

CÁLCULO DO DESVIO PADRÃO DA ÁRVORE

Em nossa aula sobre árvore de regressão, usamos a seguinte base de dados como exemplo para construção da árvore:

Aspecto	Temperatura	Umidade	Vento	Total de jogadores
Sol	Quente	Alta	Não	25
Sol	Quente	Alta	Sim	30
Nublado	Quente	Alta	Não	46
Chuva	Branda	Alta	Não	45
Chuva	Frio	Normal	Não	52
Chuva	Frio	Normal	Sim	23
Nublado	Frio	Normal	Sim	43
Sol	Branda	Alta	Não	35
Sol	Frio	Normal	Não	38
Chuva	Branda	Normal	Não	46
Sol	Branda	Normal	Sim	48
Nublado	Branda	Alta	Sim	52
Nublado	Quente	Normal	Não	44
Chuva	Branda	Alta	Sim	30

O cálculo do desvio padrão é obtido da seguinte forma:

$$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

Em que \bar{x} é a média e n é o número de objetos da base.

O desvio padrão do total de jogadores é:

Total de jogadores = {25, 30, 46, 45, 52, 23, 43, 35, 38, 46, 48, 52, 44, 30}

$$\bar{x} = (25 + 30 + 46 + 45 + 52 + 23 + 43 + 35 + 38 + 46 + 48 + 52 + 44 + 30)/14 = 39.78$$

$$\sigma = \sqrt{\frac{(25 - 39.78)^2 + (30 - 39.78)^2 + (46 - 39.78)^2 + \dots + (30 - 39.78)^2}{14}} = 9.32$$

Aspecto (Sol, Nublado, Ensolarado)

Total de Jogadores com aspecto sol = {25, 30, 35, 38, 48}

IMPACTA

$$\bar{x}_{sol} = \frac{25 + 30 + 35 + 38 + 48}{5} = 35.2$$

$$\sigma_{sol} = \sqrt{\frac{(25 - 35.2)^2 + (30 - 35.2)^2 + \dots}{5}} = 7.78$$

Total de jogadores com aspecto nublado = {46, 43, 52, 44}

$$\bar{x}_{nublado} = \frac{46 + 43 + 52 + 44}{4} = 46.25$$

$$\sigma_{nublado} = \sqrt{\frac{(46 - 46.25)^2 + (43 - 46.25)^2 + \dots}{4}} = 3.49$$

Total de jogadores com aspecto chuva = {45, 52, 23, 46, 30}

$$\bar{x}_{chuva} = \frac{45 + 52 + 23 + 46 + 30}{5} = 39.2$$

$$\sigma_{chuva} = \sqrt{\frac{(45 - 39.2)^2 + (52 - 39.2)^2 + \dots}{5}} = 10.87$$

Resumindo, temos:

Aspecto	Desvio Padrão	Objetos
Nublado	3.49	4
Chuva	10.87	5
Sol	7.78	5

Agora nós obtemos o desvio padrão ponderado:

$$\left(\frac{4}{14}\right) * 3.49 + \left(\frac{5}{14}\right) * 10.87 + \left(\frac{5}{14}\right) * 7.78 = 7.66$$

Subtraímos, então, do desvio padrão do total de jogadores o desvio ponderado:

$$9.32 - 7.66 = 1.66$$

Assim, **1.66** é o número que devemos comparar com os demais atributos para escolher a raiz. O objetivo é escolher aquele que fornecer o maior valor.

IMPACTA

Seguindo o mesmo procedimento com os outros atributos, temos:

Temperatura

Temperatura	Desvio Padrão	Objetos
Quente	8.95	4
Frio	10.51	4
Branda	7.65	6

$$Desvio\ ponderado_{temperatura} = \left(\frac{4}{14}\right) * 8.95 + \left(\frac{4}{14}\right) * 10.51 + \left(\frac{6}{14}\right) * 7.65 = 8.84$$

$$9.32 - 8.84 = 0.47$$

Umidade

Umidade	Desvio Padrão	Objetos
Alta	9.36	7
Normal	8.73	7

$$Desvio\ ponderado_{umidade} = \left(\frac{7}{14}\right) * 9.36 + \left(\frac{7}{14}\right) * 8.73 = 9.04$$

$$9.32 - 9.04 = 0.27$$

Vento

Vento	Desvio Padrão	Objetos
Sim	10.59	6
Não	7.87	8

$$Peso\ ponderado_{vento} = \left(\frac{6}{14}\right) * 10.59 + \left(\frac{8}{14}\right) * 7.87 = 9.03$$

$$9.32 - 9.03 = 0.29$$

Ao final da primeira etapa para escolher o atributo raiz temos:

Atributo	Desvio Padrão
Aspecto	1.66
Temperatura	0.47
Umidade	0.27

IMPACTA

Vento	0.29
-------	------

Uma vez que a raiz foi escolhida, prosseguimos com o mesmo cálculo, agora usando o aspecto como ponto de partida. Por exemplo, considerando o aspecto “**sol**”, calcular o desvio padrão de temperatura, umidade e vento.