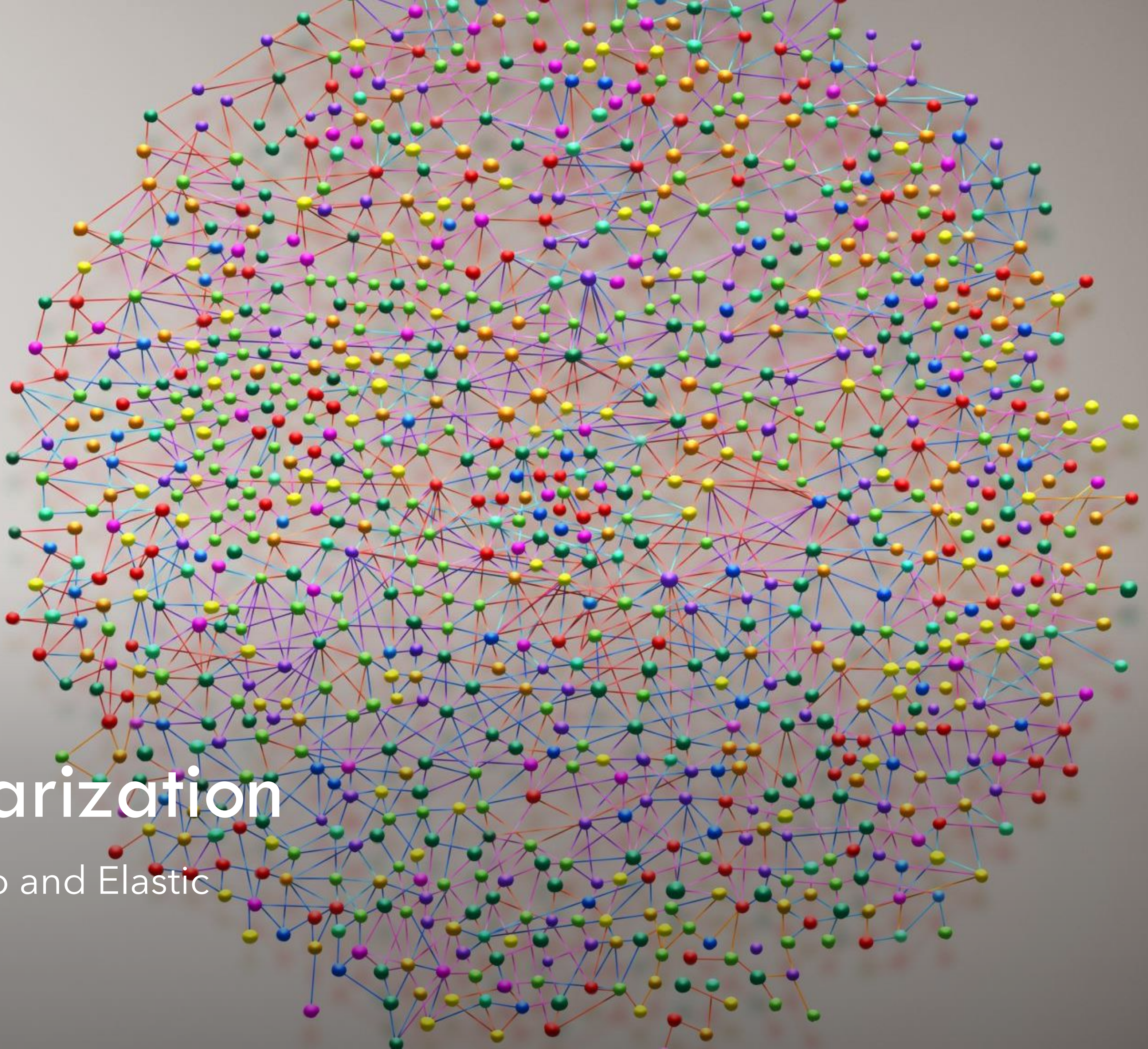
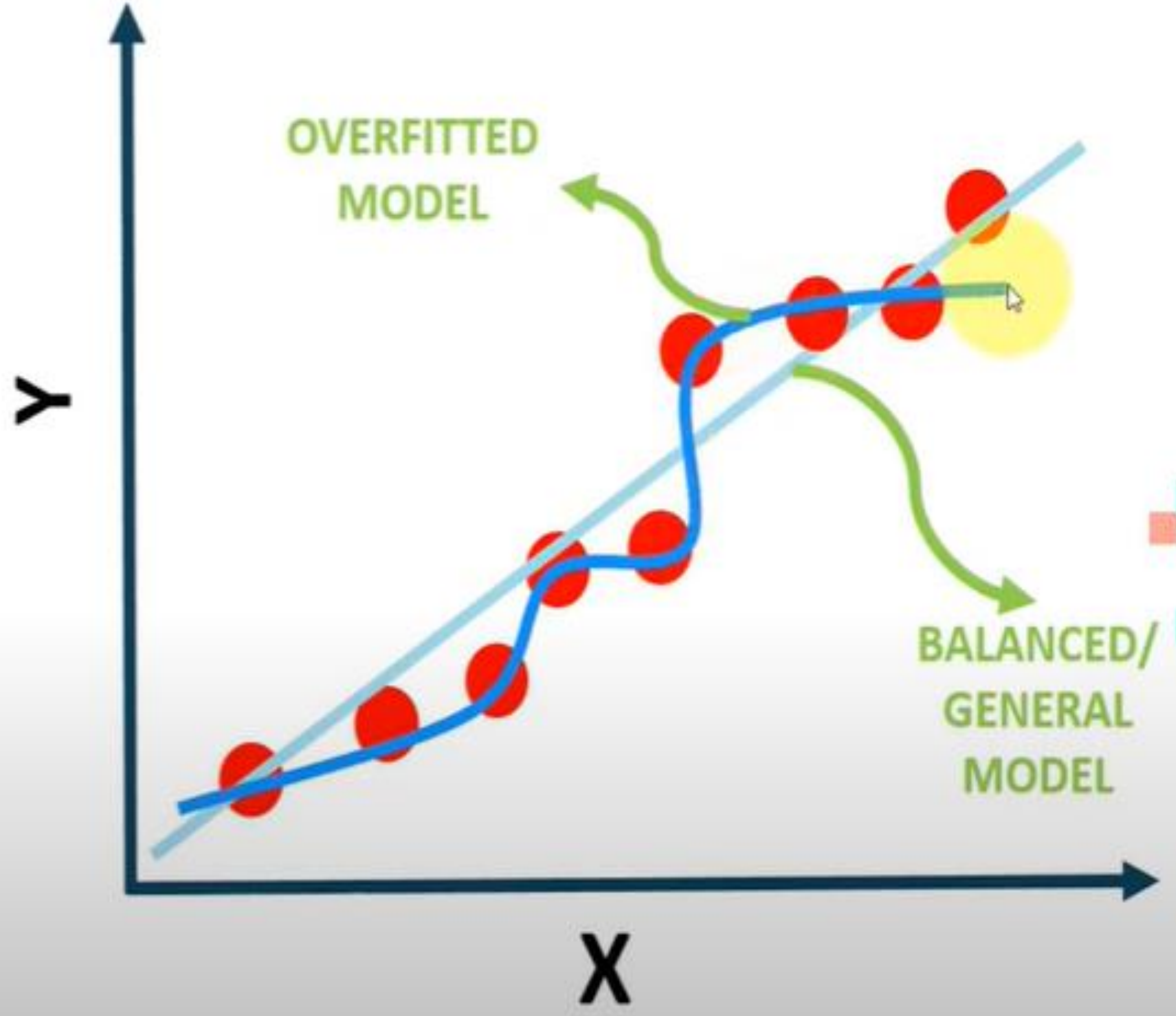


Regularization

Ridge,Lasso and Elastic

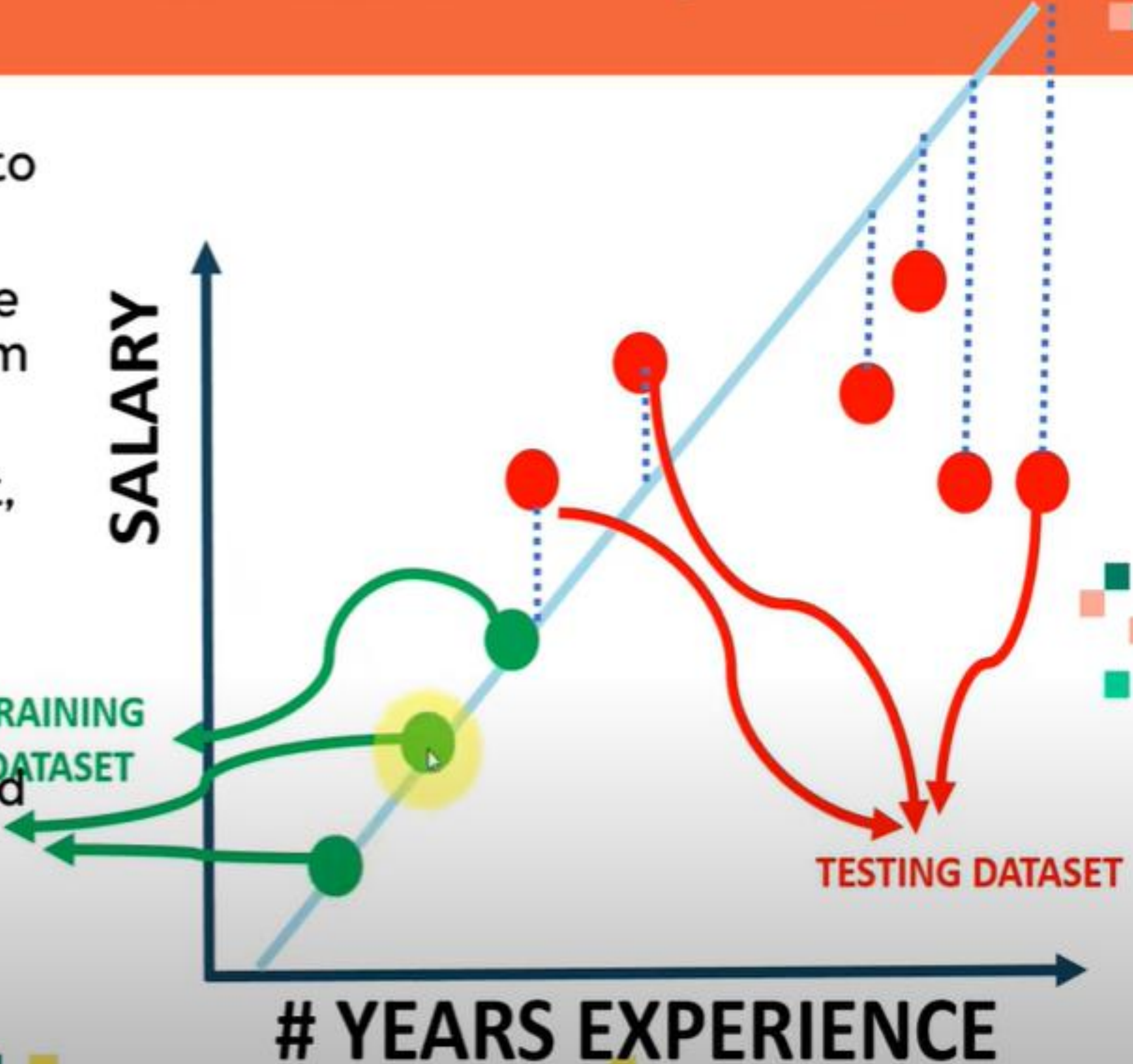
RIDGE REGRESSION (L2 REGULARIZATION): INTUITION

- Ridge regression advantage is to avoid overfitting.
- Our ultimate model is the one that could generalize patterns; i.e.: works best on the training and testing dataset
- Overfitting occurs when the trained model performs well on the training data and performs poorly on the testing datasets
- Ridge regression works by applying a penalizing term (reducing the weights and biases) to overcome overfitting.



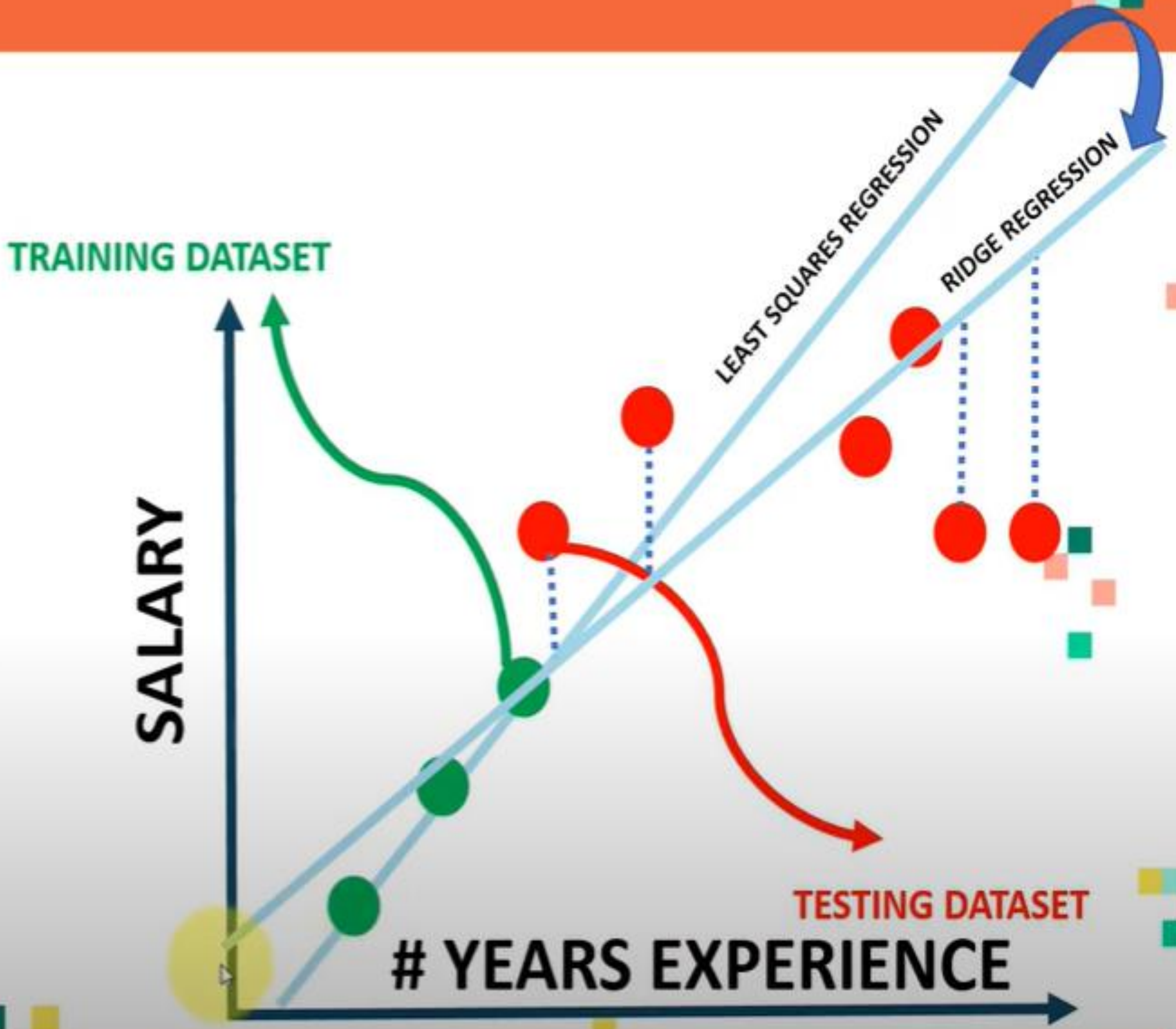
RIDGE REGRESSION (L2 REGULARIZATION): INTUITION

- Least sum of squares is applied to obtain the best fit line
- Since the line passes through the 3 training dataset points, the sum of squared residuals = 0
- However, for the testing dataset, the sum of residuals is large so the line has a high variance.
- Variance means that there is a difference in fit (or variability) between the training dataset and the testing dataset.
- This regression model is overfitting the training dataset



RIDGE REGRESSION (L2 REGULARIZATION): INTUITION

- Ridge regression works by attempting at increasing the bias to improve variance (generalization capability)
- This works by changing the slope of the line
- The model performance might be little poor on the training set but it will perform consistently well on both the training and testing datasets.



RIDGE REGRESSION (L2 REGULARIZATION): MATH

- Slope has been reduced with ridge regression penalty and therefore the model becomes less sensitive to changes in the independent variable (#Years of experience)

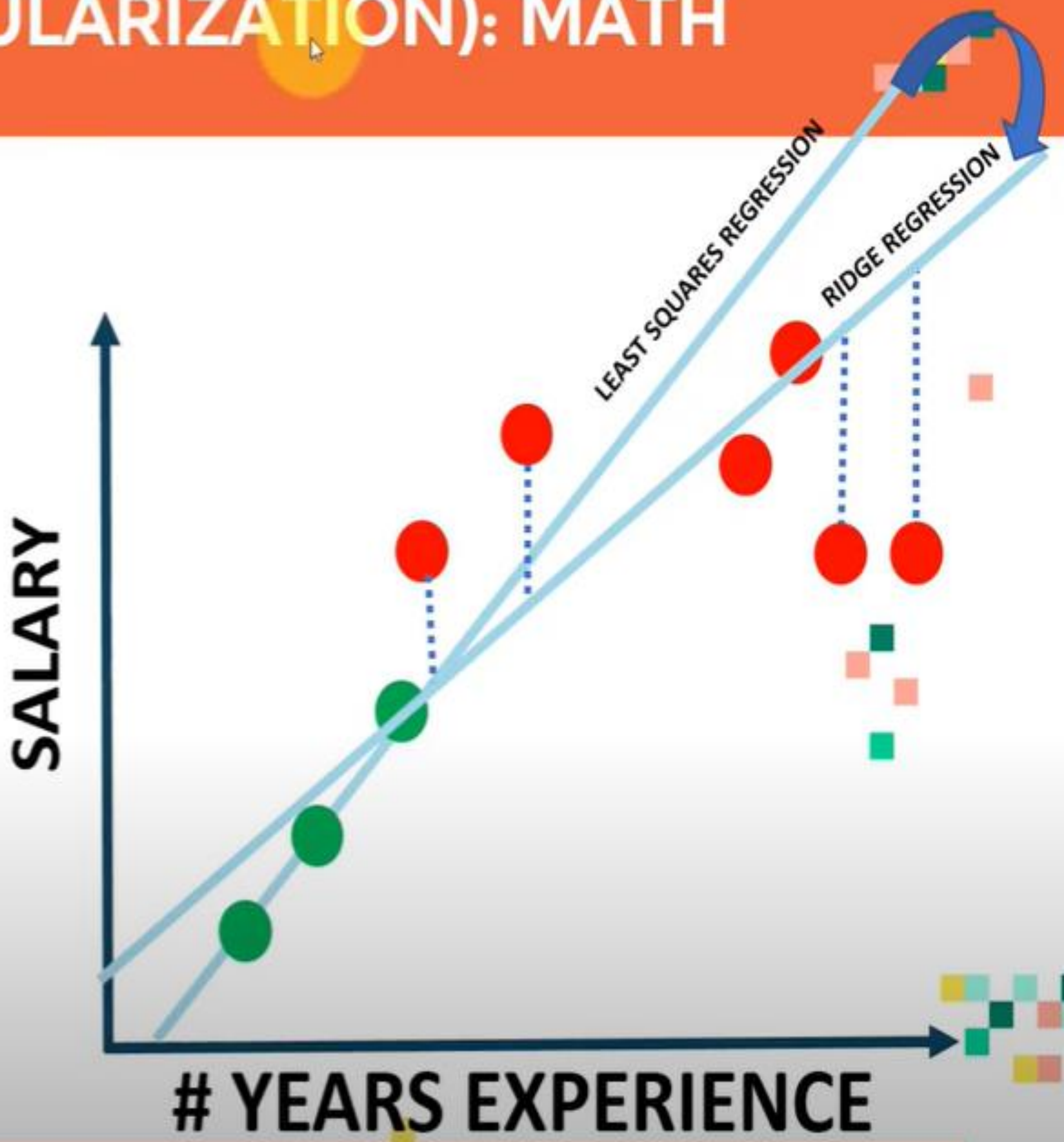
PENALTY TERM

Least Squares Regression:

$$Min(sum\ of\ the\ squared\ residuals)$$

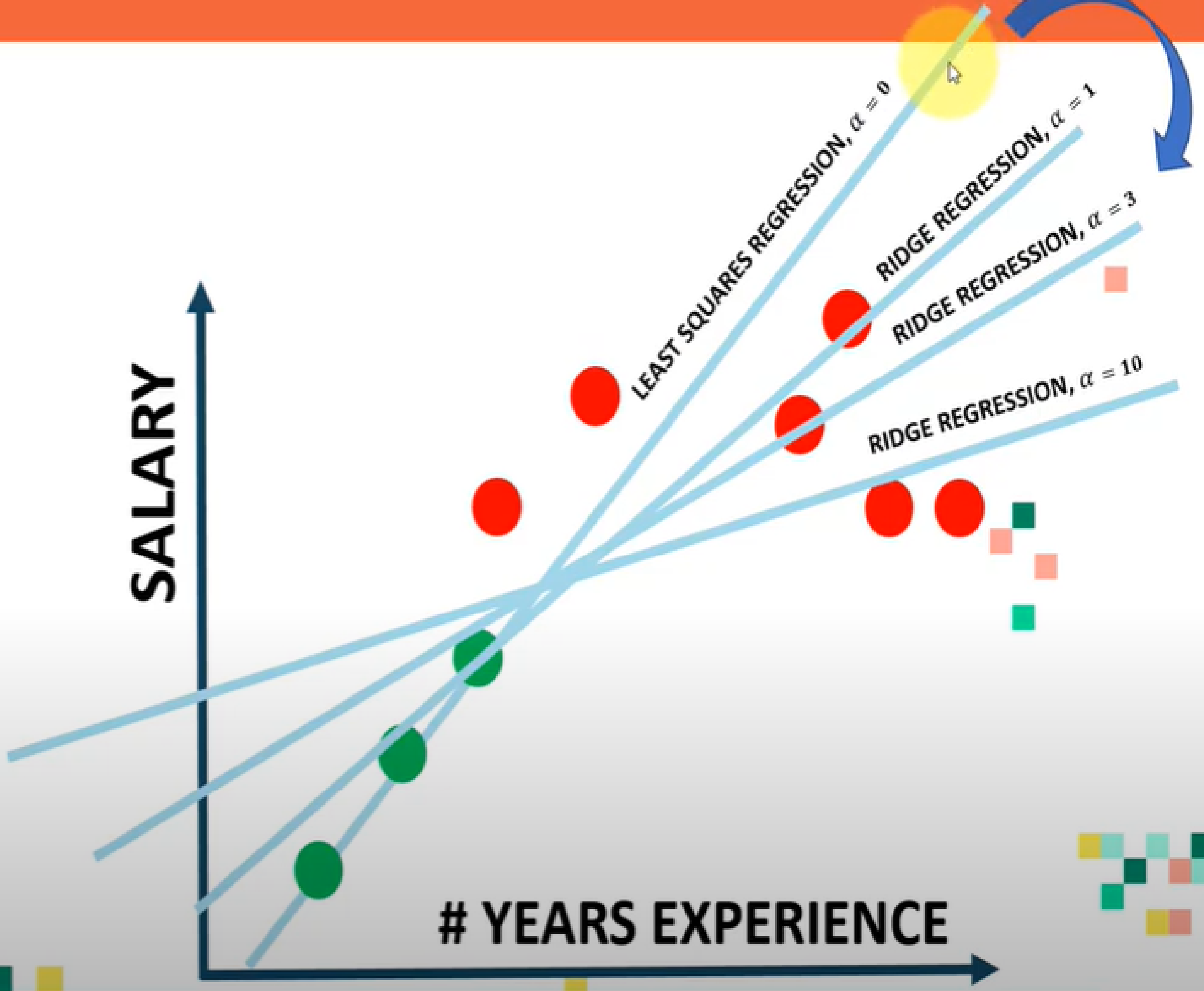
Ridge Regression:

$$Min(sum\ of\ squared\ residuals + \boxed{\alpha * slope^2})$$



RIDGE REGRESSION (L2 REGULARIZATION): ALPHA EFFECT

- As Alpha increases, the slope of the regression line is reduced and becomes more horizontal.
- As Alpha increases, the model becomes less sensitive to the variations of the independent variable (# Years of experience)



LASSO REGRESSION: MATH

- Lasso Regression is similar to Ridge regression
- It works by introducing a bias term but instead of squaring the slope, the absolute value of the slope is added as a penalty term

PENALTY TERM

Least Squares Regression:

$$Min(sum\ of\ the\ squared\ residuals)$$

Lasso Regression:

$$Min(sum\ of\ squared\ residuals + \boxed{\alpha * |slope|})$$

LASSO REGRESSION: MATH

- Lasso regression helps reduce overfitting and it is particularly useful for feature selection
- Lasso regression can be useful if we have several independent variables that are useless
- Ridge regression can reduce the slope close to zero (but not exactly zero) but Lasso regression can reduce the slope to be exactly equal to zero.

Least Squares Regression:

$$Min(sum\ of\ the\ squared\ residuals)$$

Ridge Regression:

$$Min(sum\ of\ squared\ residuals + \alpha * slope^2)$$

Lasso Regression:

$$Min(sum\ of\ squared\ residuals + \alpha * |slope|)$$

Elastic net = L1(Lasso Regression) + L2(Ridge Regression)