* Regularized linear models

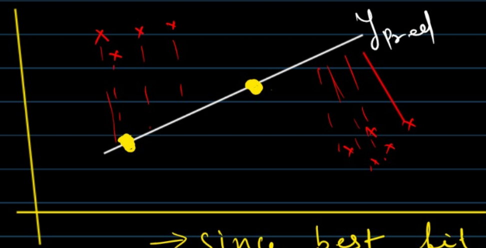
* **Regularisation** → To add something to reduce overfitting

↓

To regularise

↓

To penalise



→ Since best fit line passes through both the data point, the model is a overfitted model.

Analogy

→ wake up, take a bath, formal dress and then reach to office.

→ Say, some day you didn't took a bath and reached the office.

→ Then you will be fined | penalized

→ Similar to this in school days, if you don't bring green vegetable, you were fined.

data → train, test

train: Acc ↑ (low bias & low training error)

test: Acc ↓ (High variance & High testing error)

} → overfitting → overfitted model means model is performing well on train data and bad on test data.

→ Idea of linear Regression is to find best line.

$y = mx + c$

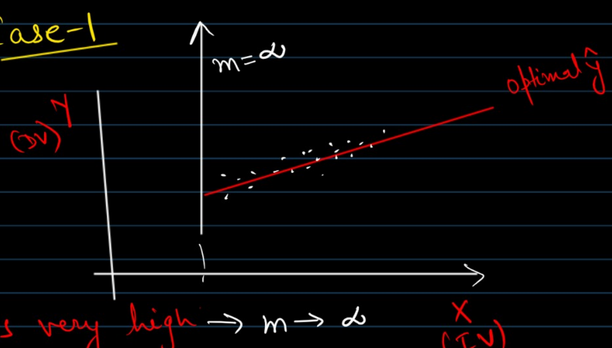
↓

m (magnitude of slope) is very very high

⇓

model has memorised the data point

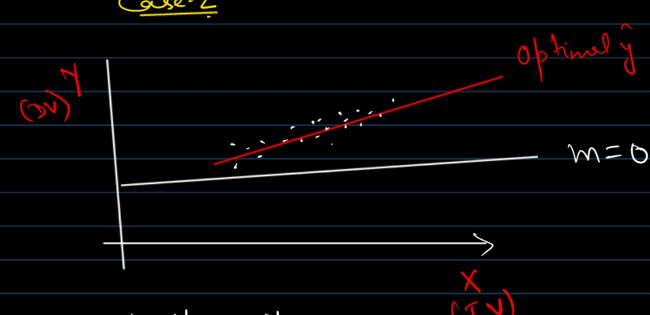
Case-1



→ m is very high → $m \to \infty$

$m = \frac{y_2 - y_1}{x_2 - x_1}$ ($x_2 - x_1 \approx 0$, $m \to \infty$ as x is not changing)

Case-2



→ $m = \frac{y_2 - y_1}{x_2 - x_1} = y_2 - y \approx 0$

for vertical line)

→ $y = mx + C$
$\Downarrow$
∞ → very very large → overfitting

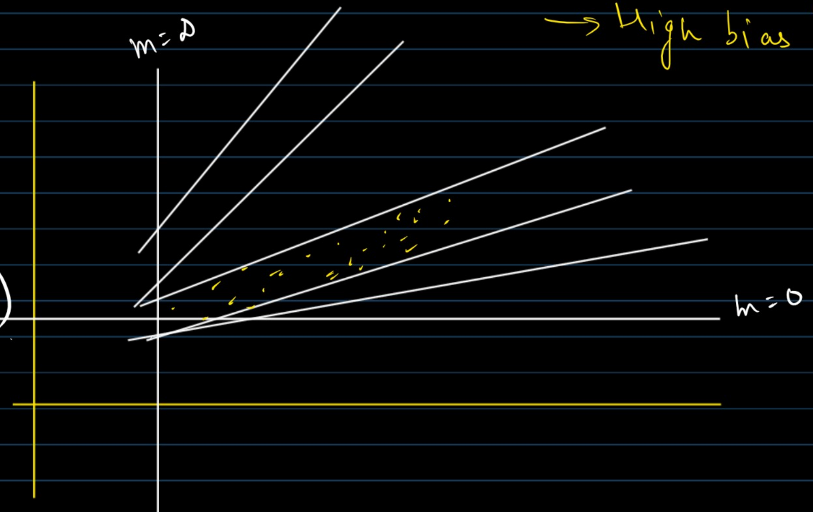
→ $y = mx + C$
↑
0

this implies no use of x in predicting y.

→ High bias model.

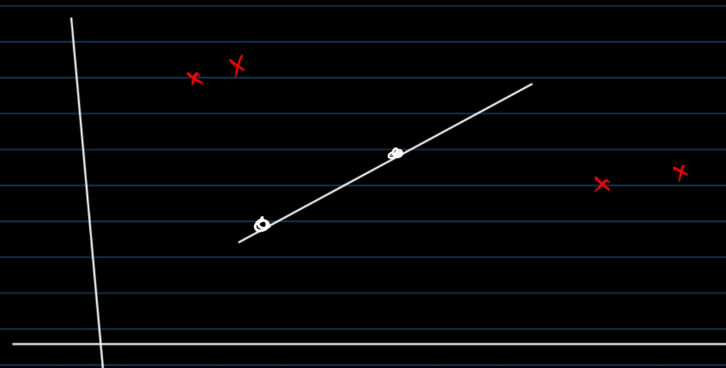
Conclusion

→ The best fit line in our case is between 0 to ∞ (Direct relationship)



Problem of Overfitting

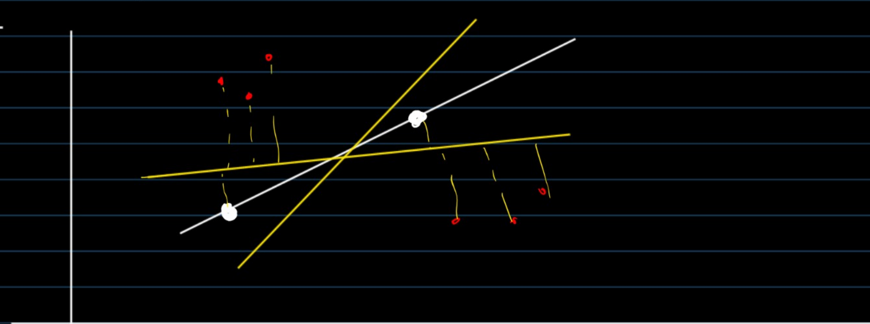
→ Model will not perform well on test data



* What is the solution to the above problem

→ we wanted a model with good accuracy → get coff's using gradient descent.

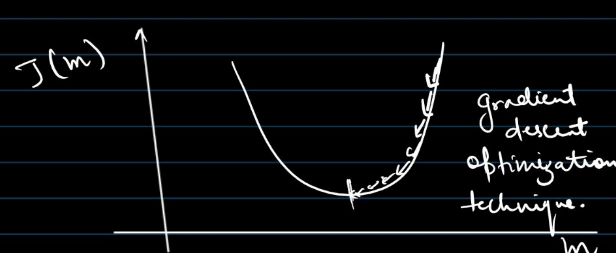
→ In Overfitting, you get 100% accuracy (difference b/w train-test accuracy > 5%)



→ So now in order to make sure that model performs well on test data (unseen) as well, we will introduce some error | penalization while training the model itself.

→ while training we will introduce some error in the model but

How you fit best-fit line



model will perform well on test data.

(Regularisation — To add something to reduce overfitting)

$\Downarrow$

Error

Standard Cost function

$$\text{Cost function} = \frac{1}{n}\sum_{i=1}^{n}\left(y_{act} - y_{pred}\right)^2$$

$y_{pred} \rightarrow \theta_0 + \theta_1 x$

$c + mx$

$\theta_1, m \rightarrow$ slope

θ_0, c – intercept

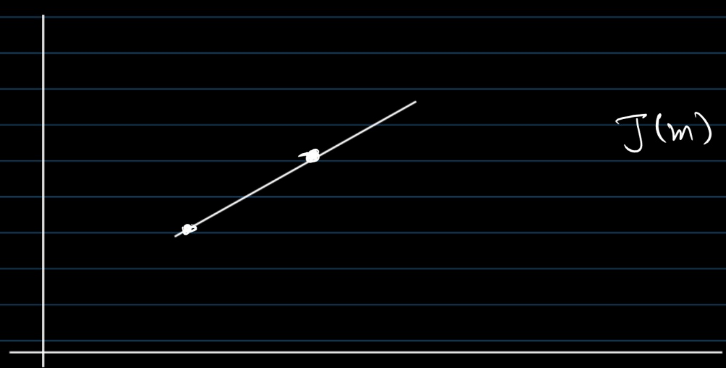
Regularized Linear model.

$$\text{Cost fun} = \frac{1}{n}\sum_{i=1}^{n}\left(y_{act} - y_{pred}\right)^2 + \lambda m^2$$

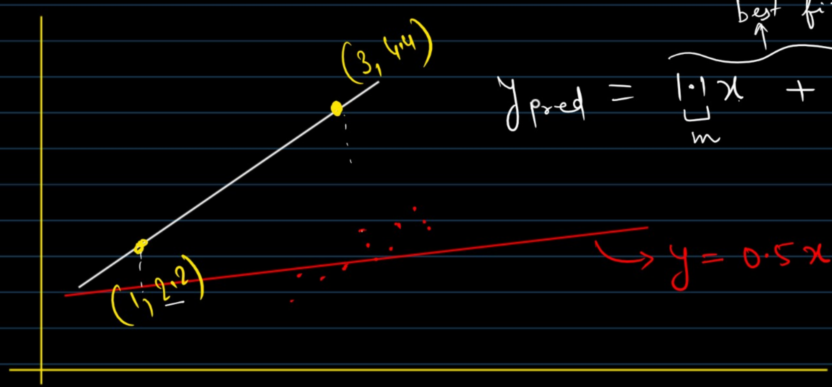
$y_{pred} = mx + c$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(y_{act} - mx_i - c\right)^2 + \lambda m^2$$

$\rightarrow$ where λ is regularisation parameter.



$$J(m) = \frac{1}{n}\sum_{i=1}^{n}\underbrace{\left(y_{act} - y_{pred}\right)^2}_{\Downarrow\ 0} + \lambda m^2$$



$$y_{pred} = \overbrace{\underbrace{1.1}_{m}x + \underbrace{2.1}_{c}}^{\text{best fit line}}$$

$\rightarrow y = 0.5x + 2.7$ (Hypothetical model)

Case-1 (Without any regularisation)

$y = 1.1x + 2.1$ $\quad \lambda = 0$

$\rightarrow$ overfitted line

$C.F = (y_{act} - y_{pred})^2$ $\quad \lambda m^2 = 0$ $\Downarrow$ No regularisation

$\Downarrow$

0

$CF = 0$

(Accuracy is 100%)

Case-2 (With regularisation) $\lambda = 1$

$y = 1.1x + 2.1$

$CF = (y_{act} - y_{pred})^2 + \lambda m^2$

$\Downarrow$

$\Rightarrow 0 + 1 \times (1.1)^2$

$\Rightarrow 1.21$

$\rightarrow$ Due to regularisation, inspite of all the dp's lying on best fit line, there is some error.

Case-3

(another hypothetical line) with $\lambda = 1$

as compared with original line

$y_{pred} = 0.5x + 2.7$

$CF = (y_{act} - y_{pred})^2 + \lambda m^2$

from original line

$y_{pred} = mx + c$

$\left\{ \begin{array}{l} y_{act} = 2.2 \\ x_{act} = 1 \end{array} \right\} \rightarrow$ first dp

$\left\{ \begin{array}{l} y_{act} = 4.4 \\ x_{act} = 3 \end{array} \right\} \rightarrow$ 2nd dp

$CF = (y_{act} - mx - c)^2 + \lambda m^2$

$\Rightarrow \underbrace{(2.2 - 0.5 \times 1 - 2.7)^2}_{1st\ dp} +$

$\underbrace{(4.4 - 0.5 \times 3 - 2.7)^2}_{2nd\ dp} +$

$\underbrace{1 \times (0.5)^2}_{\lambda m^2}$

$\Rightarrow 1.29$

$\rightarrow$ the hypothetical line will a lot of error wrt training data

$\rightarrow \lambda m^2$

$\Rightarrow$ bias $\uparrow$ (training error) variance $\downarrow$ (testing error)

$\Rightarrow \underbrace{CF + \lambda m^2}$ $\rightarrow$ Ridge Regression.

* Regularisation Type :-

1. Ridge Regression
2. Lasso Regression
3. Elastic Regression.