Name : Ahmed Gamal Hamid Zein

1. Who proposed the idea with SVM?

theory proposed by Vapnik (1982, 1995) and Chervonenkis (1974). Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier (although methods such as Platt scaling exist to use SVM in a probabilistic classification setting). SVM maps training examples to points in space so as to maximise the width of the gap between the two categories. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall.

In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.

History of SVM

- The original SVM algorithm was invented by Vladimir N. Vapnik and Alexey Ya. Chervonenkis in 1963.
- In 1992, Bernhard Boser, Isabelle Guyon and Vladimir Vapnik suggested a way to create nonlinear classifiers by applying the kernel trick to maximum-margin hyperplanes.
- The "soft margin" incarnation, as is commonly used software packages, was proposed by Corinna Cortes and Vapnik in 1993 and published in 1995.

2. What is the motivation?

The goal of SVM is to divide the datasets into classes to find a maximum marginal hyperplane (MMH). 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.

And to find a hyperplane in an N-dimensional space (N — the number of features) that distinctly classifies the data points.

3. The usage of kernels in SVM?

A kernel is a function used in SVM for helping to solve problems. They provide shortcuts to avoid complex calculations. The amazing thing about kernel is that we can go to higher dimensions and perform smooth calculations with the help of it. We can go up to an infinite number of dimensions using kernels.

And to take data as input and transform it into the required form. Different SVM algorithms use different types of kernel functions. These functions can be different types. For example linear, nonlinear, polynomial, radial basis function (RBF), and sigmoid.

4. What rules that should be fulfilled to implement a kernel function?

Kernel functions must be continuous, symmetric, and most preferably should have a positive (semi-) definite Gram matrix. Kernels which are said to satisfy the Mercer's theorem are positive semi-definite, meaning their kernel matrices have only non-negative Eigen values

5. Comparison between kernel types

- Linear Kernel
- It is the most basic type of kernel, usually one dimensional in nature. It proves to be the best function when there are lots of features. The linear kernel is mostly preferred for text-classification problems as most of these kinds of classification problems can be linearly separated.
- Polynomial Kernel
- It is a more generalized representation of the linear kernel. It is not as preferred as other kernel functions as it is less efficient and accurate.
- Sigmoid Kernel
- It is mostly preferred for neural networks. This kernel function is similar to a two-layer perceptron model of the neural network, which works as an activation function for neurons.
- Gaussian Kernel
- It is a commonly used kernel. It is used when there is no prior knowledge of a given dataset.
- ANOVA kernel
- It is also known as a radial basis function kernel. It usually performs well in multidimensional regression problems.

6. How to choose correct Kernel for an ML problem?

Always try the linear kernel first, simply because it's so much faster and can yield great results in many cases (specifically high dimensional problems). If the linear kernel fails, in general your best bet is an RBF kernel. They are known to perform very well on a large variety of problems.

7-Extending SVC for multi-class classification

Classification. SVC , NuSVC and LinearSVC are classes capable of performing binary and multi-class classification on a dataset. SVC and NuSVC are similar methods, but accept slightly different sets of parameters and have different mathematical formulations (see section Mathematical formulation).

8. The Concept of VC dimension

The VC dimension of {f(α)} is the maximum number of. training points that can be shattered by {f(α)} For example, the VC dimension of a set of oriented lines in R2 is three. In general, the VC dimension of a set of oriented hyperplanes in Rn is n+1

And if there exists a set of n points that can be shattered by the classifier and there is no set of n+1 points that can be shattered by the classifier, then the VC dimension of the classifier is n.

9. The Curse of Dimensionality

the curse of dimensionality is the problem caused by the exponential increase in volume associated with adding extra dimensions to Euclidean space.

The curse of dimensionality basically means that the error increases with the increase in the number of features. It refers to the fact that algorithms are harder to design in high dimensions and often have a running time exponential in the dimensions. A higher number of dimensions theoretically allow more information to be stored, but practically it rarely helps due to the higher possibility of noise and redundancy in the real-world data.