



Matemática do Calinski (Parte 2)

Transcrição

Faremos o mesmo procedimento para o cluster vermelho, mas dessa vez nomearemos os pontos como D, E e F. Coletaremos a matriz com o valor dos pontos, o centróide e então criaremos a matriz com as diferenças.

Matriz de Variância-Covariância

	x'	y'
D	-0,1	0,0
E	0,1	0,4
F	0,2	-0,2

	x	y
D	1,3	3,9
E	1,5	4,3
F	1,6	3,7

	x	y
ct	1,4	3,9

Multiplicaremos os dois e então teremos novamente uma matriz de variância e covariância, como aprendemos anteriormente. O valor da variância de x será de 0,6 e de y será 0,2.


Faremos o mesmo procedimento para o cluster amarelo, formaremos a matriz e coletaremos os valores de variância para x e y .


Mas agora aprenderemos uma nova forma de obter esses valores, sem necessariamente criar a matriz, mas de qualquer forma é interessante sabermos como ela se estrutura, pois é largamente utilizada em aprendizado de máquina.


A variância nada mais é que o valor da diferença entre o ponto e o centróide, ao quadrado. Como resultado para o cluster amarelo, teremos o valor 0,3. Agora que finalizamos a segunda parte da fórmula para todos os clusters, faremos a segunda, que é somar o resultado das três matrizes de variância e covariância.

Para realizar a soma corretamente, basta somarmos as cédulas que estão na mesma posição.


Soma das Matrizes

	x'	y'
x'	0,06	-
y'	-	0,36

	x'	y'
x'	0,06	-
y'	-	0,2

	x'	y'
x'	0,1	-
y'	-	0,3

Como resultado teremos :

	x'	y'
x'	0,22	-
y'	-	0,86

temores a variância de 0,22 para x e 0,86 para y. Notemos que o valor de x é inferior ao valor de y. A ideia é que o valor de x seja o menor possível, em relação a variável x temos um cluster mais compacto, o que difere no caso y não apresenta a mesma compactação.

Podemos pensar que x seja o valor total de compras, então todos os clientes que estão dentro do cluster possuem um valor mais similar e conciso dentro do grupo, já y pode ser o débito que possuem dentro da empresa de cartão de crédito.

Temos o valor de W_k , agora precisamos do Tr de W_k .

$$s = \frac{\text{tr}(B_k)}{\text{tr}(W_k)} \times 3$$

Tr é o traço, isto é, a soma de diagonais da matriz dos valores que obtemos. $0,22 + 0,86 = 1,08$. A próxima etapa é calcularmos o valor de B_k .