Date

# CORELATION COEFFICIENT :-

strongly inversely

Range (-1 — 1) strongly directly

Correlation coefficient is used to tell the measure of strength of propotionality / dependance b/w two quantities.

Std Cordinates are unitless.

Prove with an e.g that correlation - coefficient is not affected by the phenomena translation & scaling.

Scaling can make the Corelation coefficient negative. (through multiple)

correlation with outliers reduce to which we made boxplots too → why??

to better visualize the outliers.

Now we shall remove the outliers.

the does doing it would improve the correlation have good effect.

Date

# USING CORRELATION COEFFICIENT TO PREDICT.

feature y

$(Y/O)$ → (P) → predicted datapoint

at $0^{th}$ index

$(\hat{\ })^P$ → standard data point

$(\hat{Y}/O)^P$ ⇒ shows a predicted feature y at the index O.

$(\hat{Y}_i)^P = a\hat{x}_i + b$ → linear predictor → used because we saw a linear trend.

↳ using the type of predictor is found by

(i) Visualizing the scatterplot.

(ii) Value of correlation coefficient.

slope. y-intercept

Now to measure the level of error of our predictor we would compare it with actual value.

predicted.

$$v_i = (\hat{Y}_i) - (\hat{Y}_i)^P$$

Now we after collecting our all error values we take mean: mean $\{(v)\} = 0$

↓

i.e error does not exist / ideal case.

calculating ideal value of b.

Proof (imp! in slide)

Remember; mean of std. coordinate = 0.

calculating ideal value of a.

Proof (2):

$$\text{Corr coeff} = \frac{1}{N}\sum_{i=1}^{N} \hat{x}_i \hat{y}_i \Rightarrow (2)$$

; mean of the square of a std coordinate is 1.