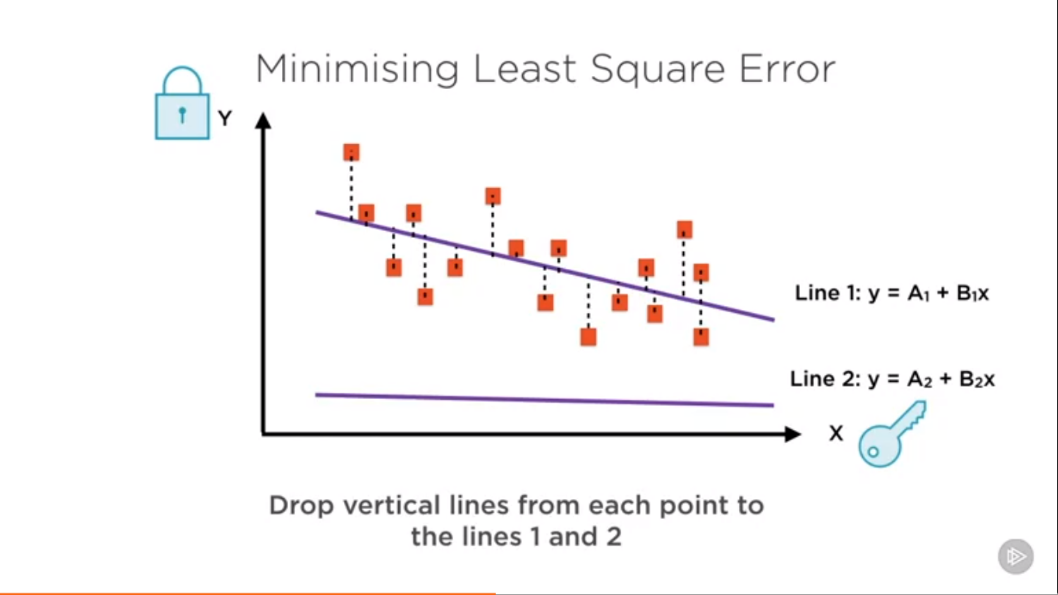
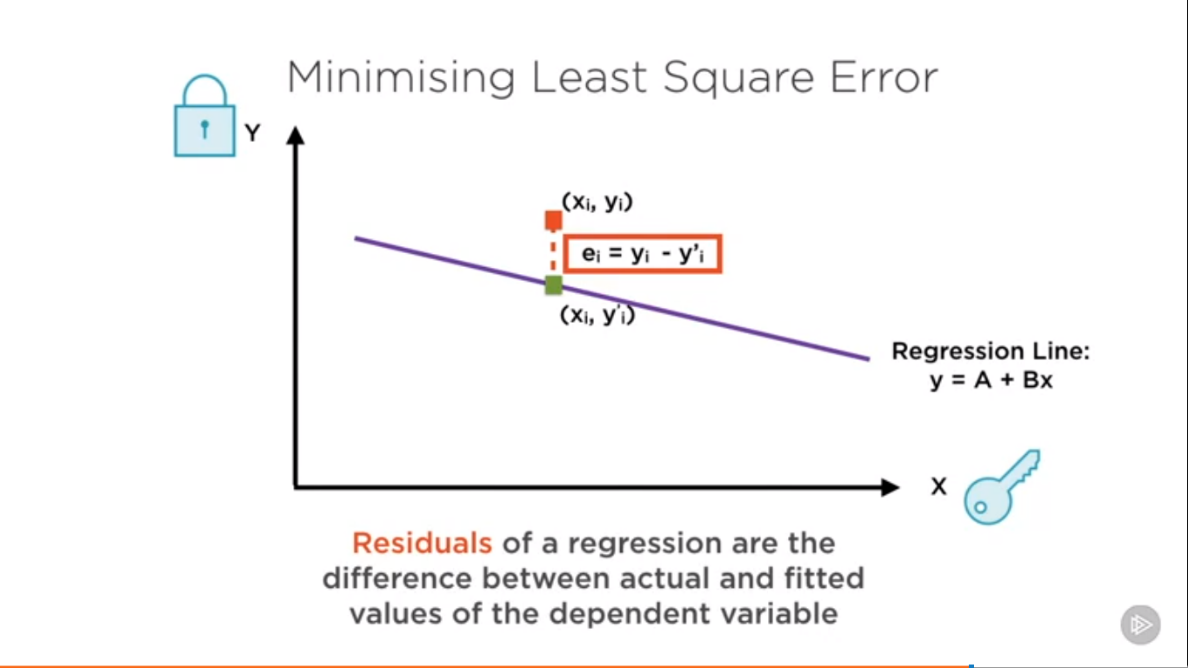
**Regression:** X causes Y. X is Independent variable and Y is Dependent Variable.



The Best line can be found using Least Square Error

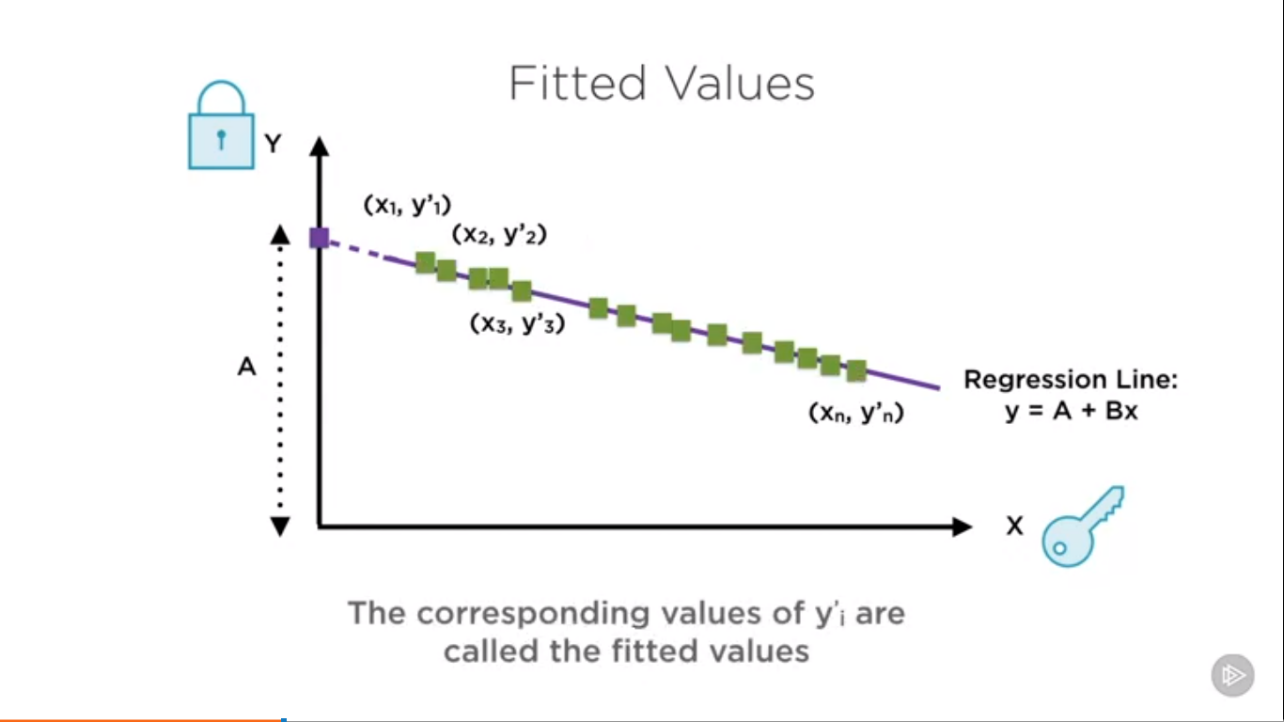


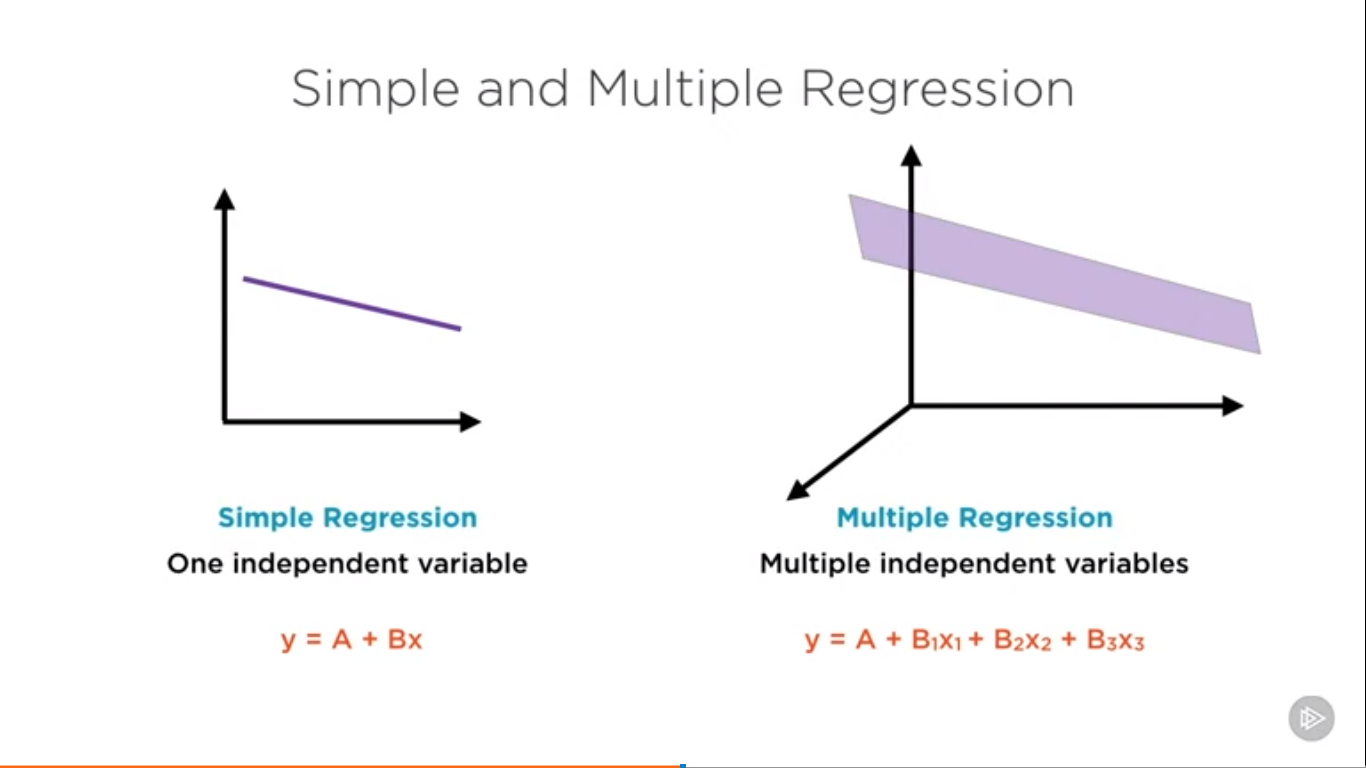
The best fit line is the one where the sum of the squares of the lengths of these dotted lines is minimum.



**Linear Regression as an Optimization Problem:**

**Regression Line:** The best fit line which minimizes the sum of squares of the errors





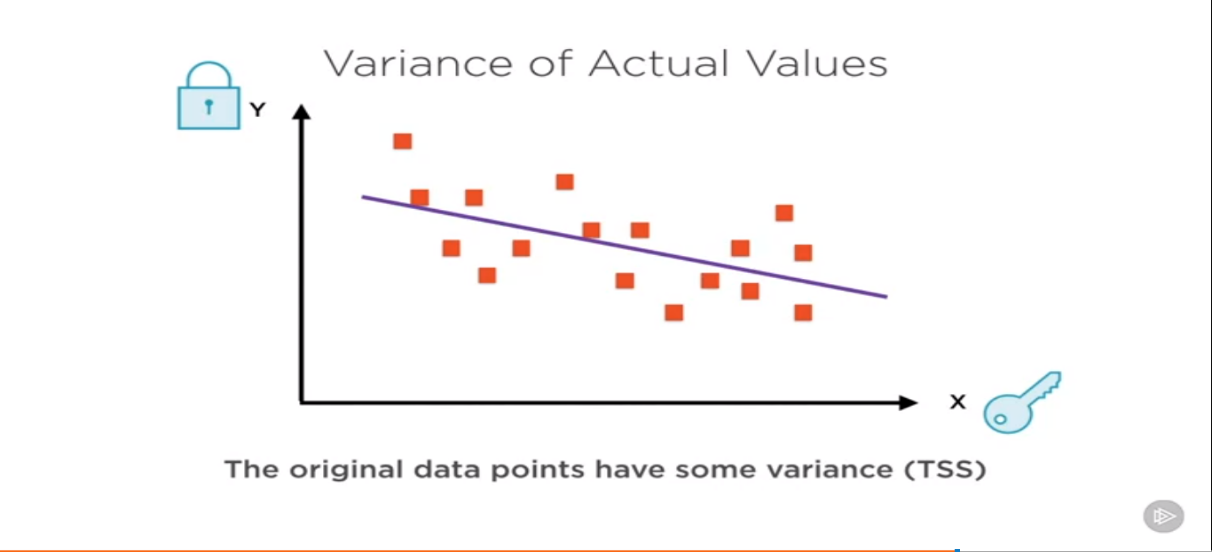
How good the Regression is: It is measured by R2

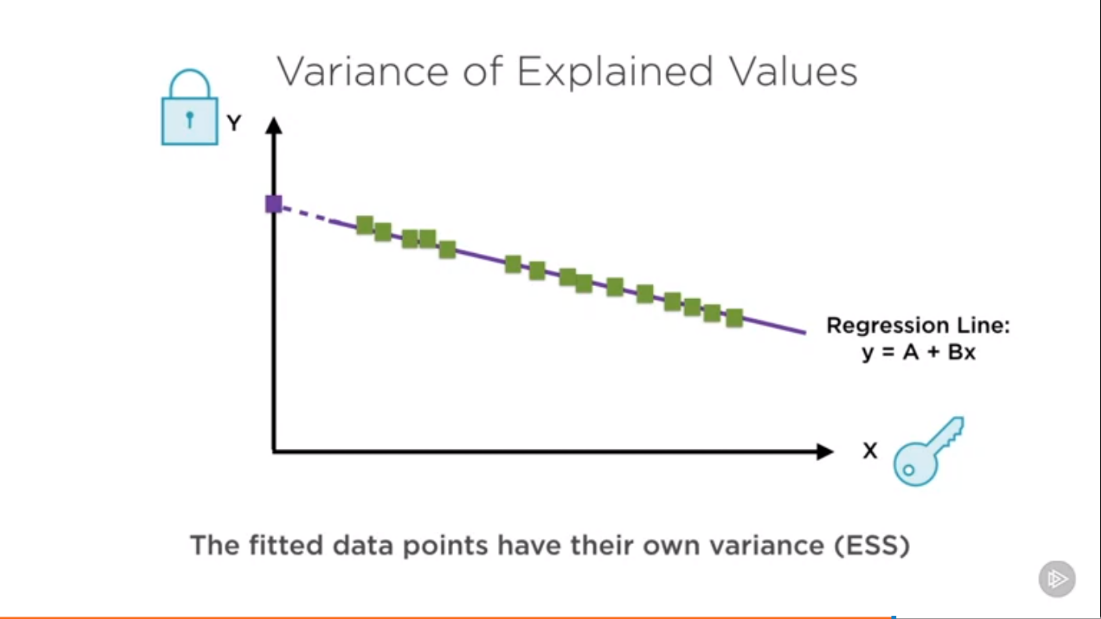
R2 = Explained Sum of Squares(ESS) / Total Sum of Squares(TSS)

ESS – Variance of fitted value

TSS – Variance of Actual value

**R2** – The percentage of total variance explained by the regression. Usually higher the R2, the better the quality of Regression (upper bound is 100%)



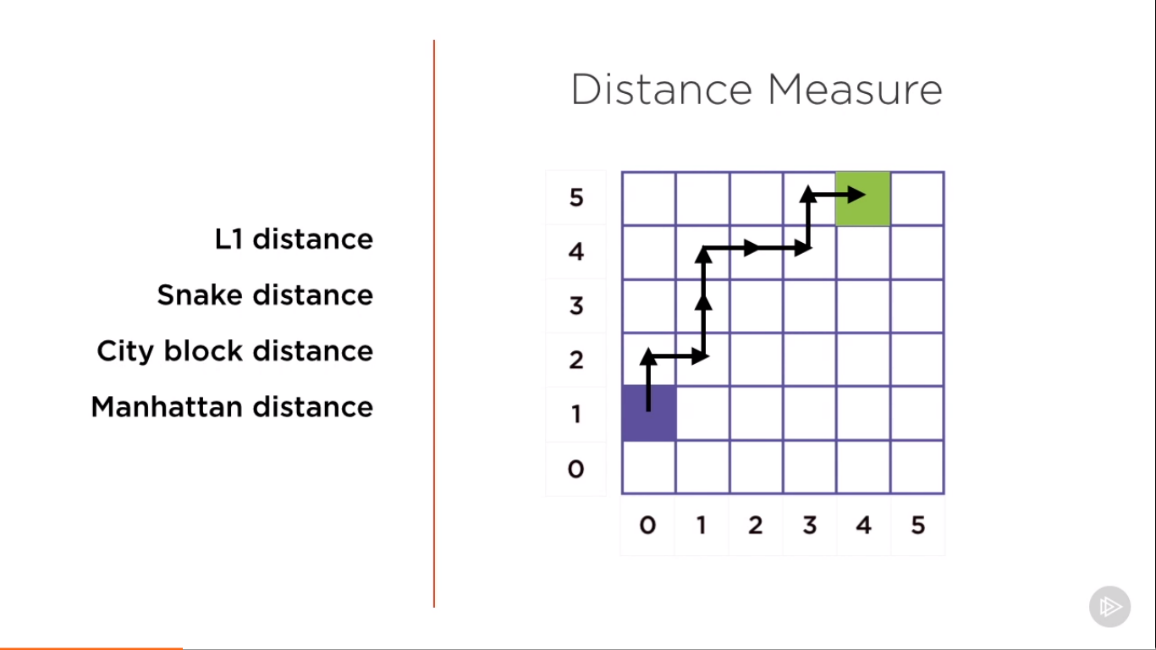


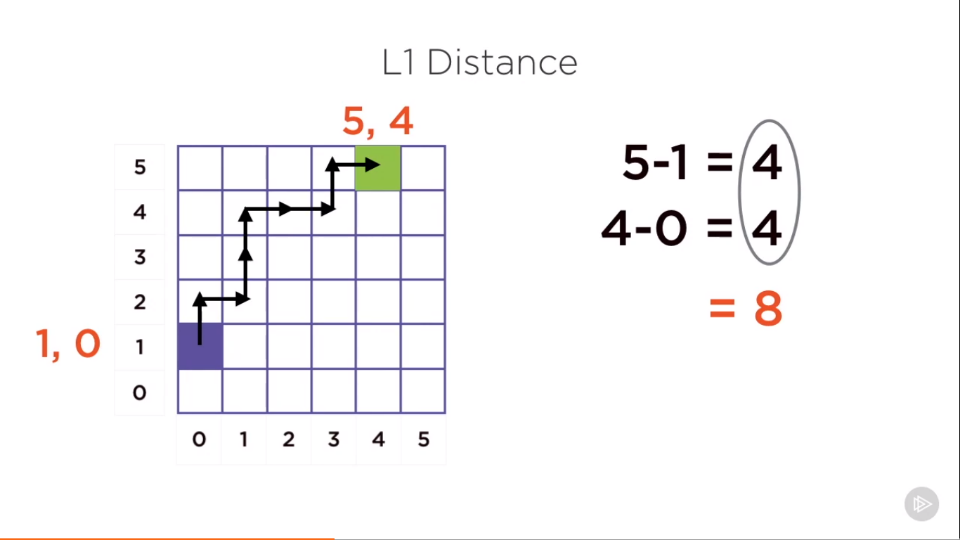
R2  = How much of the original variance is captured in the fitted values?

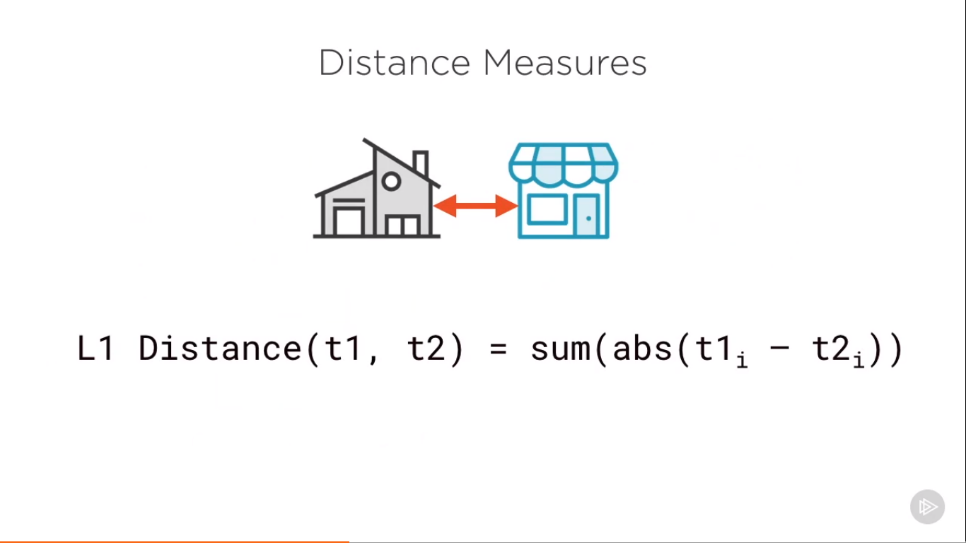
**Demo: Data Preparation:**

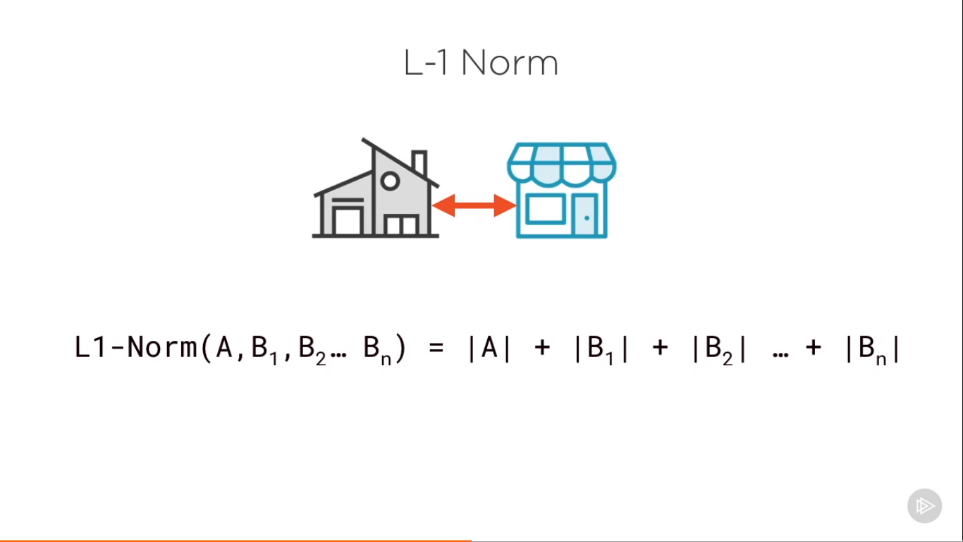
**L1 and L2 Norms:**

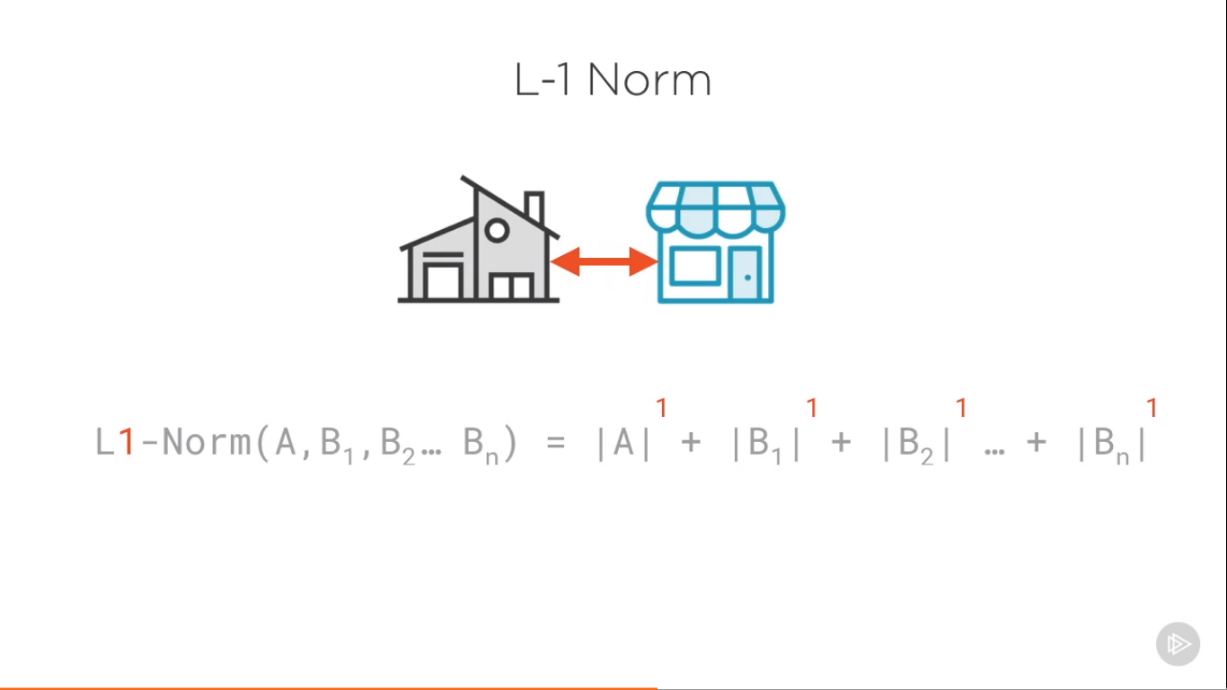
**L1 Distance**

****

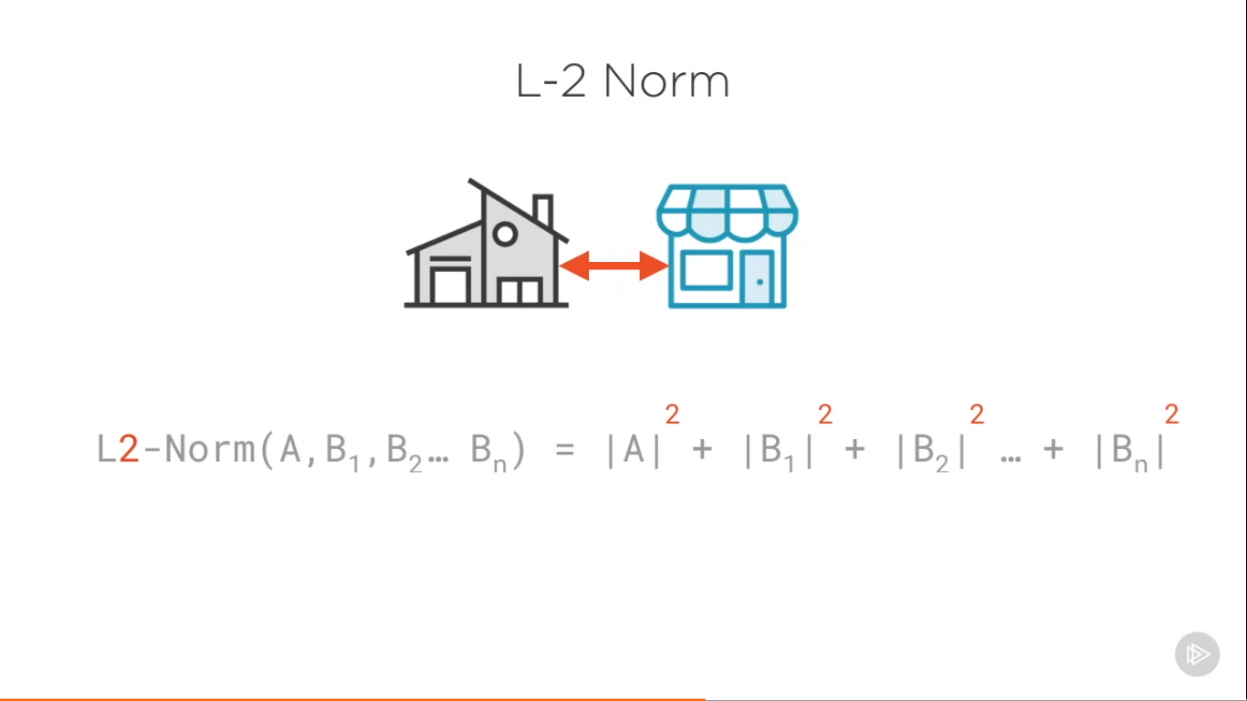
****

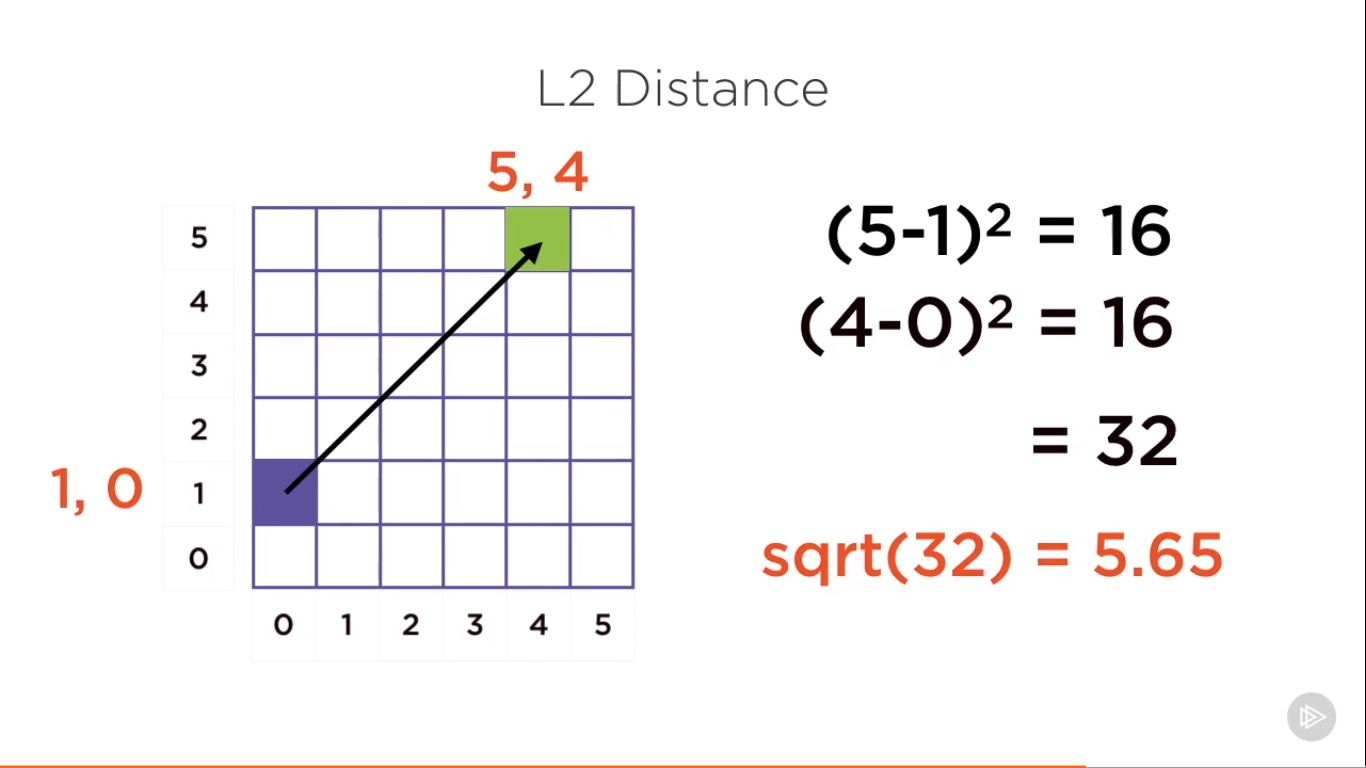
****

****

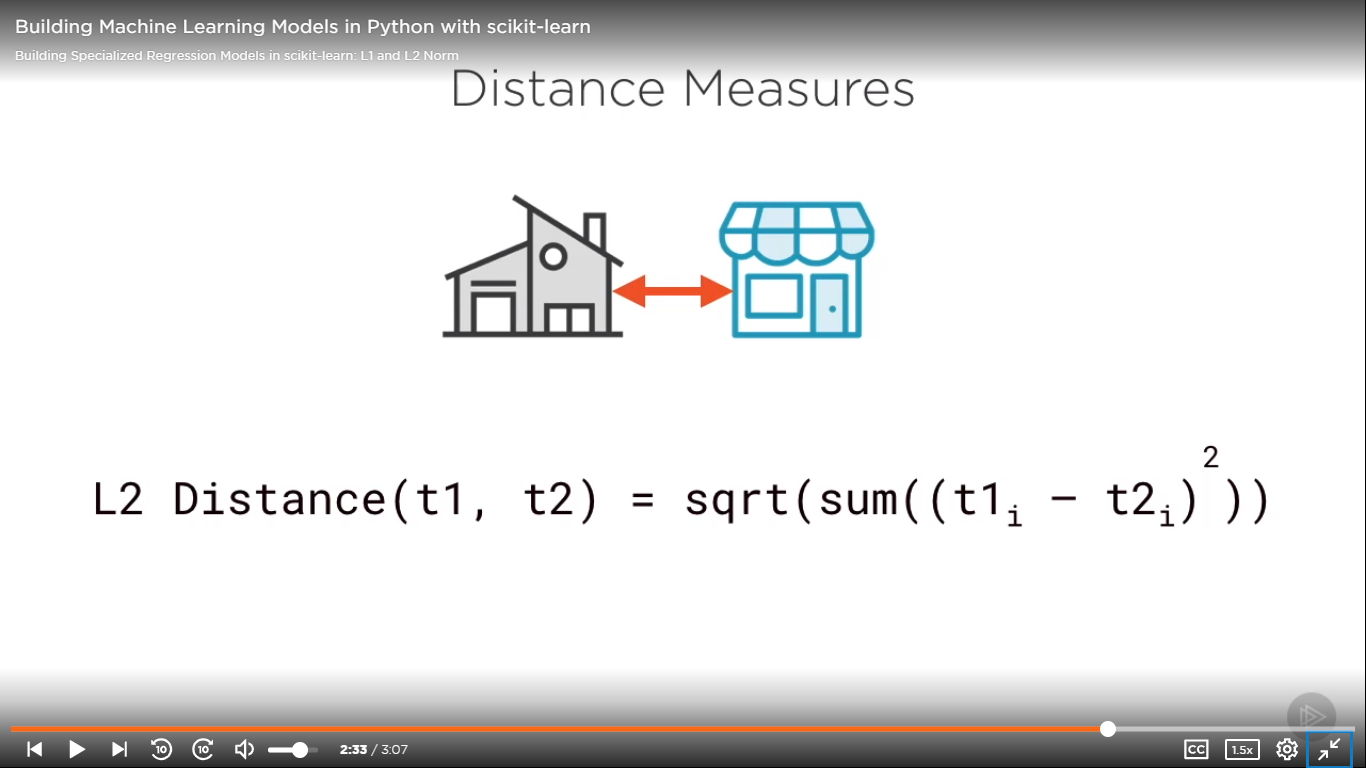
****

**L2 Distance**

****

****

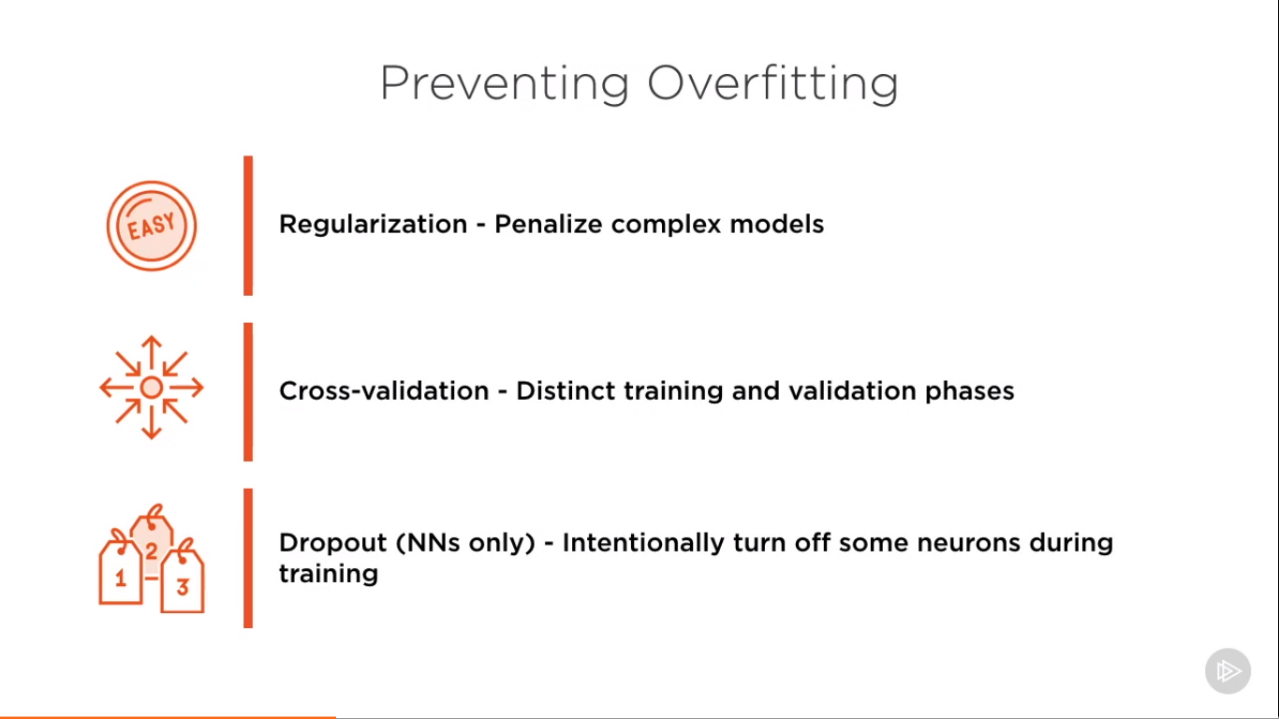
**L2 Distance = sqrt(L2 Norm)**

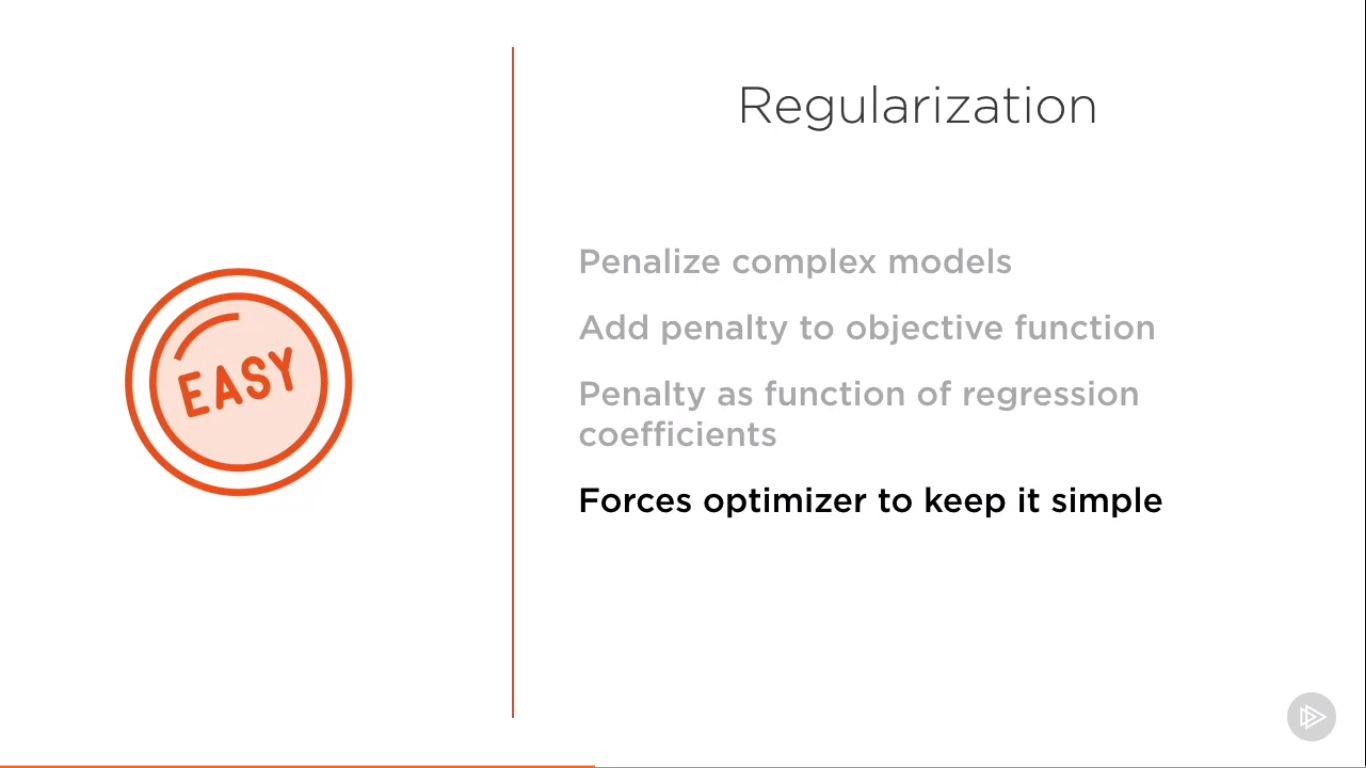
****

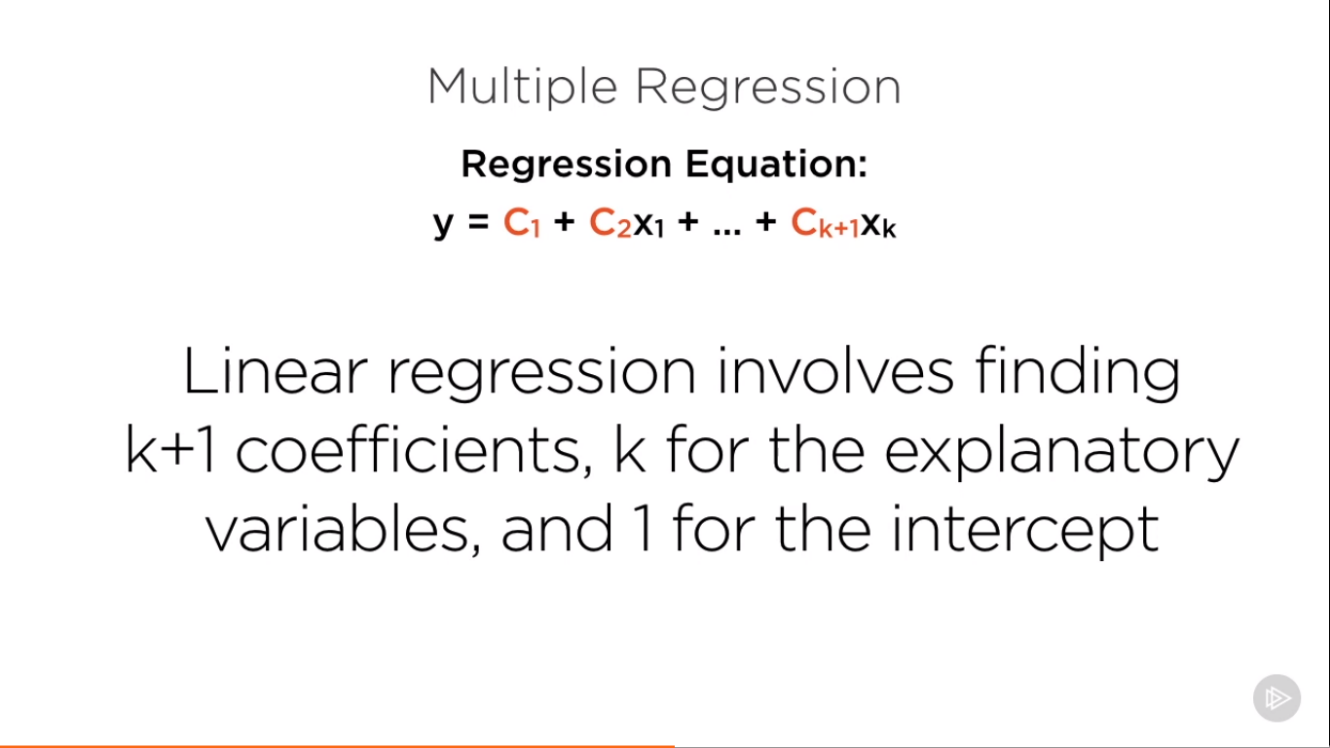
**Variance and Bias**

**Note**

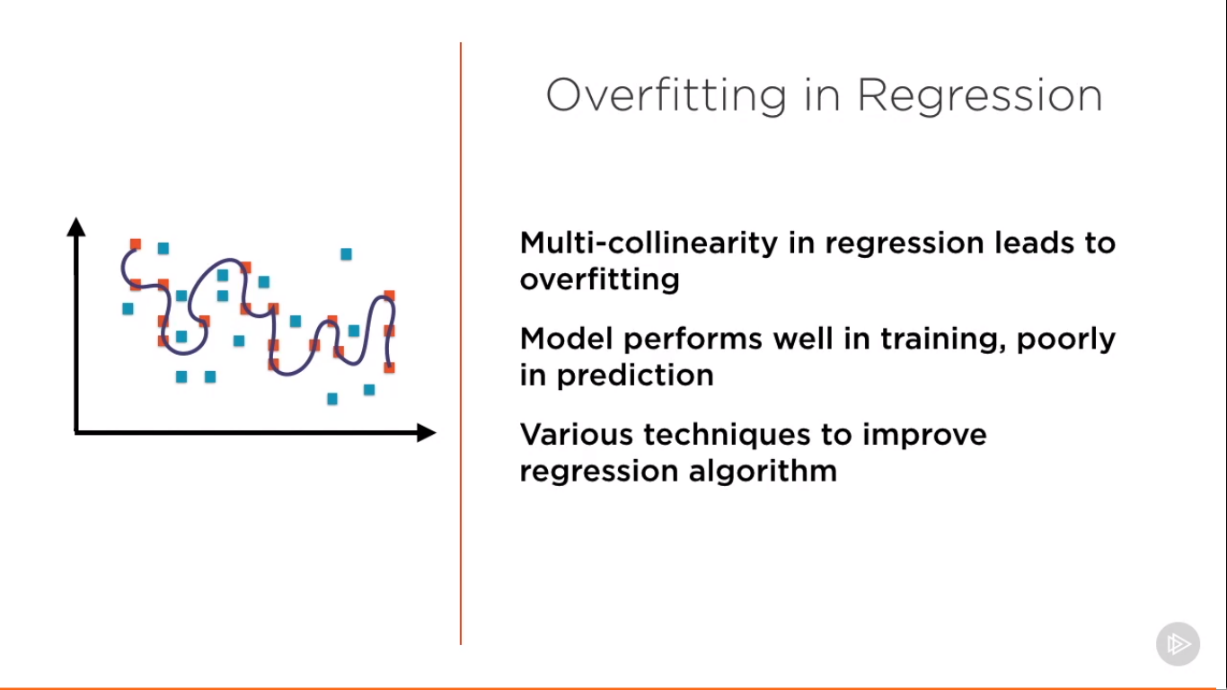
Low bias means we made very less assumption, high bias means we made more assumptions like in linear regression that one line will represent all data properly.

****

****

****

A big risk with regression is multicollinearity: X variables containing the same information.

****

**Lasso and Ridge Regression:**

**Preventing Overfitting** requires use of Regularized Regression Model

**Lasso Regression:** Penalizes large regression coefficients

**Ridge Regression:** Also penalizes large regression coefficients

**Elastic Net Regression:** Simply combines lasso and ridge

**Regularization:**

* Reduces the variance error
* But increases bias error

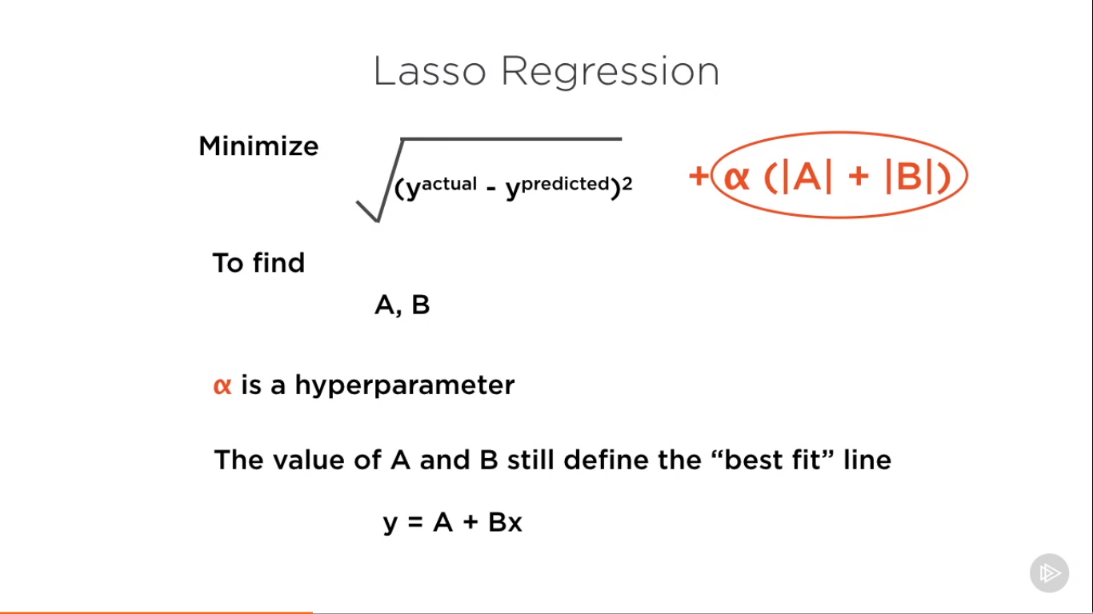
Objective function of our Least square Regression:

Minimize

For minimizing the mean square error find A and B which give the best fit line

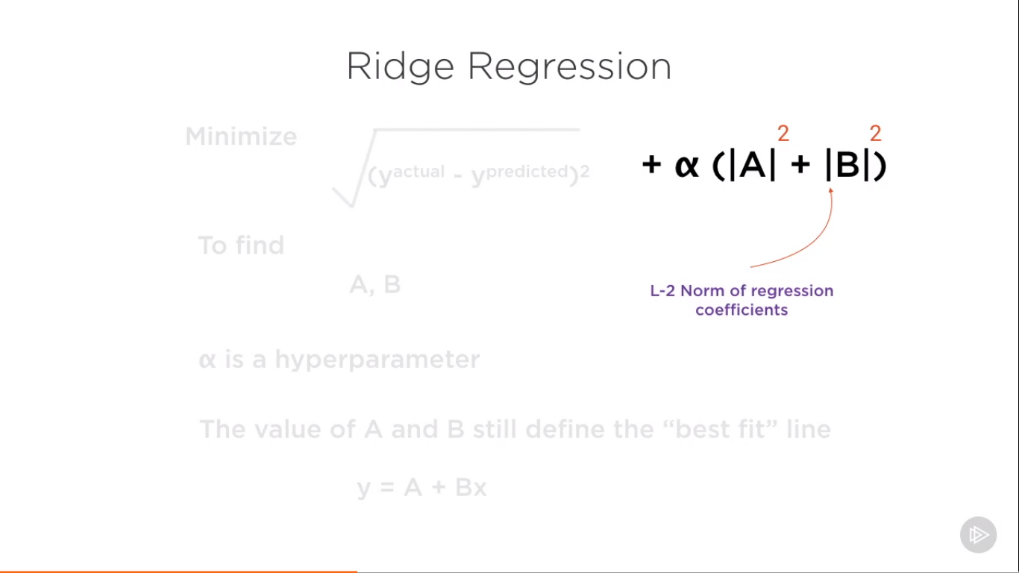
y = A + Bx

In **Lasso Regression:** It is not the mean square error that is minimized an additional term is added to the objective function which we seek to minimize



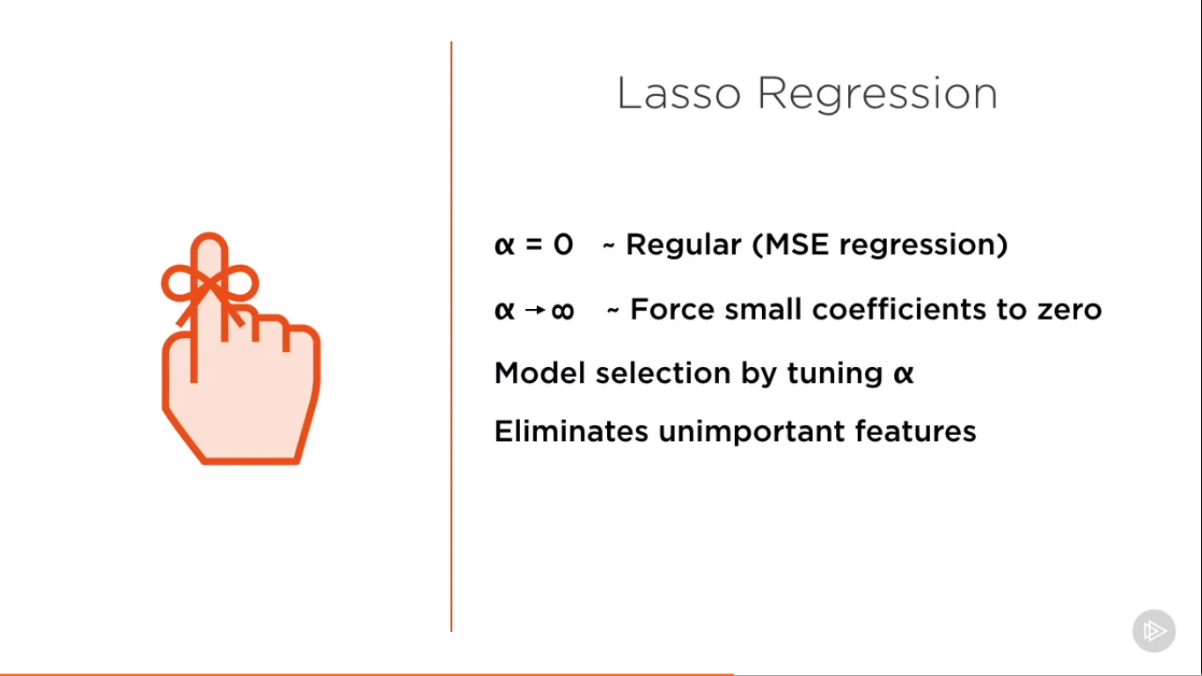
L1 norm of the coefficients (α (|A| + |B|)

**Ridge Regression** is exactly like the Lasso Regression except that the penalty function is the L2 norm of the coefficients



**Characteristics of Lasso Regression:**

* Add penalty for large coefficients
* Penalty term is L-1 norm of coefficients
* Penalty weighted by hyperparameter α which you can tweak to find the best possible model for your data



**Lasso** – Least Absolute Shrinkage and Selection Operator

Ridge Regression have exactly the same Characteristics except it uses L-2 norm for Penalty term

**Differences:**

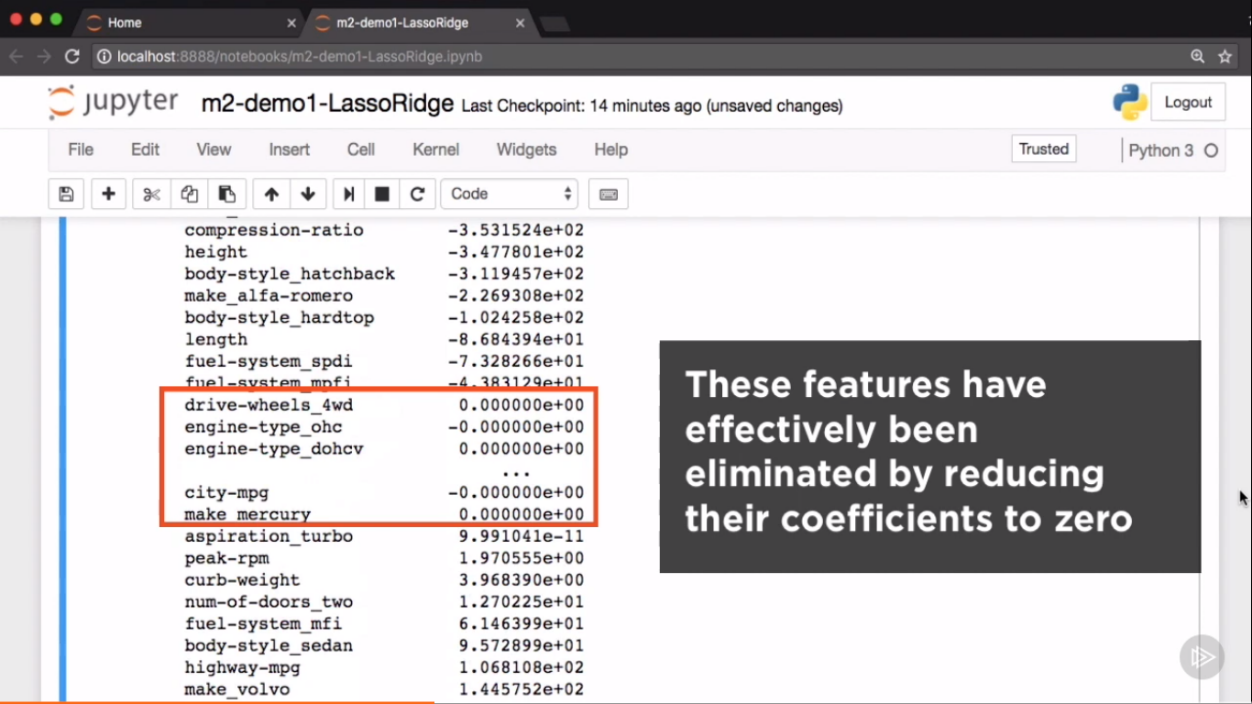
**Lasso Regression** cannot be represented as formula i.e. it does not have a closed form of solution

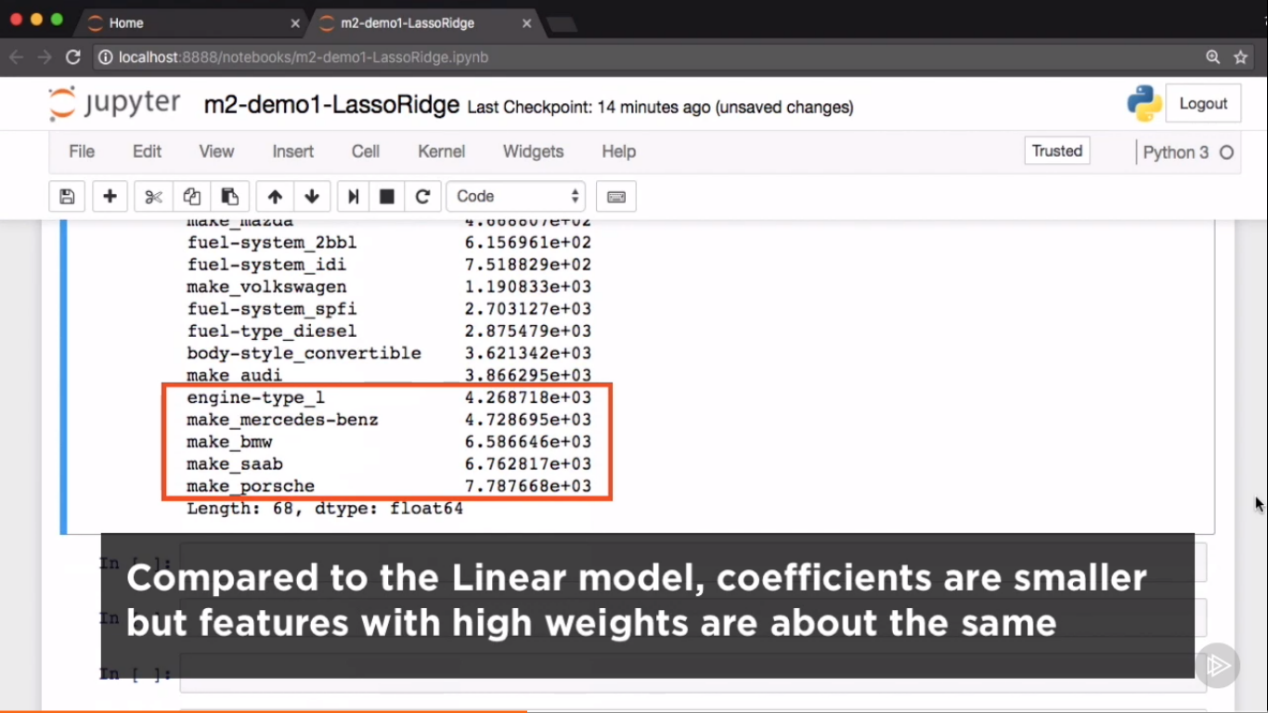
**Ridge Regression** has Closed form of solution

Unlike lasso, ridge regression will not force coefficients to 0 if you increase value of alpha to be very high.

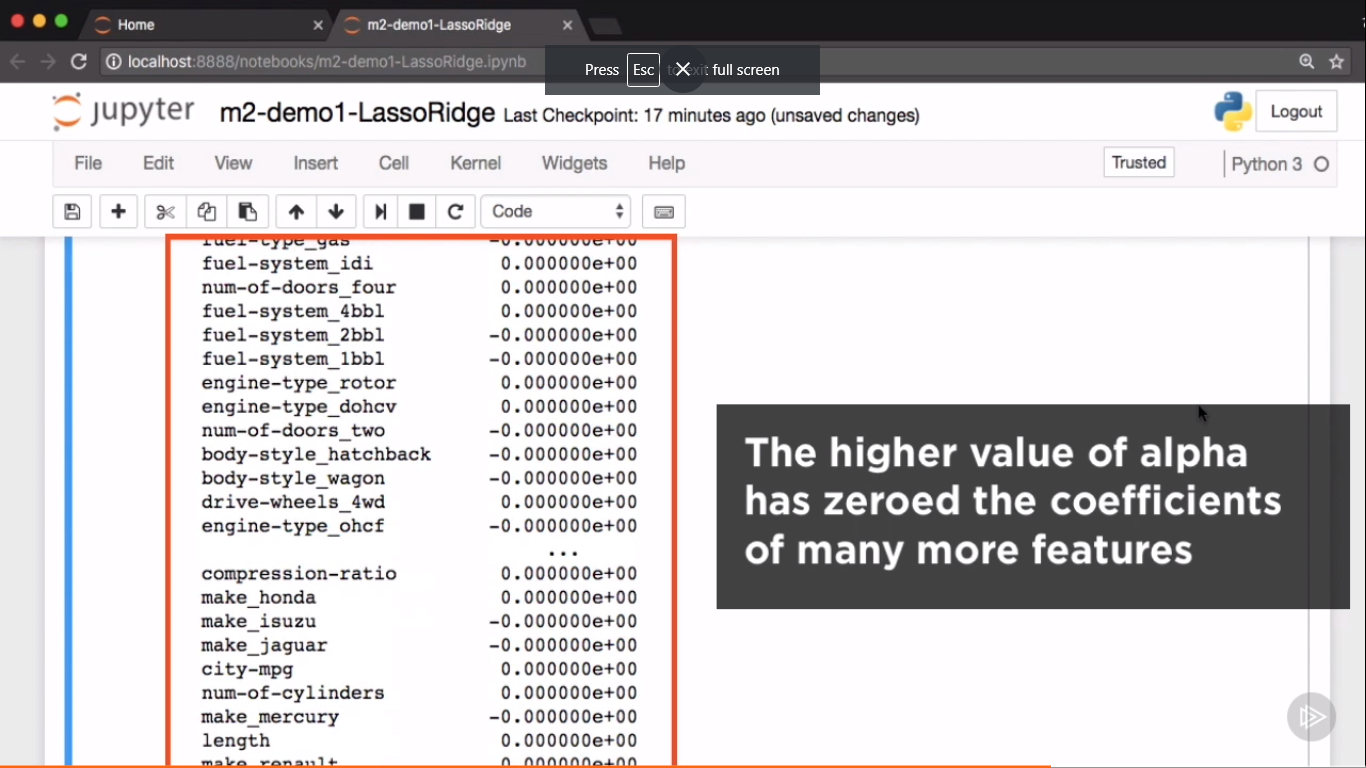
As ridge regression is closed solution it does not perform model selection.

**Demo Lasso Regression:**

****

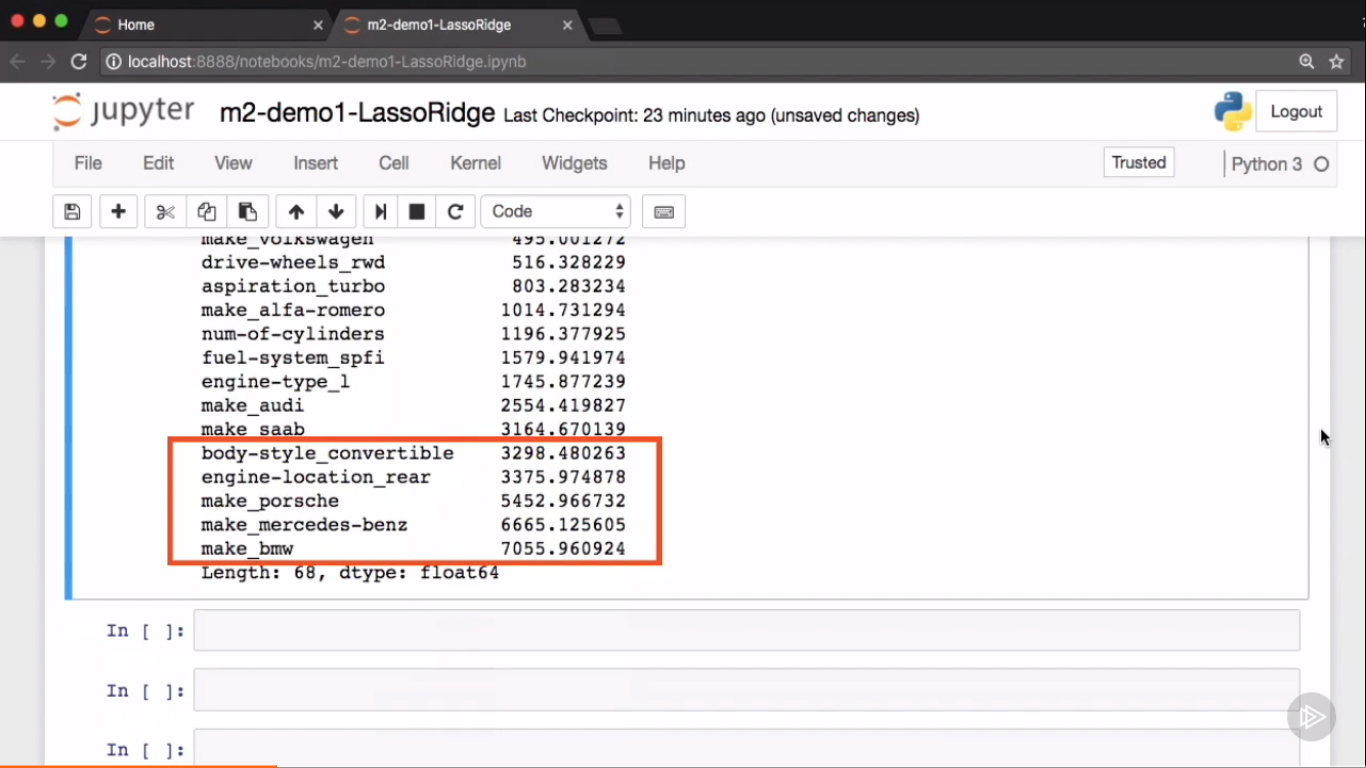
****

**If we increase the value of alpha**





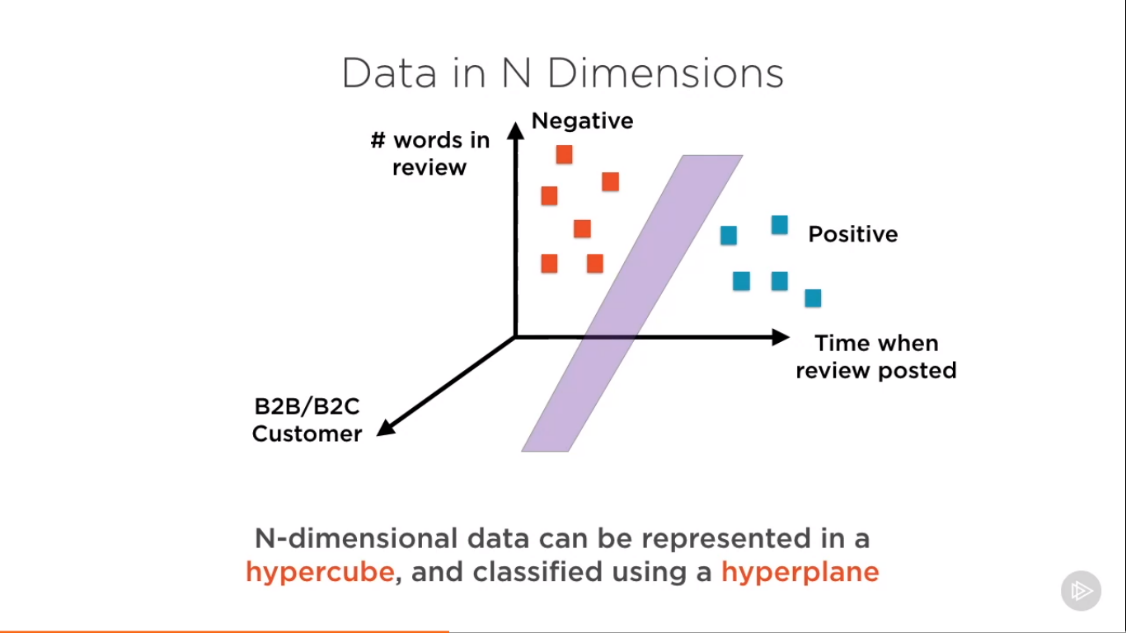
Ridge:

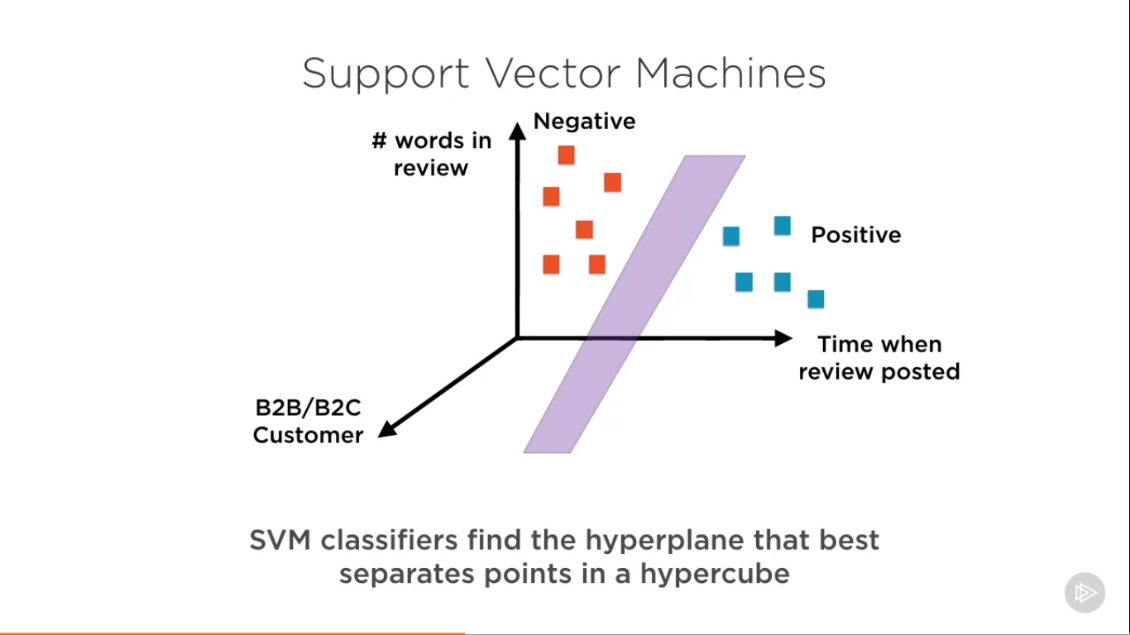


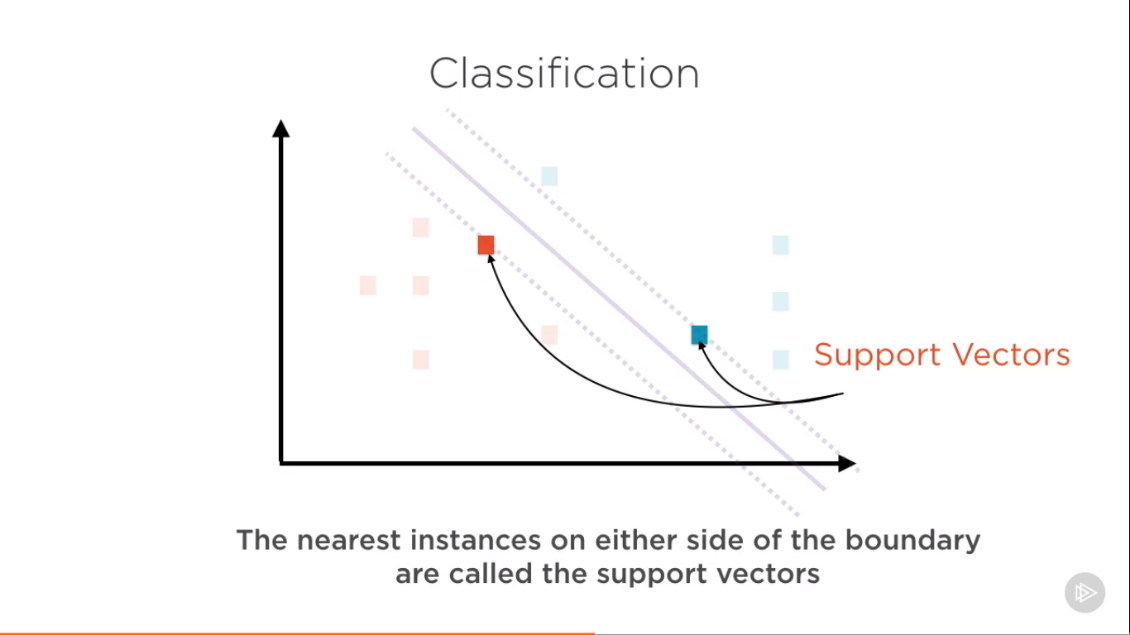
You can see clearly that high coefficient values are much lower than Lasso as well as simple linear regression

SVM’S are typically used for classification problems.

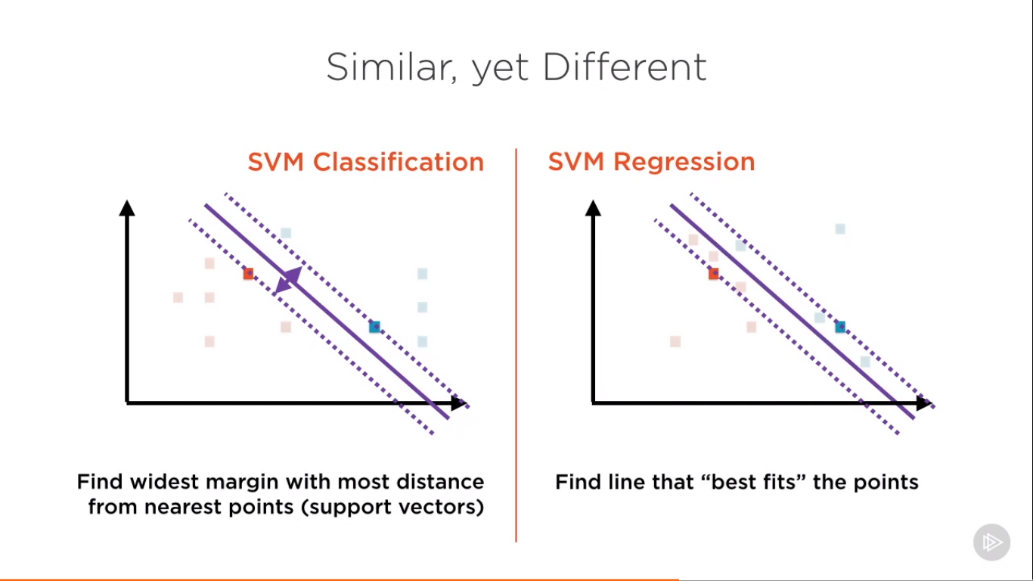
SVR’S use the same underlying principles with a **different objective function**

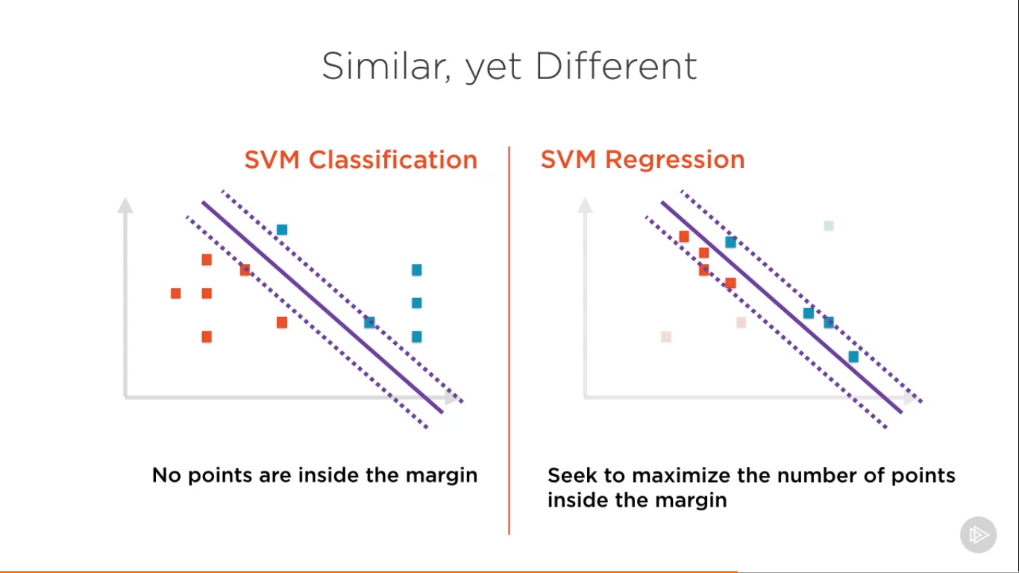






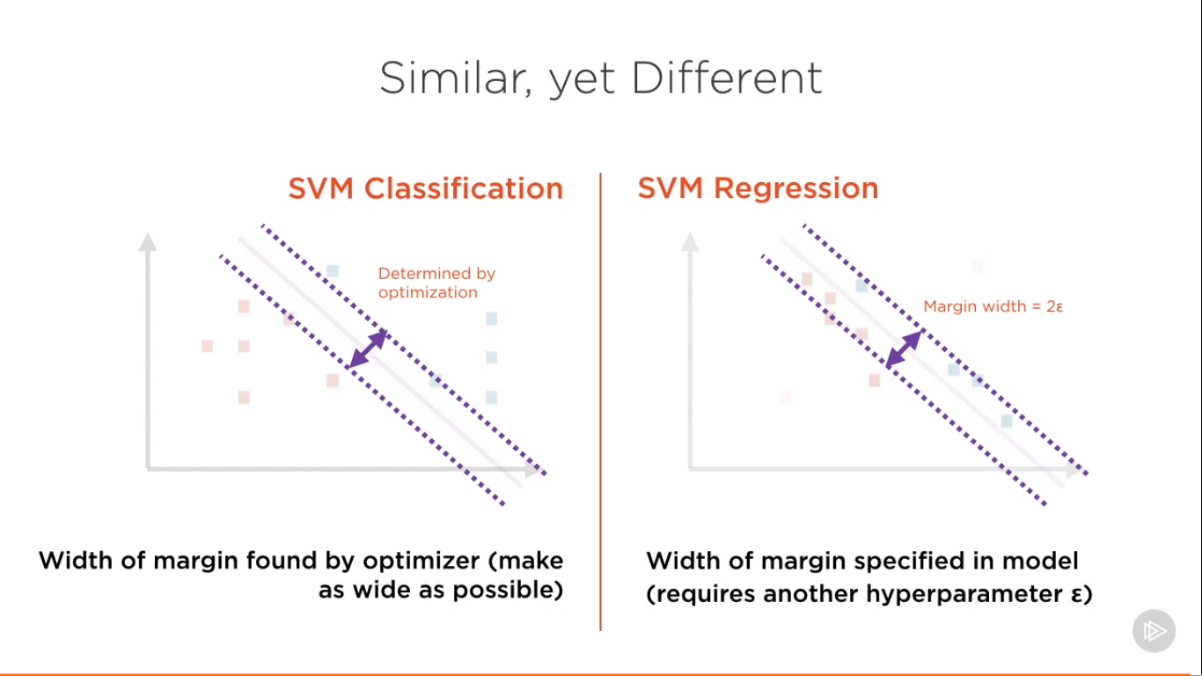








In SVM regression penalty is not given on wrong side rather it is given if how much point is away from margin.



**Implementing Support Vector Regression in scikit-learn:**