**REGRESSION METRICS ERROR**

**Mean Absolute Error (MAE):**

Absolute Error : Means hum koi sign consider nhi karenge means mere predicted y me or data points me jo distance ae ga oska koi sign nhi hoga cause existing values predicted value se bari bhi ho sakti hai or predicted values existing values se bari bhi hosakti ha tu distance positive or negative dono a sakta so hum isko absolute consider karenge.

FORMULA FOR IT : | ACTUAL VALUE – PREDICTED VALUE |

Mean absolute error aese kaam karta hai k hum har point ka predicted point se distance nikalenge. For annotation

y = actual value

Y = predicted value

Tu kch is tarha bane ga k har point apne predicted point se subtract hoga or total no of points se divide tu hamare pass mean absolute error ajai ga. Like this if there are 3 points

**(| y1 – Y1 | + | y2 – Y2 | + | y3 – Y3 |)/3**

This how I get my mean absolute error.

Main Purpose is to reduce the mean absolute error as much as possible.

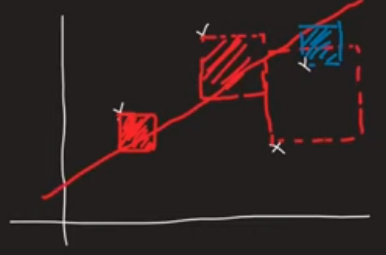
ADVANTAGE:

1. The value MAE is showing have the same unit of y means, mae ki value jo hai wo y axis par jo values hoti hai ose represent karti hai like : x axis pe cgpa hai or y par k package kitne ka lage ga in LPA tu agar MAE 1.5 ata hai tu iska mtlb hai jo package lag rha osme 1.5 lpa ka difference hai. Basically MAE hume loss btata hai.
2. Outlies efficiently handle karta hai

DISADVANTAGE:

1. Neglected values and 0th values ko graph par proper show nhi karta.

MEAN SQUARED ERROR (MSE)



Mse is same as mae , bas hum mae me mod use kar rhe the or ab yaha square use karenge,

Mae me srf distance nikal rhe the actual or predicted value ka or Mse me pora square calculate karlete hai jo oper screen shot me dekha ja sakta hai.

FORMULA FOR MSE:

If we have 3 data points,

**(( y1 – Y1)\*\*2 + (y2 – Y2)\*\*2 + (y3 – Y3 )\*\*2)/3**

ADVANTAGE

Can be use as a loss function

DISADVANTAGE

Difficult to interpret output

Not robust to outliers give deep impact on output

DIFFERENCE

When data have to many outlier use MAE, and when there are few use MSE.

ROOT MEAN SQUARED ERROR(RMSE)

Same as mse , just sqrt in a formula like

FORMULA:

**Sqrt((( y1 – Y1)\*\*2 + (y2 – Y2)\*\*2 + (y3 – Y3 )\*\*2)/3 )**

ADVANTAGE

SAME UNIT AS ON Y AXIS

DISADVANTAGE

ROBUST TO OUTLIERS

R2 SCORE (CO-EFFICIENT OF DETERMINATION)

1. R2 score is independent of units
2. R2 score ye hota hai k mera mean wali line or meri predicted regression ki line mean wali line se kitni bhetar hai.
   1. Mean wali line means srf target column ki values ka mean lelo
3. FORMULA:
   1. 1 – (SSR/SSM)\*\*2
   2. SSR: SUM OF SQUARED REGRESSION (Predicted line)
   3. SSM: SUM OF MEAN REGRESSION (Mean line)

Means ye k hum phele mean wali line se sare actual values ka distance nikalenge or phr sum kardenge (SSM) AND oske bad predicted line or actual values ka distance nikalnege and sum kardenge, phr formula put karenge and R2 square ka result ajai ga

INTERPRETATION OF VALUES COMES FROM R2 SCORE

1. Agar R2 = 0 hai iska mtlb mean wali line or regression line dono ne same plot karlia hai SSR = SSM Or ye bhi bol sakte k jitni glti mean wali line kar rhi otni hi regression bhi kar rhi hai, like this

A diagram of a graph

Description automatically generated

1. Agar R2 = 1 hai iska mtlb SSR = 0 hogaya hai or regression line ne 100% predict karlia hai.
2. SSR k 0 hone k mtlb ye hai k predicted line ne koi glti kari hi nhi hai
3. Agar R2 <1 hai means negative hai tu iska mtlb hai k model ne bhut kharab accuracy di hai means SSR ne kahi aesi jaga line plot Karli hai jaha data hi mojood nhi ya bhut hi door predict kardi line data se SSR > SSM, Like this

A graph of a line with numbers and points

Description automatically generated with medium confidence

1. Neg ane ki waja ye hai k hamara data jo hai wo non linear hai or humne ospar linear regression laga.

ADJUSTED R2 SCORE

Hota ye hai k jab r2 score nikalte waqt hum koi irrelevant column add karde tu r2 score kaam hojata hai tu is type k colums or R2 score k behaviour ko control karne k liye Adjusted R2 score ata hai.

FORMULA:

R2(adj) = 1- [(1 – R2)(n-1)/(n-1-k)]

1. R2 : simple R2 score
2. N: no of rows
3. K: input feature (Independent columns)