Date ________
Page ________

Regularization — It is a technique used to reduce the errors by fitting the function appropriately on the given training set & avoid overfitting.

Types —

i) ~~Se~~ L1 / Lasso regularization
ii) L2 / Ridge regression
iii) Elastic Net

ii) Ridge Regression (L2) — It modifies the overfitted or underfitted model by adding the penalty equivalent to the sum of the squares of the magnitude of coefficients.

lambda

$$L2 = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^i) - y^{(i)})^2 + \lambda (\text{Slope})^2$$

or

$$\Rightarrow \sum_{i=1}^{m} (y_i - \hat{y}_i)^2 + \lambda m^2$$

or

$$\Rightarrow RSS + \lambda m^2$$

i) Lasso regularization (L1) — The regularization technique perform L1. It modifies the RSS by adding the penalty equivalent to the sum of absolute value of coefficients.

Date
Page

$$L_1 = (h_\theta(x)^i - y^{(i)})^2 + \lambda\,|slope|$$

or

$$RSS + \lambda \sum^{w} |m|$$

Lasso (L_1) is help in Overfitting prevent & helps to do Feature Selection,

Ridge (L_2) help in Overfitting.

- $\hat{y}$ = y hat
- $\lambda \Rightarrow$ lambda

α = alpha
β = beta

Σ = Submission

→ Elastic Net - This method used to perform variable selection & regularization simultaneously.

It is hybrid of Lasso & Ridge

$$\rightarrow \frac{(y_i - \hat{y}_i)^2}{2n} + \lambda\left(\frac{1-\alpha}{2}\right)\Sigma\beta^2 + \alpha\Sigma|\beta|$$

$\boxed{\text{here } \beta = m}$

Q How you will able to build a robust model.

Ans - I will try to use regularization model

Date
Page

(feature Scaling) Standardization & Normalization

i) Standardization / Z-score Normalization — It typically means rescales
Standard - Scaling

data to have a mean of 0 & Standard deviation (σ) of 1. It is often called Z-score.

$$z = \frac{X - \mu}{\sigma} \quad \text{- zscore} = \frac{x - \text{mean}}{S.D}$$

∴ If the mean = 0 & SD = 1 that the output will consider as Z-score.

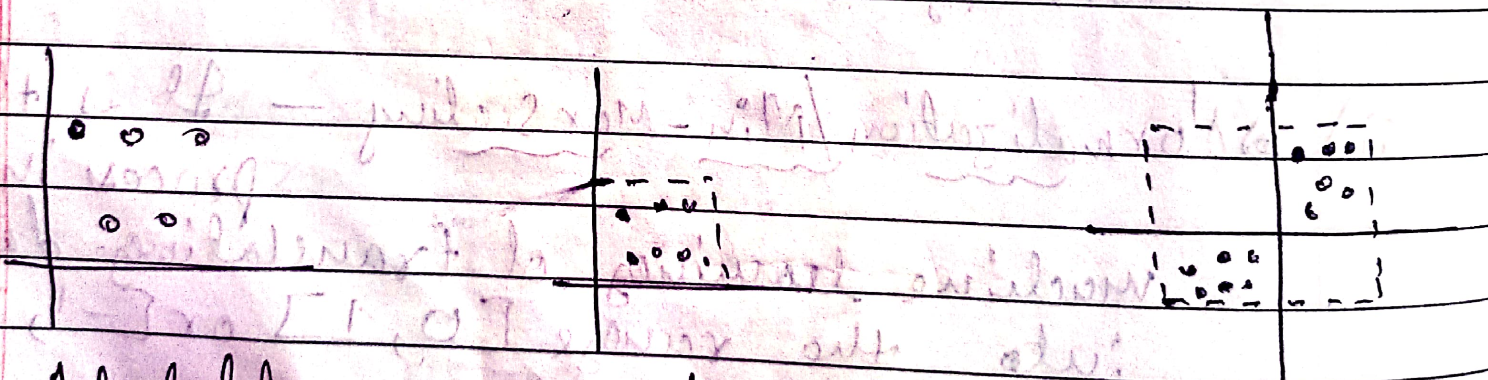
→ It is used when the feature distribution is Normaly or Gaussian distribution.

ii) Normalization / Min-Max Scaling — It is the process in machine learning of translating data into the range [0, 1] or [-1, 1]

$$x_{new} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Difference between

	Normalization (Min-max)	(Z-score) Standardization
i)	Scale value between [0, 1] or [-1, 1].	It is not bounded to a certain range.
ii)	It is really affected by outlier	Not affected or less affected by outlier
iii)	It is useful when we don't know the distribution	It is useful when the feature distribution is Normal/Gaussian
iv)	It is used when feature are of different scales	It is used when we want to ensure zero mean and unit standard deviation



Actual data   After Normalization   After Standardization

Why to use Standard Scaler-

If my data despersion is varing a lot / lot of variances. So in that case my model is not going to understand a relation with respect to independence to dependence variable in a better way. So to give a better relation among feature & label variable we use Standard Scaler.

Standardization v/s Normalization-

Most oftenly Standardization use for clustering analyses, PCA.
Normalization prefer for Image processing because pixel intensity between 0 to 255, neural network algorith require data in scale 0-1, KNN.