

Cluster Eval
F-Score

Example

X ₁
X ₁ 4
X ₂ 1.1
X ₃ 12
X ₄ 16.4
X ₅ 2.3
X ₆ 5
X ₇ 15
X ₈ 13.7
X ₉ 3.5



$x_2 \quad x_5 \quad x_9 \quad x_1 \quad x_6$

$x_3 \quad x_8 \quad x_7 \quad x_4$

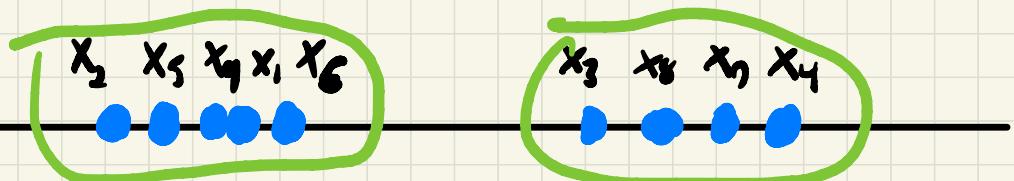
$$\Gamma = \{T_1, T_2\} = \left\{ \{x_1, x_2, x_5, x_8, x_9\}, \{x_3, x_8, x_7, x_4\} \right\}$$

$x_2 \quad x_5 \quad x_9 \quad x_1 \quad x_6$

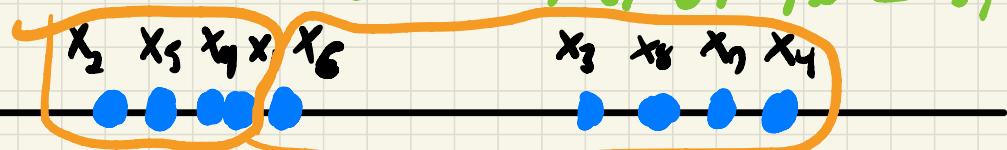
$x_3 \quad x_8 \quad x_7 \quad x_4$

$$C = \{C_1, C_2\} = \left\{ \{x_2, x_5, x_9, x_1\}, \{x_8, x_3, x_7, x_4\} \right\}$$

Contingency Table



$$T = \{T_1, T_2\} = \left\{ \{x_1, x_2, x_5, x_8, x_9\}, \{x_3, x_8, x_7, x_4\} \right\}$$



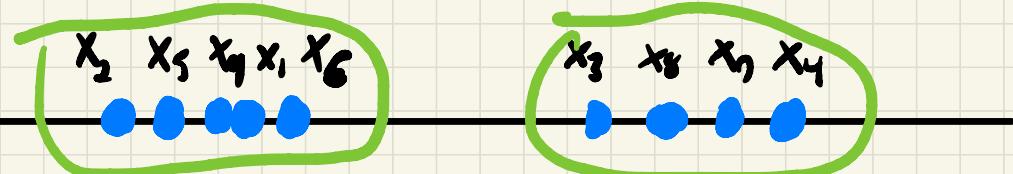
$$C = \{C_1, C_2\} = \left\{ \{x_3, x_5, x_9, x_1\}, \{x_8, x_3, x_8, x_7, x_4\} \right\}$$

Contingency Table

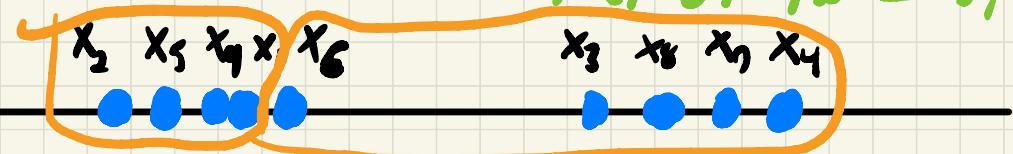
$$\text{prec}_i = \frac{1}{|C_i|} \max_{j=1}^k \{n_{ij}\}$$

C_1 | T_1 T_2 e.g. $\text{prec}_1 =$

C_2 | 4



$$T = \{T_1, T_2\} = \left\{ \{x_1, x_2, x_5, x_8, x_9\}, \{x_3, x_8, x_7, x_4\} \right\}$$



$$C = \{C_1, C_2\} = \left\{ \{x_2, x_5, x_9, x_1\}, \{x_6, x_3, x_8, x_7, x_4\} \right\}$$

Contingency Table

	T ₁	T ₂
C ₁	4	0
C ₂	1	4

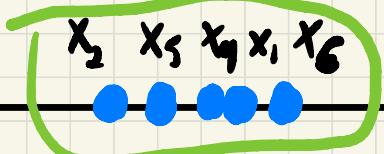
$$\text{prec}_i = \frac{1}{|C_i|} \max_{j=1}^k \{n_{ij}\}$$

$$\text{prec}_2 =$$



$$\Upsilon = \{T_1, T_2\} = \left\{ \left\{ x_1, x_2, x_5, x_8, x_9 \right\}, \left\{ x_3, x_8, x_7, x_4 \right\} \right\}$$

How can I hack precision? (Do well w/
something that is almost always wrong)



$$T = \{T_1, T_2\} = \left\{ \{x_1, x_2, x_5, x_6, x_9\}, \{x_3, x_8, x_7, x_4\} \right\}$$



$$C = \{C_1, C_2\} = \left\{ \{x_2, x_5, x_9, x_1\}, \{x_6, x_3, x_8, x_7, x_4\} \right\}$$

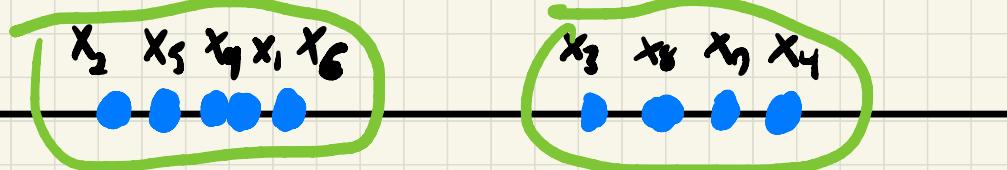
Contingency Table

	T ₁	T ₂
C ₁	4	0
C ₂	1	4

$$\text{recall}_i = \frac{n_{ij^*}}{\sum_j n_{ij}}$$

w/ $j^* = \arg \max_{j=1}^k \{n_{ij}\}$

ex



$$T = \{T_1, T_2\} = \left\{ \{x_1, x_2, x_5, x_8, x_9\}, \{x_3, x_8, x_7, x_4\} \right\}$$



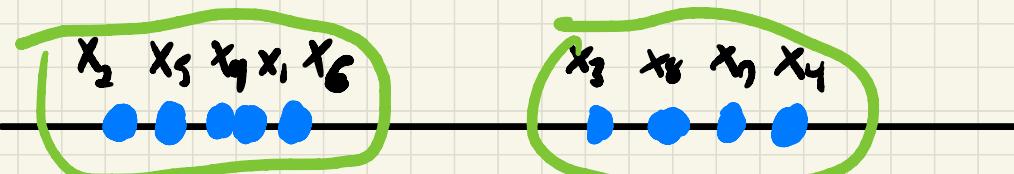
$$C = \{C_1, C_2\} = \left\{ \{x_2, x_5, x_9, x_1\}, \{x_8, x_3, x_7, x_4\} \right\}$$

Contingency Table

	T	T'
C ₁	4	0
C ₂	1	4

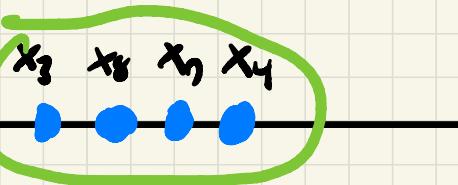
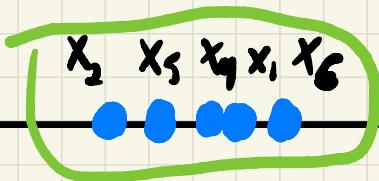
recall₂

$$\text{recall}_{i,j} = \frac{n_{ij,:}}{\|T_{j,:}\|} \quad \text{w/ } j_{\text{c}} = \arg \max_{j=1}^k \{n_{i,j}\}$$



$$T = \{T_1, T_2\} = \left\{ \{x_1, x_2, x_5, x_8, x_9\}, \{x_3, x_8, x_7, x_4\} \right\}$$

How can I hack recall?



$$T = \{T_1, T_2\} = \left\{ \left\{ x_1, x_2, x_5, x_8, x_9 \right\}, \left\{ x_3, x_8, x_7, x_4 \right\} \right\}$$

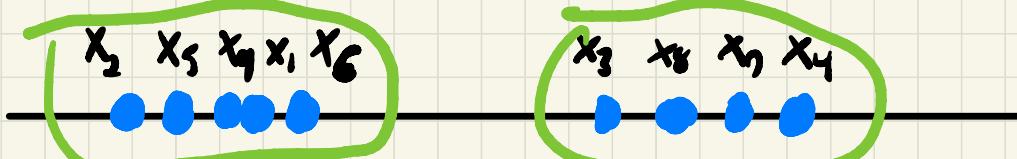


$$C = \{C_1, C_2\} = \left\{ \left\{ x_2, x_5, x_9, x_1 \right\}, \left\{ x_8, x_3, x_7, x_4 \right\} \right\}$$

Contingency Table $F_i = \frac{2(\text{prec}_i)(\text{recall}_i)}{\text{prec}_i + \text{recall}_i}$ $F_1 =$

	T ₁	T ₂
C ₁	4	0
C ₂	1	4

$$F = \sum_{i=1}^k F_i$$



$$\mathcal{T} = \{T_1, T_2\} = \left\{ \{x_1, x_2, x_5, x_7, x_8\}, \{x_3, x_8, x_7, x_4\} \right\}$$



$$\mathcal{C} = \{C_1, C_2\} = \left\{ \{x_1, x_5\}, \{x_9, x_1, x_6, x_3, x_8, x_9, x_4\} \right\}$$

What happens to F-Score?