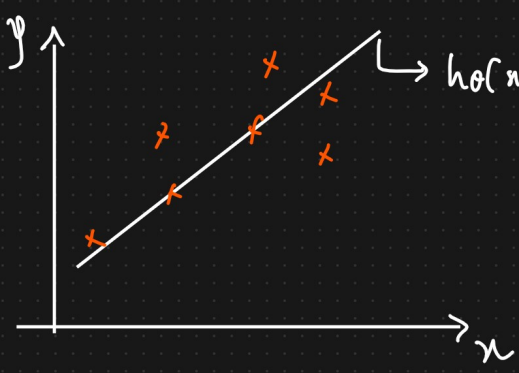
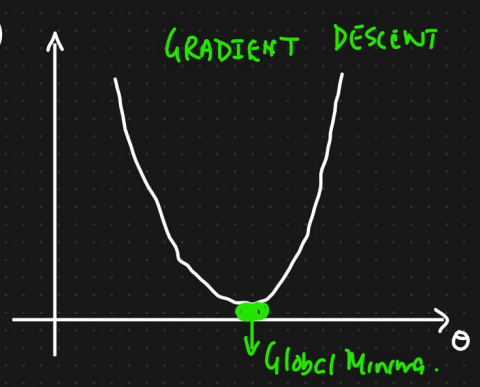
Ridge Regression, Lasso Regression, Elasticnet Regression

write i/w question: types of performance metrix usage? vs

Linear Regression



Independent features



$$\text{Cost fn} = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x)^{(i)} - y^{(i)}\right)^2$$

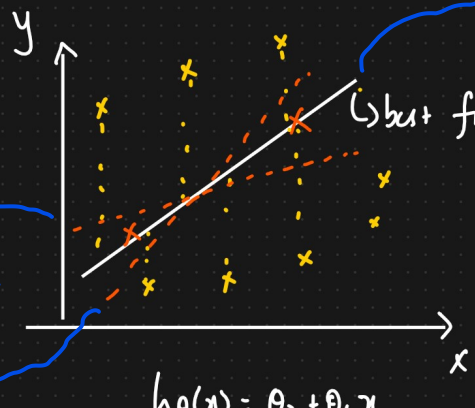
⇓

Mean Squared Error

① Ridge Regression aka (L2 Regularization) → used to Reduce Overfitting

we can also say it like hyperparameter tuning the linear regression

Overfitting



Traindata → Acc ↑ → low Bias

Test data → Acc ↓ → High Variance

to reduce this overfitting - we use ridge regression

OR

$h_\theta(x) = \theta_0 + \theta_1 x_1$

$\lambda = 0$

$\lambda = 1$

$\lambda = 10$

Hyperparameter

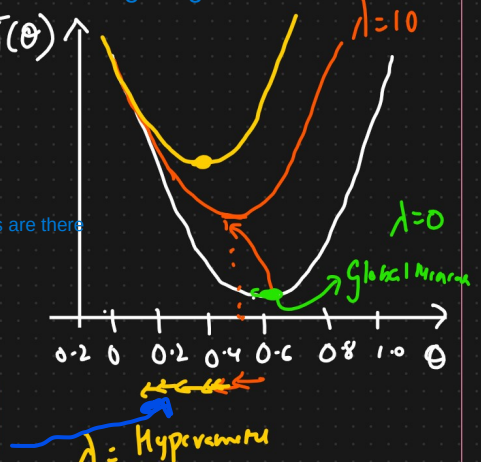
$$\text{Cost fn} = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x)^{(i)} - y^{(i)}\right)^2 + \lambda \sum_{i=1}^{n}(\text{slope})^2$$

n - how many slopes are there

this should never become 0, as it would mean that it's overfitting

$$= 0 + (1)\left[(\theta_1)^2\right]$$

$\lambda \uparrow$ Slope ↓



I/W Question: what is the relationship between lambda and slope

$\lambda = 1$

> 0

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3$$

0.24

assuming

$$= 0.34 + 0.52 x_1 + 0.48 x_2 + 0.24 x_3$$

it means if x1 moves by 1, y will move by 0.52, similarly with x2, x3

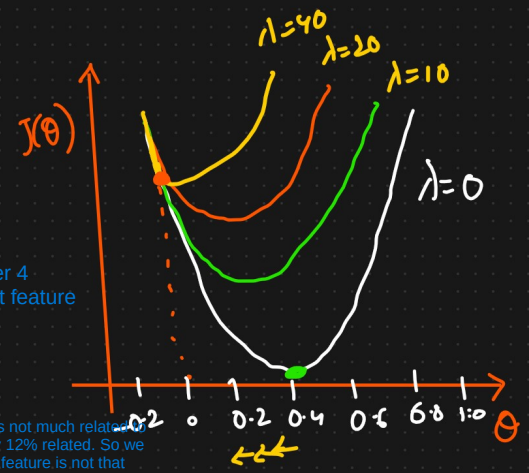
$= 0.34 + 0.40x_1 + 0.38x_2 + \boxed{0.14x_3}$

after applying ridge regression, values will decrease. the value will never become 0

② Lasso Regression (L_1 Regularization) → Feature Selection

here also theta value decreases, but at one point of time it will become 0. it will remove that feature which is less correlated

$$\text{Cost fn} = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x)^{(i)} - y^{(i)}\right) + \boxed{\lambda \sum_{i=1}^{n} |\text{slope}|}$$



$$h_\theta(x) = \theta_0 + \theta_1x_1 + \theta_2x_2 + \theta_3x_3 + \theta_4x_4$$

let's consider 4 independent feature

$$h_\theta(x) = 0.52 + 0.65x_1 + 0.72x_2 + 0.34x_3 + \boxed{0.12x_4}$$

you can see x4 is not much related to output as its only 12% related. So we can say that this feature is not that important

Lasso Regression → 0

after applying lasso regression, it will remove that feature by making it's coeff. to 0

$$= 0.52 + 0.51x_1 + 0.60x_2 + 0.14x_3 + \boxed{0 \times x_4}$$

③ ElasticNet Regression

→ ① Reduce Overfitting

→ ② Feature Selection

$$\text{Cost fn} = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x)^{(i)} - y^{(i)}\right)^2 + \boxed{\lambda_1 \sum_{i=1}^{m} (\text{slope})^2} + \boxed{\lambda_2 \sum_{i=1}^{m} |\text{slope}|}$$

m = total number of slopes

λ_1 → Reduce Overfitting

λ_2 → Feature Selection

WE ARE

{ Hyperparameter Tuning the Linear Regression }