3a. lets try to visualize it graphically,

→ actual line

→ $|y_i - w^T x| > 0$ for $\theta$ very small

→ $|y_i - w^T x| > 0$ for $\theta$ very large
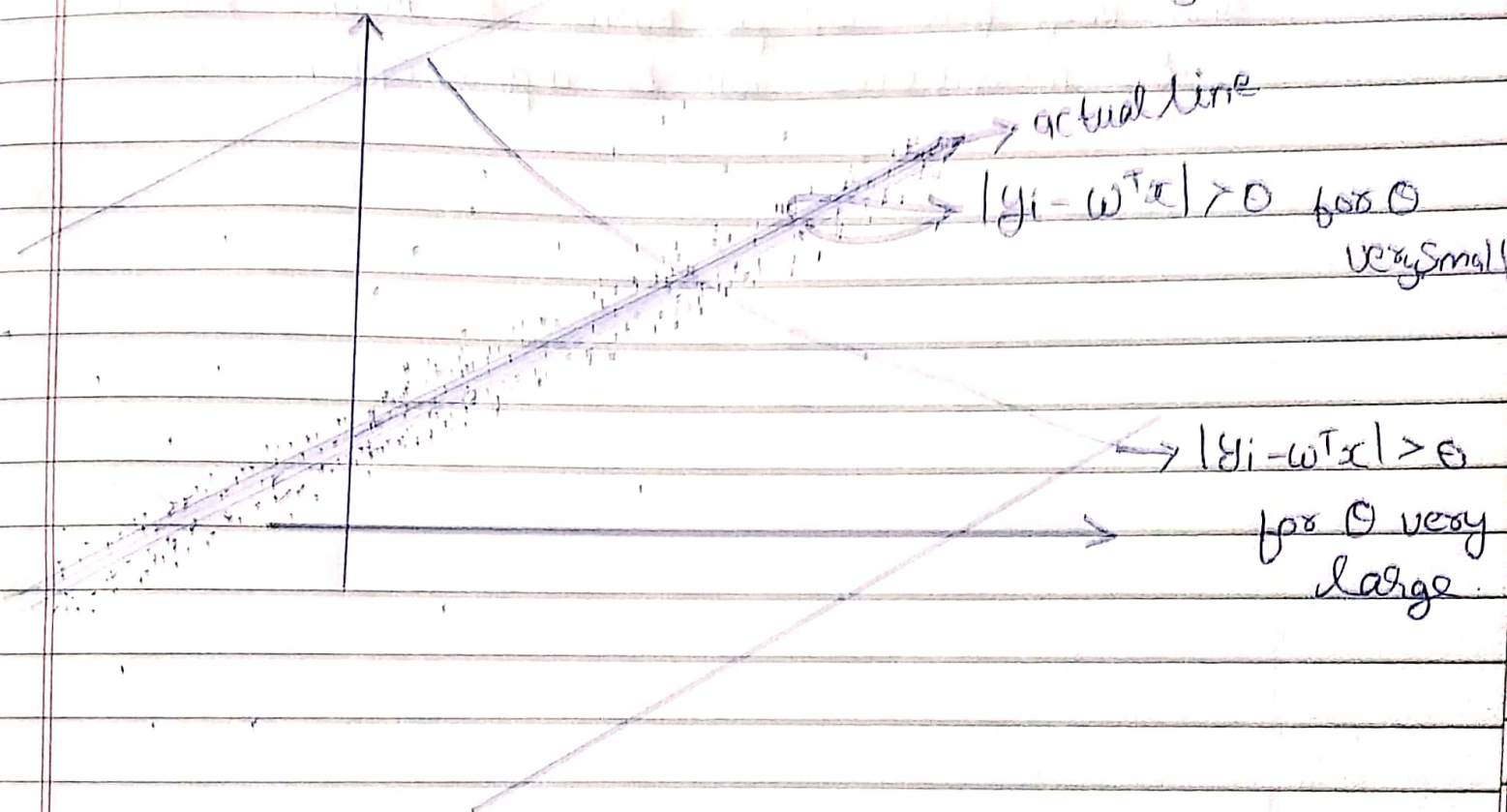
from the visualization we can say that if
we do $|y_i - w^T x| > 0$ for very small theta,
most of the most are classified as outliers
and hence there are very few points only left
to make the model, and due to that the noise
in those heavely affect the model.

for $\theta$ very large all the points are considered
for making the model, so even the outliers are also
accounted there fore the model performs very bad very
large theta.

b) a simple way to determine effective $\theta$ would be
to calculate the mean error and take that
as $\theta$, because outliers affect ~~outliers affects~~
the means the most, and this will make sure that
most of the points will come into the mean region
and very few will go ouside the mean. heance
$\theta = mean(err)$ will be a good option.