Math related to making new feature

Tuesday, 26 August 2025 10:05 am

Working with linear combinations.

$$y = ax_1 + bx_2$$

$$= c(x_1 - x_2) + d(x_1 + x_2)$$ ← double the average.

$$= (c+d)x_1 + (-c+d)x_2$$

swap variable to formulate or interpret.

$\therefore$

$$a = c+d \quad , \quad b = -c+d$$

can sub them in and determine the contribute to the regression.

$$2c = a - b \Rightarrow c = \tfrac{1}{2}(a-b)$$

$$2d = a + b \Rightarrow d = \tfrac{1}{2}(a+b)$$

sub them in

$$z = T_x$$

can write model in terms of

$z \Rightarrow$ mean / difference.

$x \rightarrow$ original

why?

is the change important or the raw absolute variables more impact

Ax Ay | Az (mean) | Ad (different) |

$2 \times 15 = 30$.

```
column1 = [" ", " ", " ", " "]  x
colm 2 = [" ", " ", " ", " "]  y
new_pd = pd.DataFrame(len)   ← size.
for _ in range(len(colm)):
    c3 = df[column1[_]] - df[column2[_]]
    c4 = df[column1[_]] + df[column2[_]]
```

x	y	z	d

30 +

Az	Ad	Bz	Bd	Cz