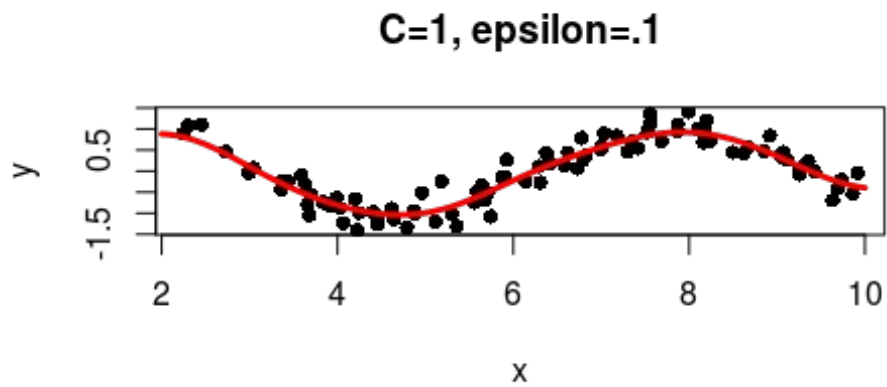## 1. a) Loading data and observing the scatter plot
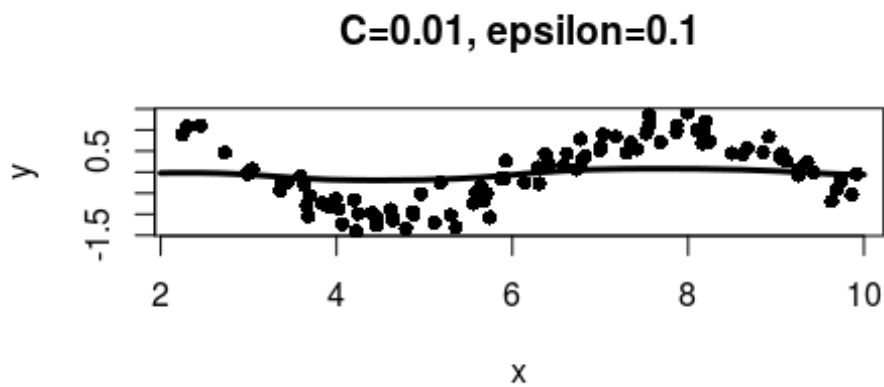
Scatter plot
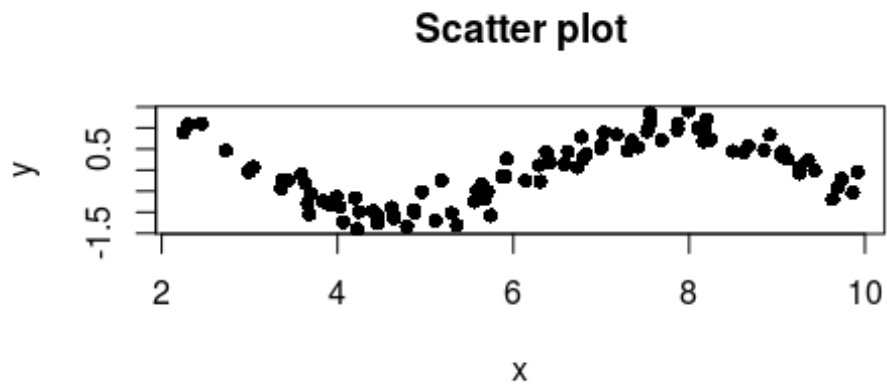
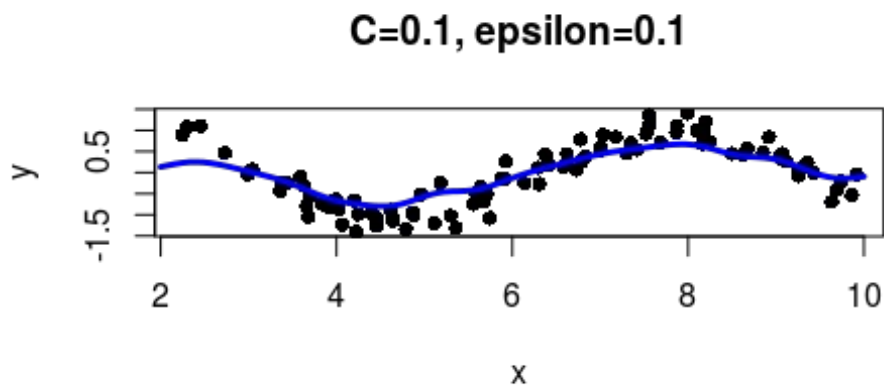Fixing epsilon = 1 and changing Cost parameter

Red line in the following figure shows the best fit to the data.

C=1, epsilon=.1
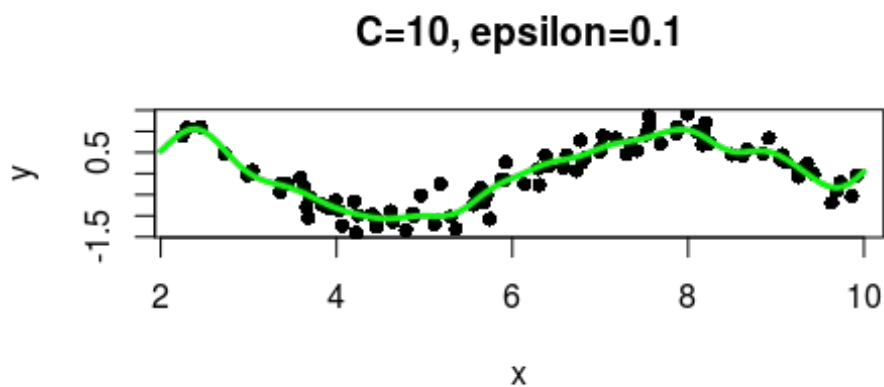
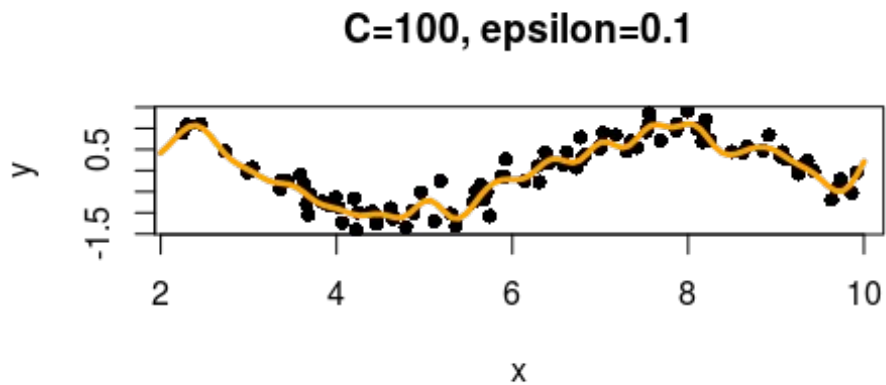When the cost parameter is very low the model under fits the data

C=0.01, epsilon=0.1

when the cost increases it becomes little better.

C=0.1, epsilon=0.1

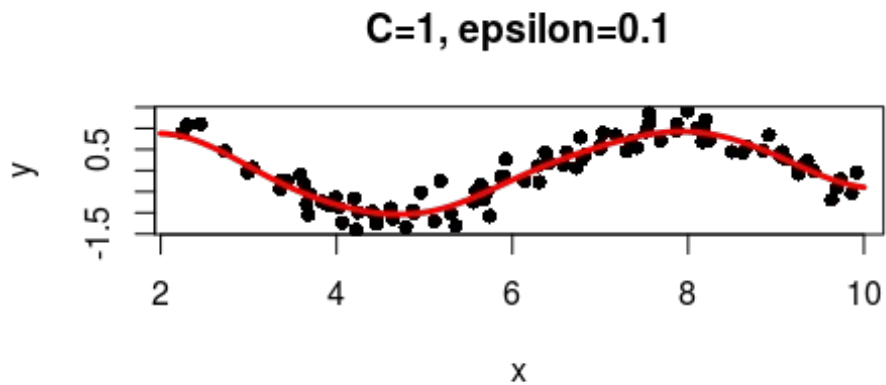And when increased more than the best value of 1, it started to become over fitted.

C=10, epsilon=0.1

Very high cost parameter overfitted the data as below.
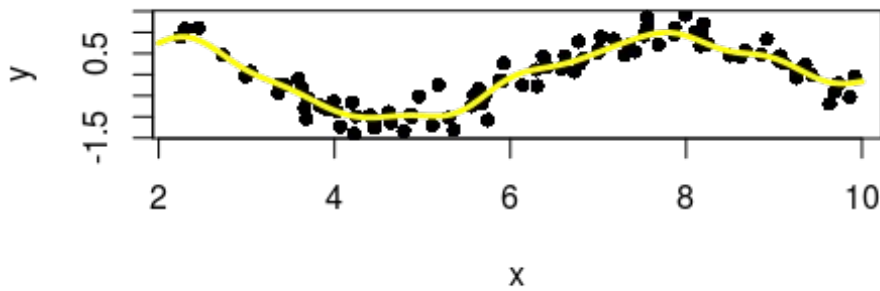
C=100, epsilon=0.1

Same as the cost parameter it is possible to change the epsilon which changes the number of support vectors selected to fit the model.

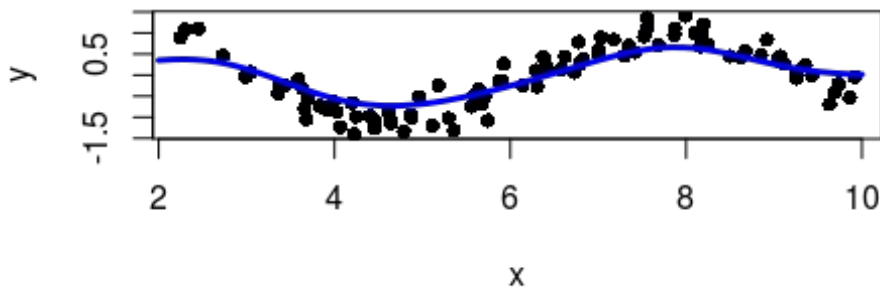Following is the best predicted line with epsilon = 0.1

C=1, epsilon=0.1

when epsilon is small number of support vectors used to find the prediction line is large . So the model over-fits.

C=1, epsilon=0.01

when the epsilon is large number of support vectors used to find the prediction line is small. So the model under-fits.
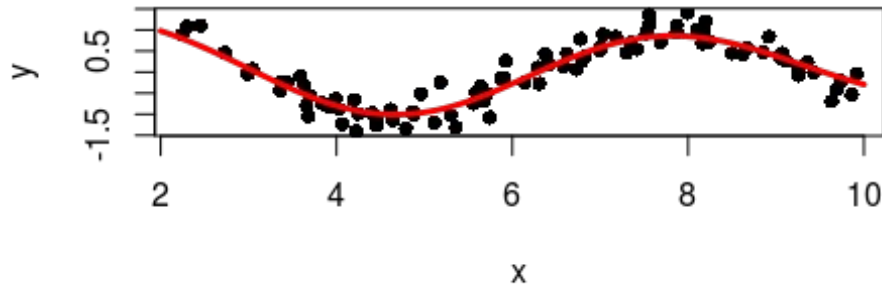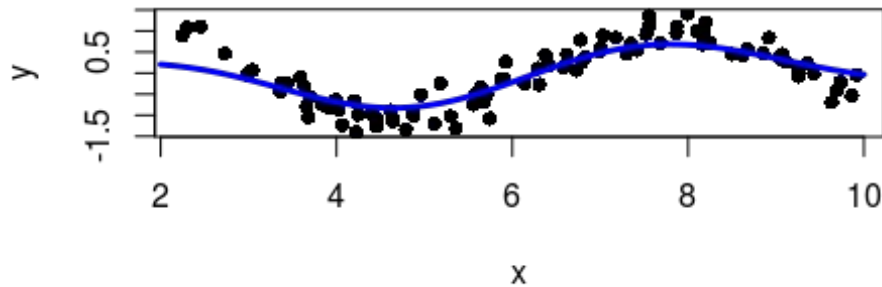
C=1, epsilon=1

b) Trying different values of sigma to check how the model behaves.

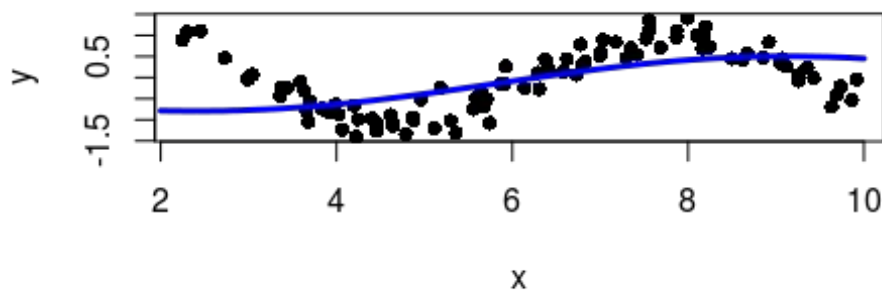When sigma become low, the model underfits the data as the figures below.
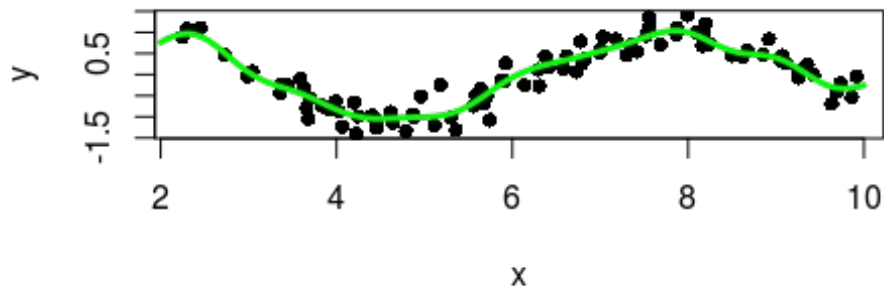
C=1, epsilon=0.1,sigma=0.5

C=1, epsilon=0.1,sigma=1
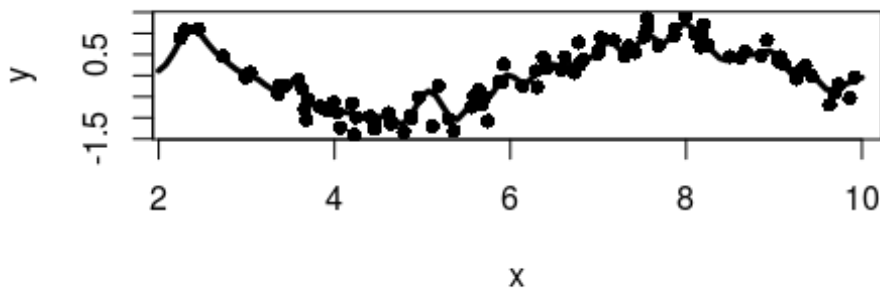

C=1, epsilon=0.1,sigma=0.1

when sigma values are too high, the model overfits the data as the figures below shows.

C=1, epsilon=0.1,sigma=10

C=1, epsilon=0.1,sigma=100

Explanation:

Epsilon > epsilon determines how many data points will be used in the support vector to contribute to the regression line. the lower the epsilon many data-points and higher the epsilon low data points to impact the prediction equation. epsilon high means number of support vectors are low to compute the model. so under-fits when epsilon too low too many data points so it over fits.

Sigma > sima parameter controls the scale. string "automatic" calculate a good sigma value for the Gaussian RBF and high sigma value over-fits the data and ow sigma under-fits the data.

Cost > When the cost is large, the model becomes very flexible and the cost is small the model will "stiffen" and become less likely to over-fit but more likely to under-fit.