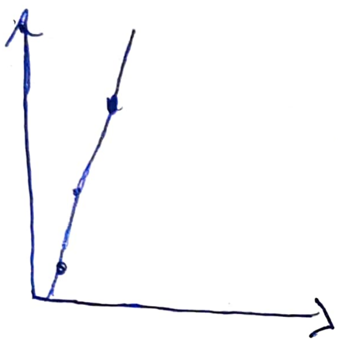
RIDGE and Lasso Regression

to understand RigdGe regression lets understant with example but before example why we have come up with RIDGE and Lasso Regression.

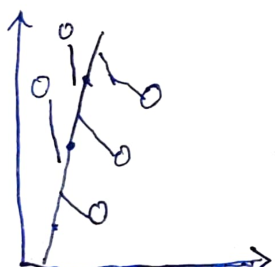
when you feed test data in machine learning it can lead to the problem of overfitting underfitting problem. Lets take one example.



→ let say I had data for traning and it choses this best fit line to predicts the values but what happens when . . . .

→ sum of Residuals will be

$$\sum_{i=1}^{n} (y - \hat{y})^2 = 0$$



when you feed test data and leads to increase in sum of square errors it will lead to poorly predicted values (overfitting and underfitting)

• → training data

o → test data.

* in order to solve this problem we will use Rigde and lasso regression. lets see how it works

all we will focus on sum of residuals inorder to reduce overfitting and underfitting problem.

$(y - \hat{y})^2$ in this formula we will just add $\lambda \times (\text{slope})^2$

Now this lamba if we will increase the slope of the line will decrease so generally we take $\lambda = 1$

Now we will try to reduce the value of this

$$(y-\hat{y})^2 + \underline{\lambda \times (\text{slope})^2} \rightarrow \underline{\text{penalty}}$$

To understand more on this lets paused for now and lets understand on sweets spots

sweets – when model not underfitting or overfitting

overfitting ———— o ———— underfitting

sweet spots.

So to find the sweet spots we use regularization

regularization
- overfitting L_1 – Lasso (least absolute shrinkage..)
- L_2 – Ridge

Lasso (least absolute shrinkage and selection operators)

1) how to choose value for λ (lambda)
you have to tune with the coefficient when
λ = increase the coefficient also = decrease

the essence of coefficient will remain.

L_1 and L_2 regularization is to tackle multi collinearity issue.

in lasso regression the formula will be $y - \hat{y} +$
$\lambda \times |$slope$|$ (magnitude means addition of coefficient)

★ but why lasso not Rigde.

① in order to get less value in Rigde algorithm as we increase λ for Rigde Regression penalty (L2 penalty (the slope) the optimal slope get closer and closer to 0 but it doesn't equal to 0 - the optimal slopes shift towards 0 but we retain nice parabola shape (graph between SSR and $\lambda \times$ slope^2 and slope values).

② in contrast when we increase the lasso penalty aka the L1 penalty ($|$slope$|$) aka absolute value penalty the optimal value shift towards 0

L2 Rigde	L1 losso
Smoths out weight loss curve but and reduce overfitting but higher lamber can kill model training	dragging to zero useful for learnable ignoring of variable useful for high dimensional data at time.

* Regularization is a technique used to used to reduce the error by fitting the function appropriately on the given training set and avoid overfitting and underfitting.

two technique — L1, L2 → which add penalty in loss function slopes