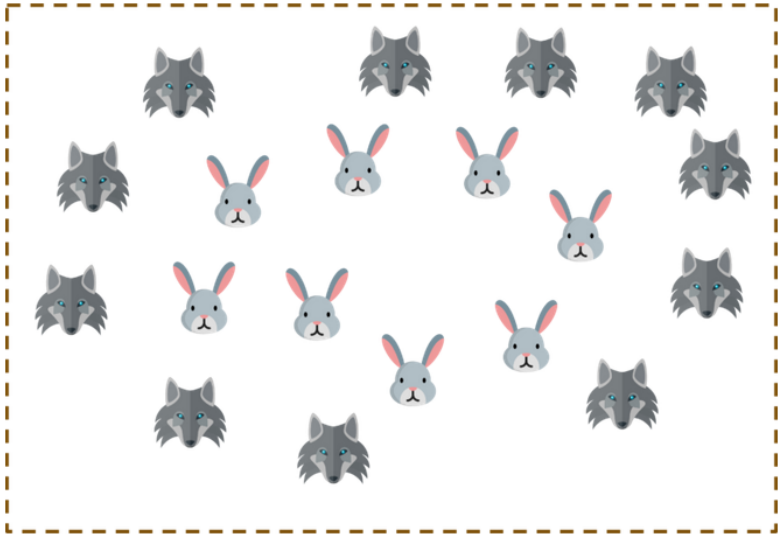




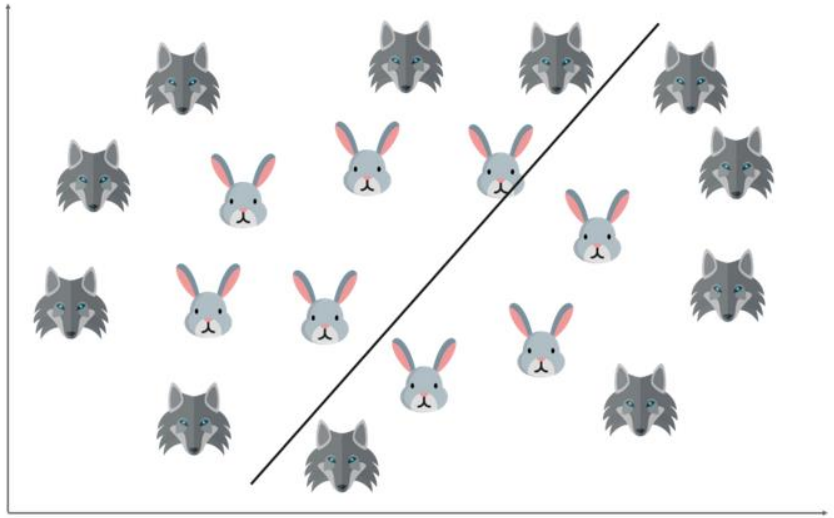
How does SVM work? – Support Vector Machine In R

Source: Lateef, J. “Support Vector Machine In R: Using SVM To Predict Heart Diseases”. May 15, 2020, accessed on September 29, 2022. https://www.edureka.co/blog/support-vector-machine-in-r/

Non-linear SVM?



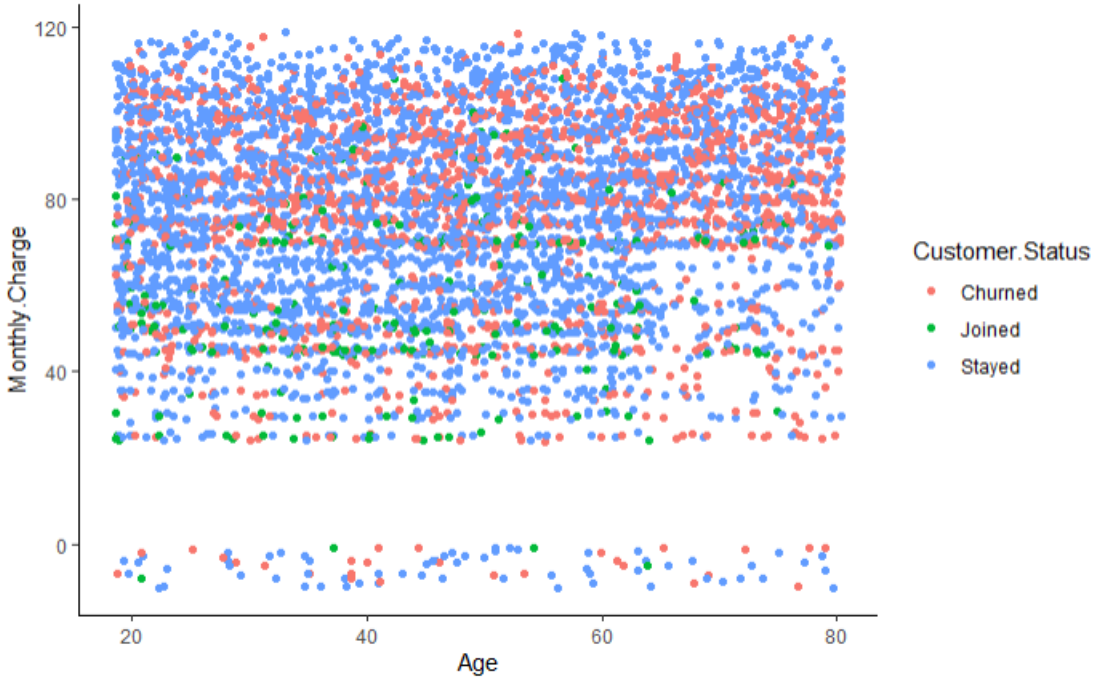
How does SVM work? – Support Vector Machine In R



How does SVM work? – Support Vector Machine In R

Source: Lateef, J. "Support Vector Machine In R: Using SVM To Predict Heart Diseases". May 15, 2020, accessed on September 29, 2022. https://www.edureka.co/blog/support-vector-machine-in-r/

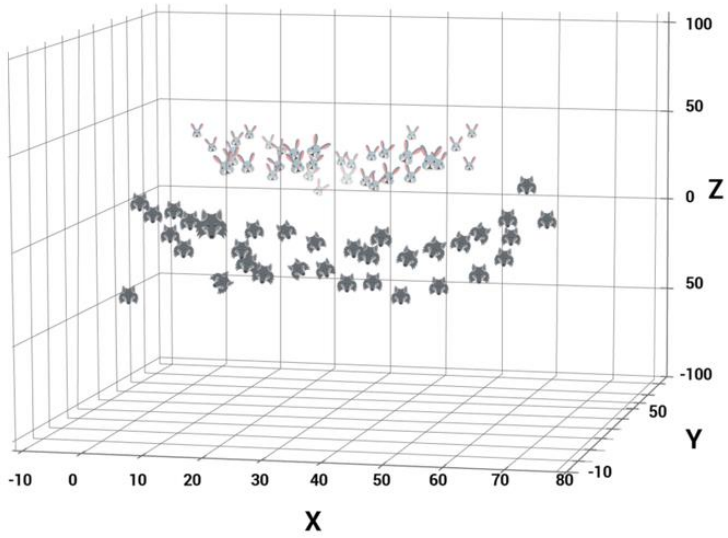
Non-linear SVM?



Customer Status' in the telecom customer dataset

What is a kernel trick?

- “SVM can be used for classifying non-linear data by using the kernel trick.”
- The kernel trick means transforming data into another dimension that has a clear dividing margin between classes of data.
- After which you can easily draw a hyperplane between the various classes of data.
- Transforming the two variables x and y into a new feature space involving a new variable z. Basically, in this example, we’re visualizing the data on a 3-dimensional space.”



Non-linear Support Vector Machine – Support Vector Machine In R

- Among different types of kernel, radial basis (RBF) is generally the most popular one.

Source: Lateef, J. “Support Vector Machine In R: Using SVM To Predict Heart Diseases”. May 15, 2020, accessed on September 29, 2022. https://www.edureka.co/blog/support-vector-machine-in-r/

SVM - Standardization

- “All kernel methods are based on distance.
- Scaling of the variables is required.
- If we do not standardize our variables to comparable ranges, the variable with the largest range will completely dominate in the computation of the kernel matrix. For example, we have two variables - X1 and X2. Values of variable X1 lies between 0 and 100 whereas values of X2 lies in range of 100 and 10000. In this case, variable X2 would dominate variable X1.
- The z-score and min-max are the two popular methods to standardize variables.”

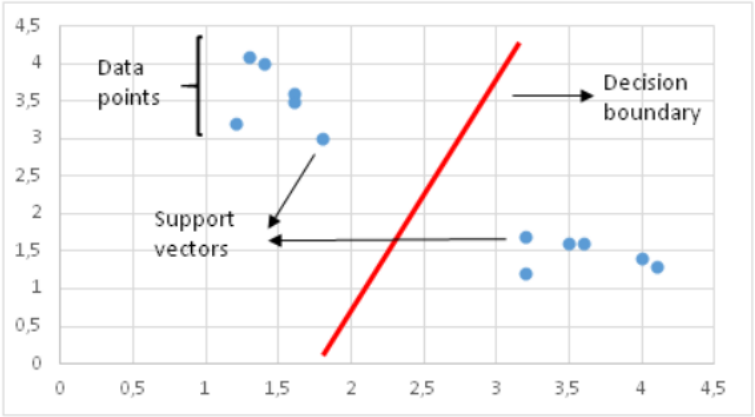
Source: Bhalla, D. “Support Vector Machine Simplified using R”., accessed on September 29, 2022. https://www.listendata.com/2017/01/support-vector-machine-in-r-tutorial.html

Tuning hyperparameters

- SVM have some parameters that can be adjusted which are called hyperparameters.
- We need to set hyperparameters before we train the models.
- Hyperparameters are very critical in building robust and accurate models. They help us find the balance between bias and variance and thus, prevent the model from overfitting or underfitting.
- To be able to adjust the hyperparameters, we need to understand what they mean and how they change a model."

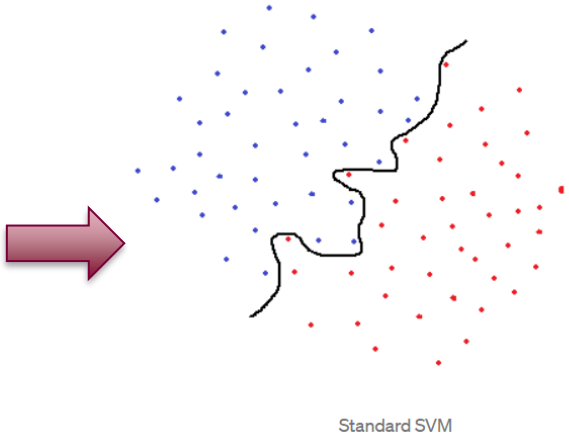
Source: Yildirim, S. "Hyperparameter Tuning for Support Vector Machines — C and Gamma Parameters". May 31, 2020. , accessed on October 16, 2022. https://towardsdatascience.com/hyperparameter-tuning-for-support-vector-machines-c-and-gamma-parameters-6a5097416167

Tuning hyperparameters – C and gamma



Linearly separable data points

But real data is noisy and not linearly separable



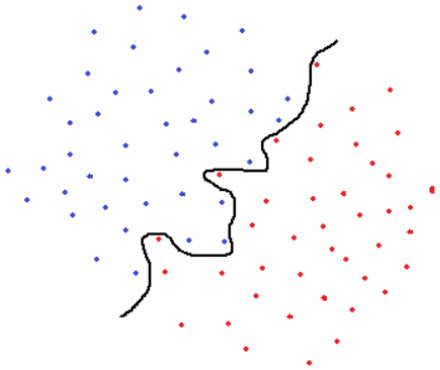
Standard SVM

"This is a too specific classification and highly likely to end up overfitting.

Source: Yildirim, S. "Hyperparameter Tuning for Support Vector Machines — C and Gamma Parameters". May 31, 2020. , accessed on October 16, 2022. https://towardsdatascience.com/hyperparameter-tuning-for-support-vector-machines-c-and-gamma-parameters-6a5097416167

Tuning hyperparameters – C and gamma

- An overfit SVM achieves a high accuracy with training set but will not perform well on new, previously unseen examples.
- This model would be very sensitive to noise and even very small changes in data point values may change the classification results.
- The SVM that uses this black line as a decision boundary is not generalized well to this dataset.

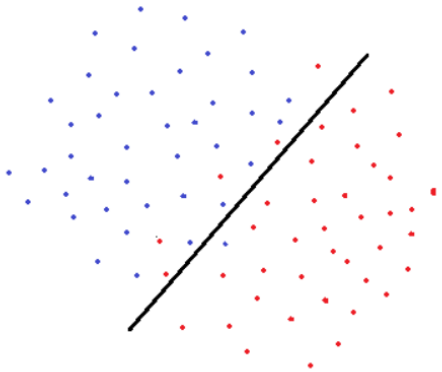


Standard SVM

Source: Yildirim, S. “Hyperparameter Tuning for Support Vector Machines — C and Gamma Parameters”. May 31, 2020. , accessed on October 16, 2022. https://towardsdatascience.com/hyperparameter-tuning-for-support-vector-machines-c-and-gamma-parameters-6a5097416167

Tuning hyperparameters – C and gamma

- To overcome this issue, in 1995, Cortes and Vapnik, came up with the idea of “soft margin” SVM which allows some examples to be misclassified or be on the wrong side of decision boundary.
- Soft margin SVM often result in a better generalized model.
- There are some misclassified points but we end up having a more generalized model.

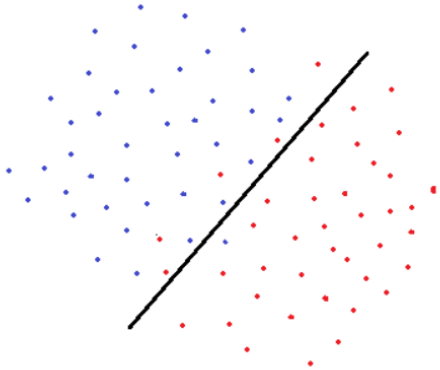


Soft margin SVM

Source: Yildirim, S. “Hyperparameter Tuning for Support Vector Machines — C and Gamma Parameters”. May 31, 2020. , accessed on October 16, 2022. https://towardsdatascience.com/hyperparameter-tuning-for-support-vector-machines-c-and-gamma-parameters-6a5097416167

Tuning hyperparameters – C and gamma

- A soft margin SVM tries to solve an optimization problem with the following goals:
    - Increase the distance of decision boundary to classes (or support vectors)
    - Maximize the number of points that are correctly classified in the training set

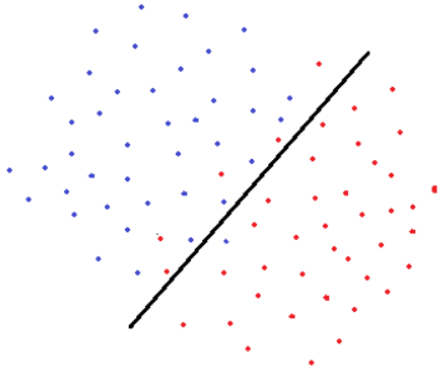


Soft margin SVM

Source: Yildirim, S. “Hyperparameter Tuning for Support Vector Machines — C and Gamma Parameters”. May 31, 2020. , accessed on October 16, 2022. https://towardsdatascience.com/hyperparameter-tuning-for-support-vector-machines-c-and-gamma-parameters-6a5097416167

Tuning hyperparameters – C and gamma

- There is obviously a trade-off between these two goals.
- Decision boundary might have to be very close to one particular class to correctly label all data points in training set. However, in this case, accuracy on test dataset might be lower because decision boundary is too sensitive to noise and to small changes in the independent variables.
- On the other hand, a decision boundary might be placed as far as possible to each class with the expense of some misclassified exceptions. This trade-off is controlled by c parameter.

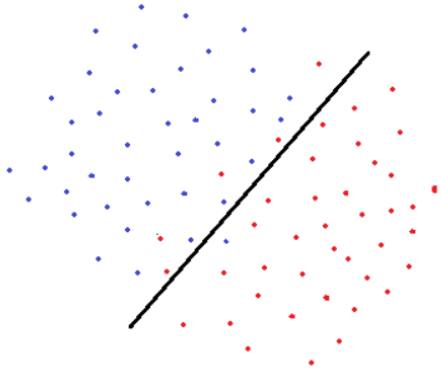


Soft margin SVM

Source: Yildirim, S. “Hyperparameter Tuning for Support Vector Machines — C and Gamma Parameters”. May 31, 2020. , accessed on October 16, 2022. https://towardsdatascience.com/hyperparameter-tuning-for-support-vector-machines-c-and-gamma-parameters-6a5097416167

Tuning hyperparameters – C and gamma

- “C parameter adds a penalty for each misclassified data point.
    - If c is **small**, the **penalty for misclassified** points is low so a **decision boundary with a large margin** is chosen at the expense of a greater number of misclassifications.
    - If c is **large**, SVM tries to minimize the number of misclassified examples due to **high penalty** which results in a decision boundary with a smaller margin.
- Penalty is not same for all misclassified examples. It is directly proportional to the distance to decision boundary.”



Soft margin SVM

Source: Yildirim, S. “Hyperparameter Tuning for Support Vector Machines — C and Gamma Parameters”. May 31, 2020. , accessed on October 16, 2022. https://towardsdatascience.com/hyperparameter-tuning-for-support-vector-machines-c-and-gamma-parameters-6a5097416167

Tuning hyperparameters – C and gamma

Minimize the error

gamma <dbl>	cost <dbl>	error <dbl>
0.5	4	0.2581992
1.0	4	0.2665830
2.0	4	0.2794938
0.5	8	0.2593319
1.0	8	0.2715625
2.0	8	0.2901335
0.5	16	0.2652235
1.0	16	0.2747325
2.0	16	0.3003242

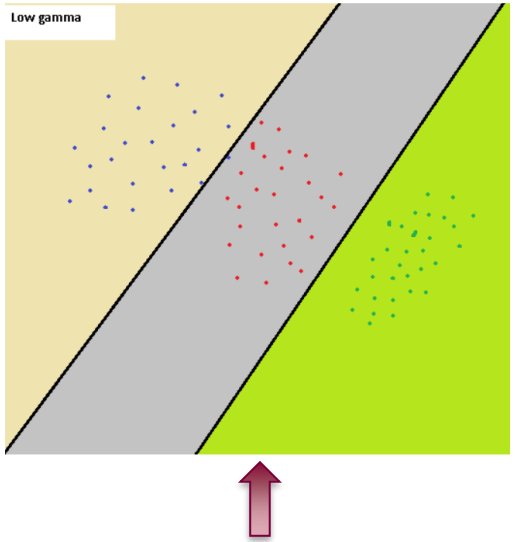
Source: Yildirim, S. “Hyperparameter Tuning for Support Vector Machines — C and Gamma Parameters”. May 31, 2020. , accessed on October 16, 2022. https://towardsdatascience.com/hyperparameter-tuning-for-support-vector-machines-c-and-gamma-parameters-6a5097416167

Tuning hyperparameters – C and gamma

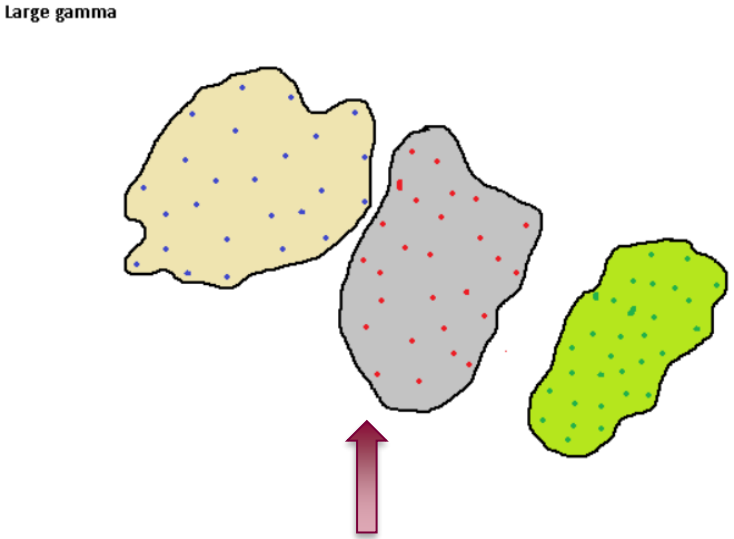
- Gamma is used when data points that are not linearly separable and are transformed using kernel functions so that they become linearly separable.
- One of the commonly used kernel functions is radial basis function (RBF).
- Gamma parameter of RBF controls the distance of influence of a single training point --- how far the influence of a single training example reaches, with low values meaning ‘far’ and high values meaning ‘close’.
- Low values of gamma indicates a large similarity radius which results in more points being grouped together.
- For high values of gamma, the points need to be very close to each other in order to be considered in the same group (or class). Therefore, models with very large gamma values tend to overfit.

Source: Yildirim, S. “Hyperparameter Tuning for Support Vector Machines — C and Gamma Parameters”. May 31, 2020. , accessed on October 16, 2022. https://towardsdatascience.com/hyperparameter-tuning-for-support-vector-machines-c-and-gamma-parameters-6a5097416167

Tuning hyperparameters – C and gamma



As the gamma decreases, the regions separating different classes get more generalized.



A small noise may cause a data point to fall out of a class. Large gamma values are likely to end up in overfitting.

Source: Yildirim, S. “Hyperparameter Tuning for Support Vector Machines — C and Gamma Parameters”. May 31, 2020. , accessed on October 16, 2022. https://towardsdatascience.com/hyperparameter-tuning-for-support-vector-machines-c-and-gamma-parameters-6a5097416167

Tuning hyperparameters – C and gamma

- For a linear kernel, we just need to optimize the c parameter.
- However, if we want to use an RBF kernel, both c and gamma parameter need to optimized simultaneously.
- If gamma is large, the effect of c becomes negligible. If gamma is small, c affects the model just like how it affects a linear model.
- Typical values for c and gamma are as follows. However, specific optimal values may exist depending on the application:

$$0.0001 < \text{gamma} < 10 \quad \text{and} \quad 0.1 < c < 100$$

Source: Yildirim, S. “Hyperparameter Tuning for Support Vector Machines — C and Gamma Parameters”. May 31, 2020. , accessed on October 16, 2022. https://towardsdatascience.com/hyperparameter-tuning-for-support-vector-machines-c-and-gamma-parameters-6a5097416167

Materials are available at: shorturl.at/ajuW8

https://drive.google.com/drive/folders/1Al9IZ-Hs5gt0NJ3sPe-nrehuSeL6bYI2?usp=sharing

Thank you!

Questions: jamals16@mcmaster.ca