|  |  |
| --- | --- |
| Name Of The Student | Vaishnavi G |
| Internship Project Topic | Build a Classification Model for Drug Trials Dataset |
| Name of the Organization | TCS iON |
| Name of the Industry Mentor | Himdweep Walia |
| Name of the Institute | SRM Institute of Science and Technology |

|  |  |  |
| --- | --- | --- |
| Date | Day # | Hours Spent |
| 7/11/2022 | 27 | 5 hours |
| Activities done during the day:  **Support Vector Machine (SVM) Algorithm Works In Machine Learning:**  **What Is Support Vector Machine (Svm)?**    A support vector machine is a machine learning model that is able to generalise between two different classes if the set of labelled data is provided in the training set to the algorithm. The main function of the SVM is to check for that hyperplane that is able to distinguish between the two classes.  There can be many hyperplanes that can do this task but the objective is to find that hyperplane that has the highest margin that means maximum distances between the two classes, so that in future if a new data point comes that is two be classified then it can be classified easily.  **Types of SVM**  **SVM can be of two types:**  **Linear SVM:**  Linear SVM is used for linearly separable data, which means if a dataset can be classified into two classes by using a single straight line, then such data is termed as linearly separable data, and classifier is used called as Linear SVM classifier.  **Non-linear SVM:**  Non-Linear SVM is used for non-linearly separated data, which means if a dataset cannot be classified by using a straight line, then such data is termed as non-linear data and classifier used is called as Non-linear SVM classifier.  **How Does Svm Works?**    **1. Linearly Separable Data**    Let us understand the working of SVM by taking an example where we have two classes that are shown is the below image which are a class A: Circle & class B: Triangle. Now, we want to apply the SVM algorithm and find out the best hyperplane that divides the both classes.  IMG_256  SVM takes all the data points in consideration and gives out a line that is called ‘Hyperplane’ which divides both the classes. This line is termed as ‘Decision boundary’. Anything that falls in circle class will belong to the class A and vice-versa.  IMG_256  There can be many hyperplanes that you can see but the best hyper plane that divides the two classes would be the hyperplane having a large distance from the hyperplane from both the classes. That is the main motive of SVM to find such best hyperplanes.  IMG_256  There can be different dimensions which solely depends upon the features we have. It is tough to visualize when the features are more than 3.  IMG_256  Consider we have two classes that are red and yellow class A and B respectively. We need to find the best hyperplane between them that divides the two classes.  IMG_256  Soft margin permits few of the above data points to get misclassified. Also,it tries to make the balance back and forth between finding a hyperplane that attempts to make less misclassifications and maximize the margin.    **2. Linearly Non-separable Data**  If the data is non linearly separable as shown in the above figure then SVM makes use of kernel tricks to make it linearly separable. The concept of transformation of non-linearly separable data into linearly separable is called Cover’s theorem - “given a set of training data that is not linearly separable, with high probability it can be transformed into a linearly separable training set by projecting it into a higher-dimensional space via some non-linear transformation”. Kernel tricks help in projecting data points to the higher dimensional space by which they became relatively more easily separable in higher-dimensional space.  IMG_256  **Advantages of SVM:**   * Effective in high dimensional cases * Its memory efficient as it uses a subset of training points in the decision function called support vectors * Different kernel functions can be specified for the decision functions and its possible to specify custom kernels   **The followings are important concepts in SVM −**   * **Support Vectors −** Datapoints that are closest to the hyperplane is called support vectors. Separating line will be defined with the help of these data points. * **Hyperplane −** As we can see in the above diagram, it is a decision plane or space which is divided between a set of objects having different classes. * **Margin −** It may be defined as the gap between two lines on the closet data points of different classes. It can be calculated as the perpendicular distance from the line to the support vectors. Large margin is considered as a good margin and small margin is considered as a bad margin.   Reference:  <https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_classification_algorithms_support_vector_machine.htm/> | | |