**Support Vector Regression**

* In SVR, along with a regression line, we also pass a tube surrounding the Regression Line.
* The mentioned tube has a width of epsilon, and the width is measured vertically along the axis, and not perpendicular to the tube.
* We can think of the tube as a margin of error we are allowing our model to have, and not care about any errors within the tube.
* At the same time, we could have points outside the E-Insensitive Tube, and for that we do care about the error. And the error will be measured as the distance between the point and the tube itself, not the regression line.
* When labeling the slack errors, if it’s above the Tube, there’s no star raised to it, it it’s below the tube, there is a star raised to it.
* And support vector regression just cares about the errors outside the tube.
* SVR gives us the flexibility to define how much error is acceptable in our model and will find an appropriate line (or hyperplane in higher dimension) to fit the data.

Diagram

Description automatically generated with medium confidence

**Why is it called a Support Vector Regression?**

* Basically, every point on the plot is a vector. It can be represented as a vector in a 2D space, or multiple 2D space if we have more features.
* But the points outside the tube, the ones that are highlighted in red, they are the support vector. Because they are dictating how the tube is created. So basically, they are supporting for formatting the structure of the tube. And that’s why they’re called support vectors. Thus, it is a Support Vector Regression.

**Note** –

* When implementing SVR, we need to apply feature scaling to the dataset elements as there is no explicit equation of the dependent variable with respect to the features (independent variables), and mostly there are no features multiplying the features, and therefore not compensating with lower values for features taking high values.
* We apply feature scaling to the dependent and independent variables if the independent variable is not binary. Sometimes the independent variable could be something like salary, which would be beyond the range of feature scaling. For example – when looking at some database where we predict whether a customer will purchase something or not, we encode the purchase decision as categorical variable and it results it 0 and 1, which is in the range of the feature scaled results. And thus, we don’t need to feature scale the independent variable. But sometimes, the dependent variable could be something like salary, which needs to be feature scaled, as it is beyond the range of feature scaled values of the independent variables.
* Unlike other regression models, where dependent variables and features have an explicit relation – the outcome of the dependent variable is dependent on independent variables, SVR model doesn’t have the same case – the equation has an implicit relation between independent and dependent variables.
* We can choose between a linear kernel and a non-linear kernel when implementing SVR.

One of the disadvantages of SVR is that it has a hard time catching outliers to make predictions.