**The algorithms that suffer from features of significantly different scales in training set:**

1. K-nearest neighbors:

K-nearest neighbors relies on computing distances between data points. When features have different scales, those with larger scales might dominate the distance computations, leading to biased results. Features with larger scales can have a larger impact on the distance calculation, making the algorithm sensitive to the choice of scaling.

1. Support Vector Machines (SVM):

SVM aims to find the hyperplane that best separates classes. As SVM seeks to maximize the margin orthogonal to this hyperplane, features with larger scales might disproportionately influence the decision boundary. This can result in the model being biased towards features with larger scale

**How to fix features of different scales**

Normalize the features to bring them to a similar scale.

This can be done by:

* Min-max scaling: scale the features to a range.
* Standardization: transform the data to have a mean of 0 and a standard deviation of 1.

**Why gradient descent is preferred rather than normal equations in estimating parameters in linear regression:**

* **Scalability:**gradient descent tends to perform better for large datasets where matrix operations become computationally expensive.
* **Avoiding Matrix inversion**: as dataset size grows, matrix inversion in normal equations becomes increasingly expensive, whereas gradient descent does not involve this step.