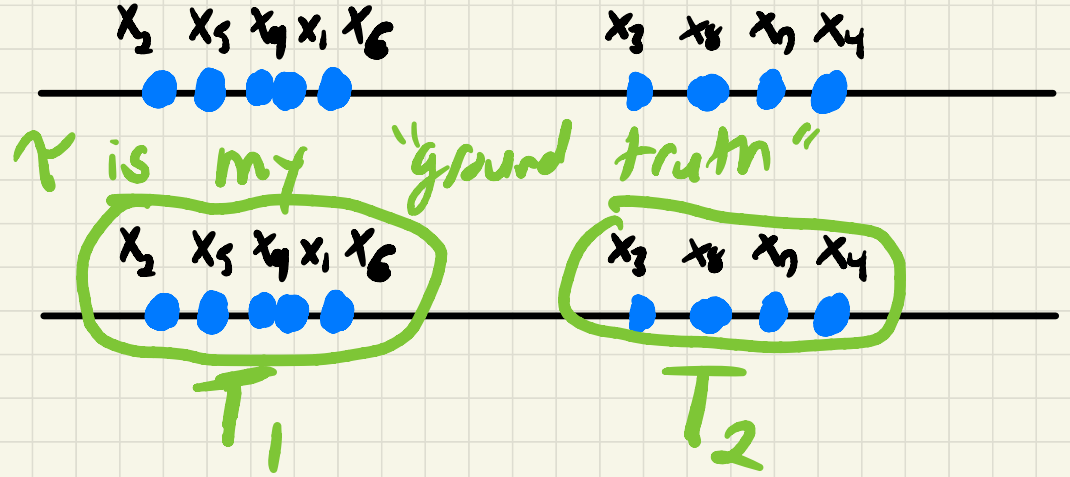
Cluster Eval
F-Score

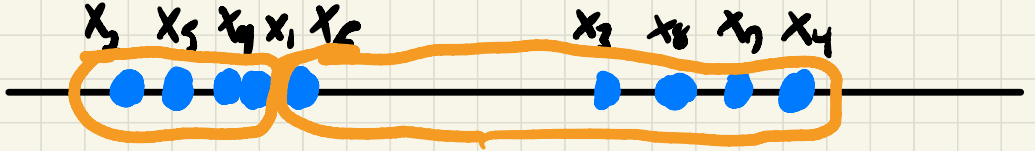
Example

	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

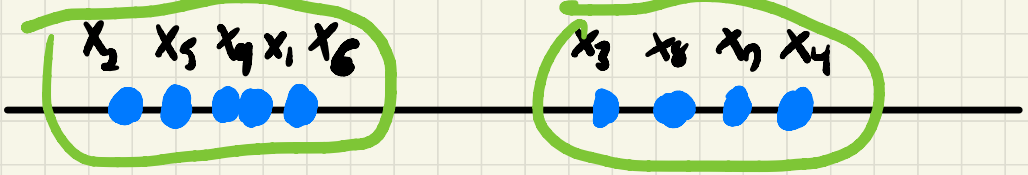


$T = \{T_1, T_2\} = \{\{x_2, x_5, x_9, x_1, x_6\}, \{x_3, x_8, x_7, x_4\}\}$

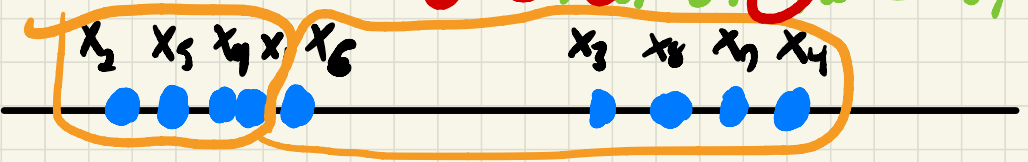
C be computed



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



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



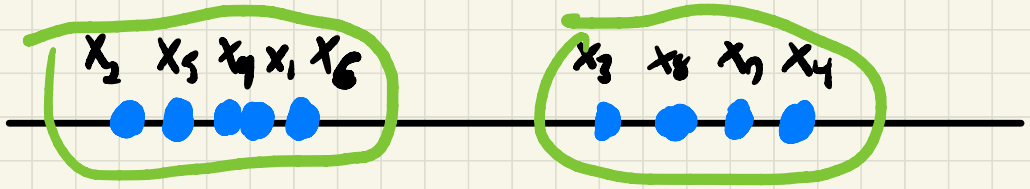
$C = \{C_1, C_2\} = \{\{x_2, x_5, x_9, x_1\}, \{x_6, x_3, x_8, x_7, x_4\}\}$

Contingency Table

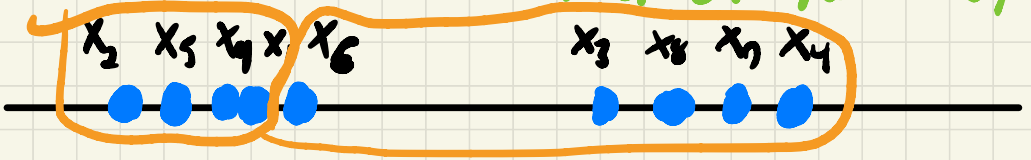
	T_1	T_2
C_1	4	0
C_2	1	4

$$n_{ij} = |C_i \cap T_j|$$

$$n_{11} = |C_1 \cap T_1|$$



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



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

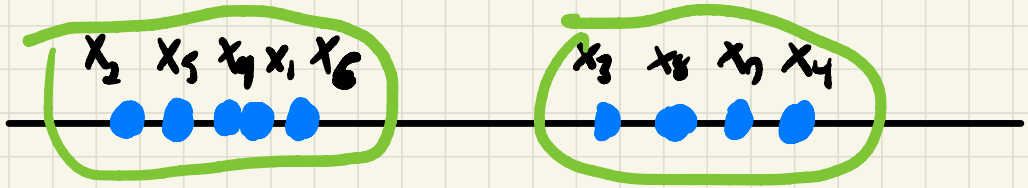
Contingency Table

	T_1	T_2
C_1	4	0
C_2	1	4

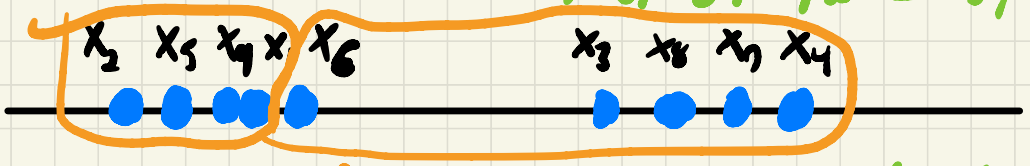
$$prec_i = \frac{1}{|C_i|} \max_{j=1}^{k} \{n_{ij}\}$$

e.g. $prec_1 = \frac{1}{|C_1|} \max\{n_{11}, n_{12}\}$

$= \frac{1}{4} \max\{4, 0\} = \frac{1}{4}(4) = 1$



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



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

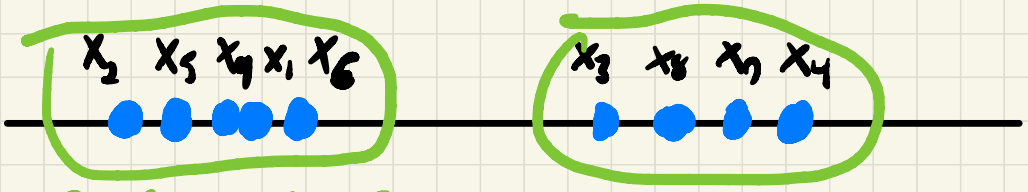
Contingency Table

	T_1	T_2
C_1	4	0
C_2	1	4

$$prec_i = \frac{1}{|C_i|} \max_{j=1}^{k} \{n_{ij}\}$$

$$prec_2 = \frac{1}{|C_2|} \max\{n_{21}, n_{22}\}$$

$$= \frac{1}{5} \max\{1, 4\} = \frac{1}{5} 4 = \frac{4}{5} = .8$$



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

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

$C = \{C_1, \ldots, C_n\}$

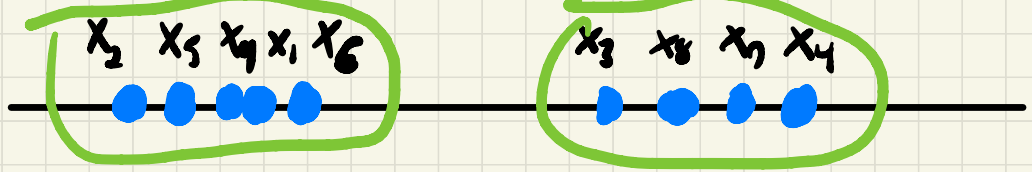
s.t $C_i = \{x_i\}$

$$prec_1 = \frac{1}{|c_1|} \max_{j=1}^{k} \{n_{ij}\} = \frac{1}{1} \cdot \max\{1, 0\} = 1$$

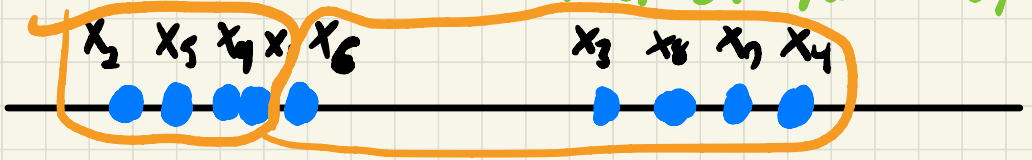
	T_1	T_2
C_1	1	0
C_2	1	0
C_3	0	1
C_4	1	0

$prec_i = 1$

$\forall i$ ∴



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



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

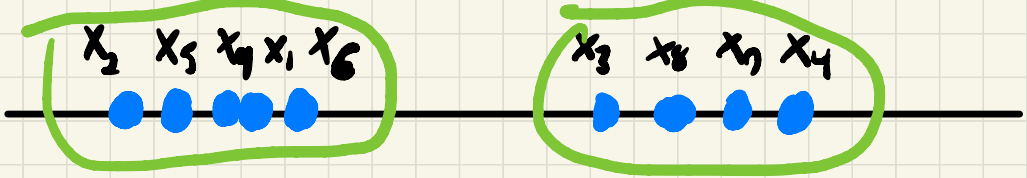
Contingency Table

	T_1	T_2
C_1	4	0
C_2	1	4

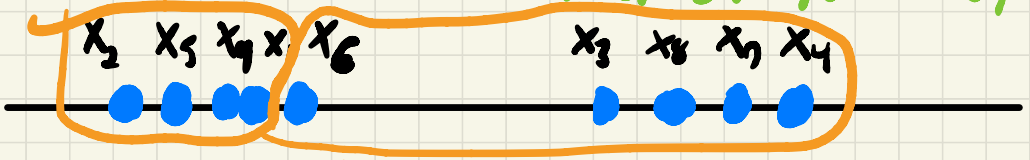
$$\text{recall}_i = \frac{n_{ij_i}}{|T_{j_i}|} \quad \text{w/ } j_i = \underset{j=1}{\overset{k}{\text{argmax}}} \{n_{ij}\}$$

ex $j_1 = \text{argmax}\{4, 0\} = 1$

$\text{recall}_1 = \frac{n_{11}}{|T_1|} = \frac{4}{5} = .8$ (n_{11}, n_{12})



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



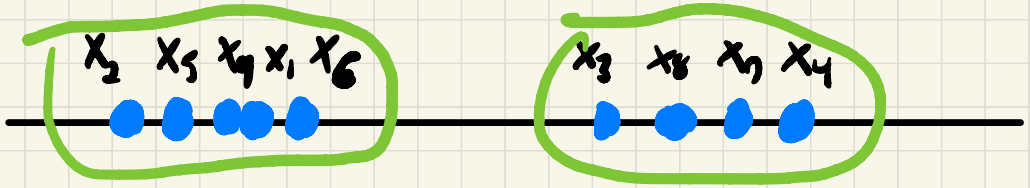
$$\mathcal{C} = \{C_1, C_2\} = \{\{x_2, x_5, x_9, x_1\}, \{x_6, x_3, x_8, x_7, x_4\}\}$$

Contingency Table

	T_1	T_2
C_1	4	0
C_2	1	4

$$\text{recall}_i = \frac{n_{ij_i}}{|T_{j_i}|} \quad \text{w/ } j_i = \underset{j=1}{\overset{k}{\operatorname{argmax}}} \{n_{ij}\}$$

recall_2

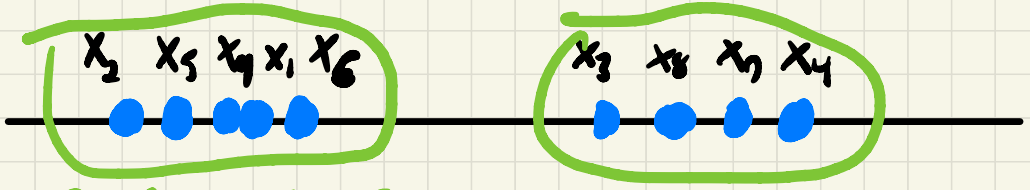


$$T = \{T_1, T_2\} = \{\{x_1, x_2, x_5, x_6, x_9\}, \{x_3, x_8, x_7, x_4\}\}$$

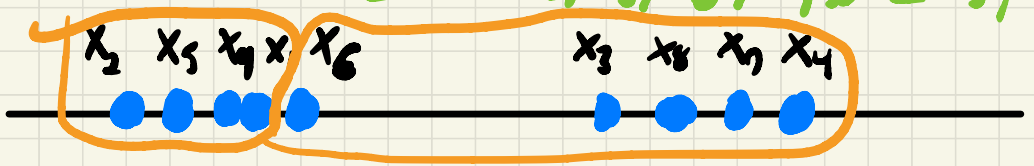
How can I hack recall?

	T_1	T_2
G_1	5	4

$$recall_1 = \frac{n_{11}}{|T_1|} = \frac{5}{5} = 1$$



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



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

Contingency Table

	T_1	T_2
C_1	4	0
C_2	1	4

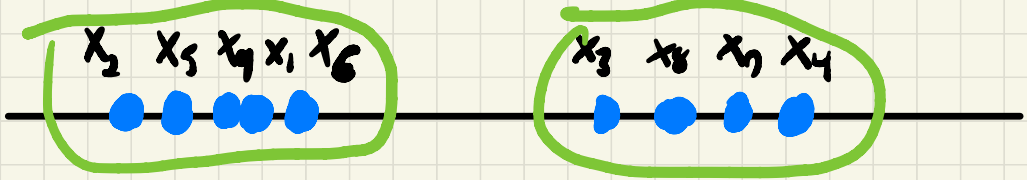
$$F_i = \frac{2(\text{prec}_i)(\text{recall}_i)}{\text{prec}_i + \text{recall}_i}$$

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

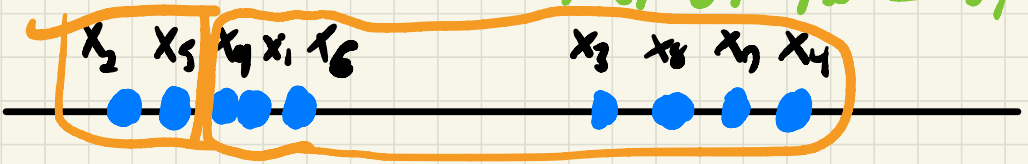
$$F_1 = \frac{(2)(1)(.8)}{1+.8} = .89$$

$$F_2 = \frac{(2)(.8)(1)}{.8+1} = .89$$

$$F = \frac{1}{2}(.89+.89) = .89$$



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



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

What happens to F-Score?

$F_1 = .57$

$F_2 = .73$

$F = \frac{1}{2}(.57 + .73) = .65$