Evaluating clustering

CSCI 347

Adiesha Liyana Ralalage

Clustering validation

- It is important to develop approaches to assess clustering results.
- This process has three main tasks:
  - Clustering evaluation: assess goodness or quality of the clustering
  - Clustering stability: assess the sensitivity of the clustering results to various algorithmic parameters.
  - Clustering tendency: assessing the suitability of applying clustering in the first place, e.g., whether the data has inherent grouping structures.

Clustering validation

- There are three types of validation.
- **External:** External validation measures employ criteria that are not inherent to the dataset. E.g., class labels for each point.
- **Internal:** Internal validation measures employ criteria that are derived from the data itself.
  - we can use intra-cluster and inter-cluster distances to obtain measures of cluster compactness (e.g., how similar are the points in the same cluster) and separation (e.g., how far apart are the points in different clusters).
- **Relative:** Relative validation measures aim to directly compare different clustering, usually those obtained via different parameter settings for the same algorithm.

Evaluating clustering: the F-Score

CSCI 347

Adiesha Liyana Ralalage

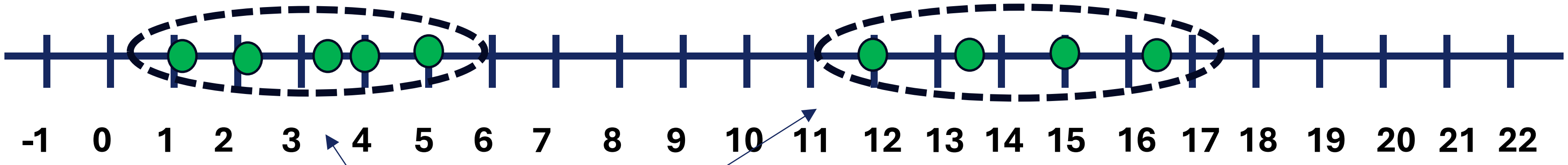
F-Score

- External measure.
- Assumes that the correct or ground-truth clustering is known a priori.
- Ground truth plays the role of external information to evaluate clustering.
- External evaluation measures try capture the extent to which points from the same partition appear in the same cluster, and the extent to which points from different partitions are grouped in different clusters.

Example

- 1-dimensional 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



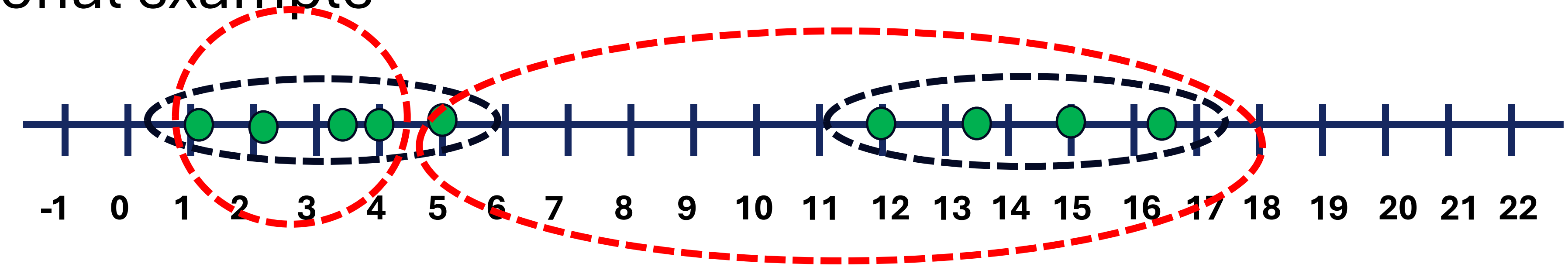
Looks like the true clusters.

Example

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

- 1-dimensional 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



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

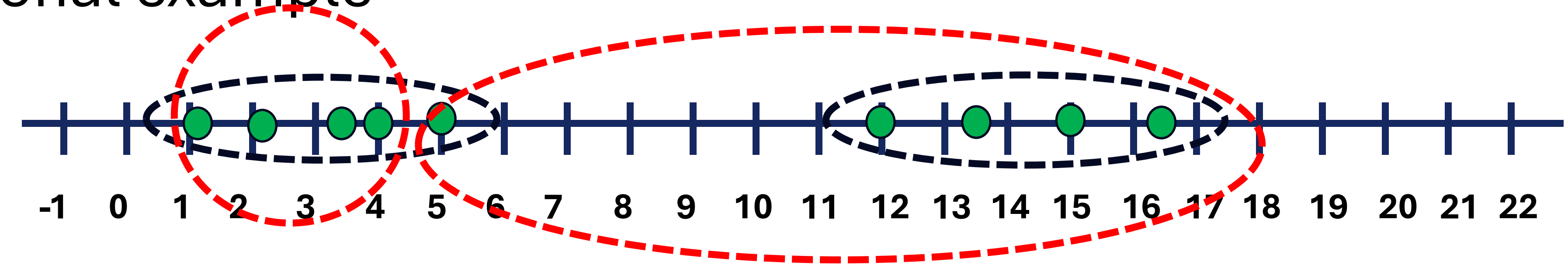
Precision

$$n_1 = 4 \qquad n_2 = 5$$

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

- 1-dimensional 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



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

Contingency table		
	T_1	T_2
C_1	4	0
C_2	1	4

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

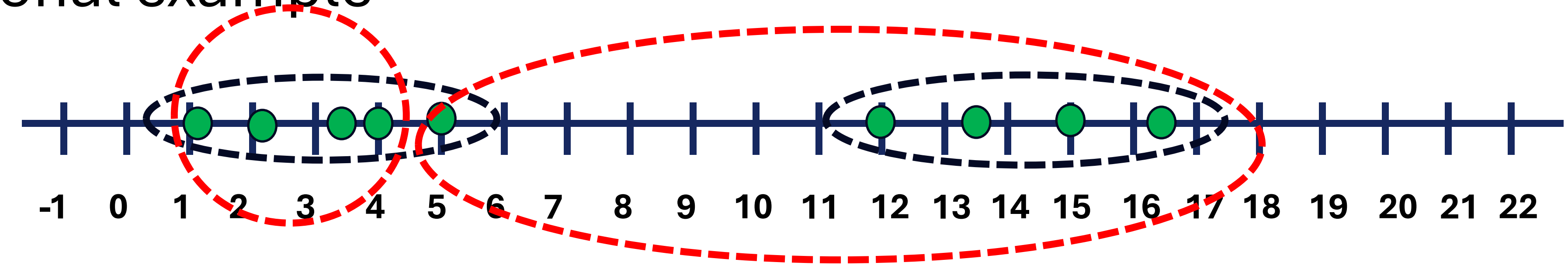
Precision

$$n_1 = 4 \qquad n_2 = 5$$

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

- 1-dimensional 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



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

Contingency table		
	T_1	T_2
C_1	4	0
C_2	1	4

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

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

quantifies the extent to which a cluster C_i contains entities from only one partition

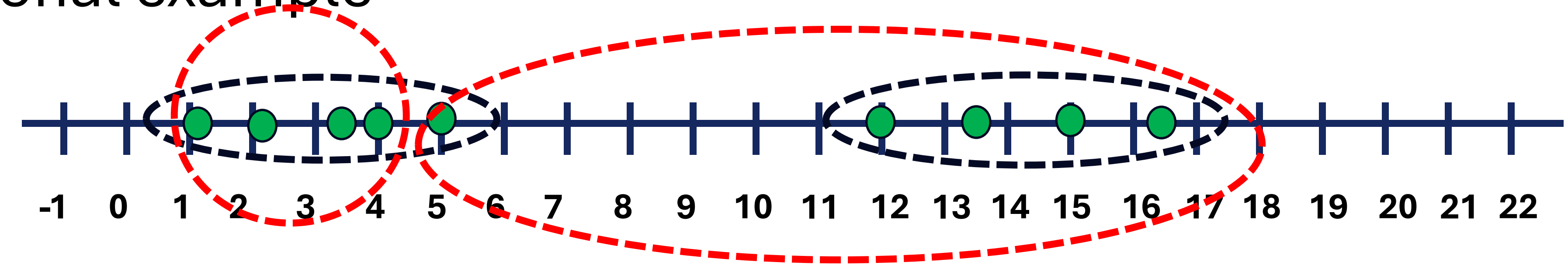
Precision

$$n_1 = 4 \qquad n_2 = 5$$

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

- 1-dimensional 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



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

Contingency table		
	T_1	T_2
C_1	4	0
C_2	1	4

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

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

$$prec_1 = \frac{1}{n_1} max\{n_{11}, n_{12}\}$$

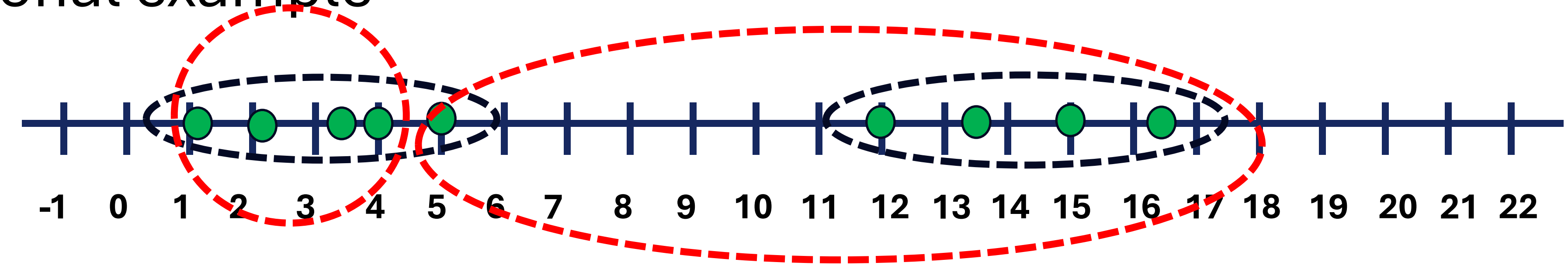
Precision

$$n_1 = 4 \qquad n_2 = 5$$

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

- 1-dimensional 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



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

Contingency table		
	T_1	T_2
C_1	4	0
C_2	1	4

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

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

$$prec_1 = \frac{1}{n_1} max\{n_{11}, n_{12}\}$$

$$prec_1 = \frac{1}{4} max\{4, 0\}$$

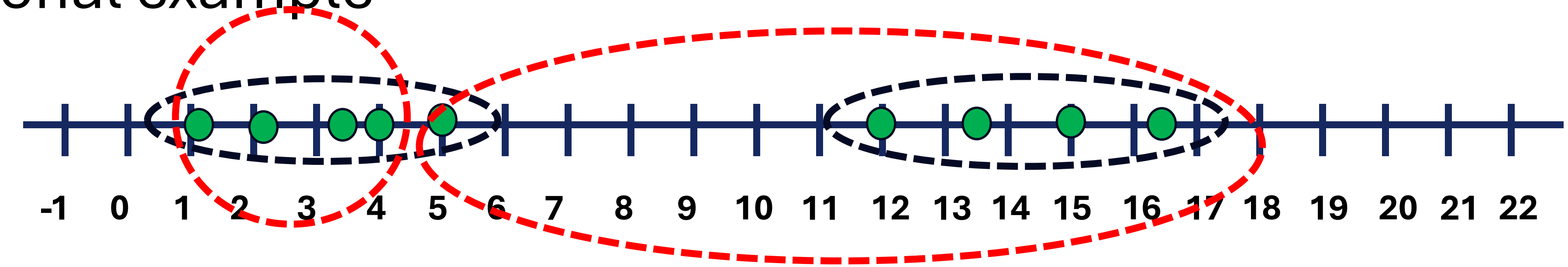
Precision

$$n_1 = 4 \qquad n_2 = 5$$

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

- 1-dimensional 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



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

Contingency table		
	T_1	T_2
C_1	4	0
C_2	1	4

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

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

$$prec_1 = \frac{1}{n_1} max\{n_{11}, n_{12}\}$$

$$prec_1 = \frac{1}{4} max\{4, 0\} = \frac{1}{4} \cdot 4 = 1$$

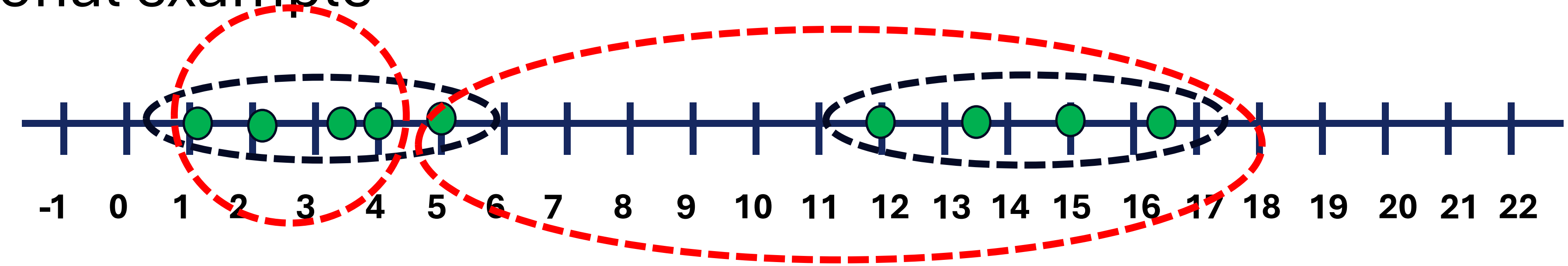
Precision

$$n_1 = 4 \qquad n_2 = 5$$

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

- 1-dimensional 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



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

Contingency table		
	T_1	T_2
C_1	4	0
C_2	1	4

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

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

$$prec_2 = \frac{1}{n_2} max\{n_{21}, n_{22}\}$$

$$prec_2 = \frac{1}{5} max\{1, 4\} = \frac{1}{5} \cdot 4 = 0.8$$

Precision

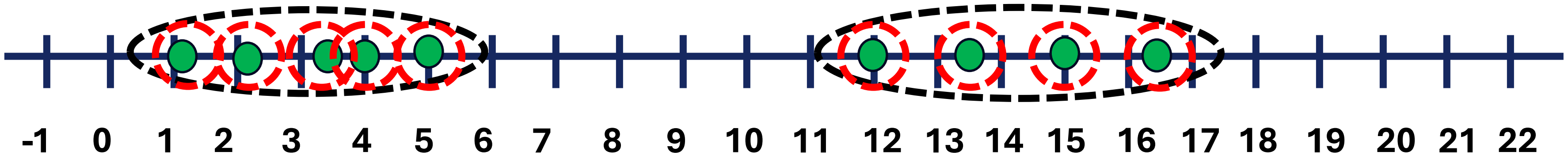
$$\mathcal{C} = \{C_1, C_2, C_3, C_4, C_5, C_6, C_7, C_8, C_9\}$$
$$= \{\{x_1\}, \{x_2\}, \{x_3\}, \{x_4\}, \{x_5\}, \{x_6\}, \{x_7\}, \{x_8\}, \{x_9\}, \}$$

- 1-dimensional 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

Contingency table

	T_1	T_2
C_1	1	0
C_2	1	0
C_3	0	1
C_4	0	1
C_5	1	0
C_6	1	0
C_7	0	1
C_8	0	1
C_9	1	0



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

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

$$prec_1 = \frac{1}{n_1} max\{n_{11}, n_{12}\}$$

$$prec_1 = \frac{1}{1} max\{1, 0\} = \frac{1}{1} \cdot 1 = 1$$

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

Precision

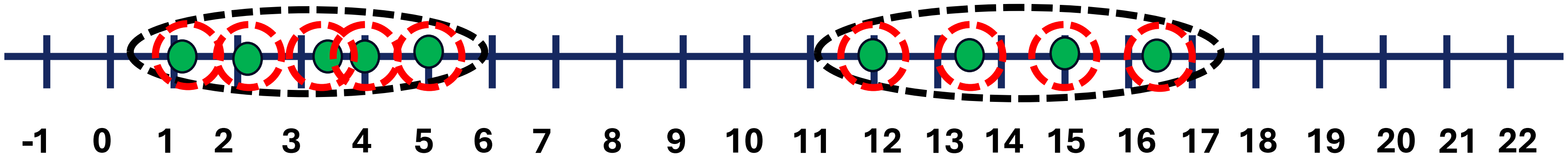
$$\mathcal{C} = \{C_1, C_2, C_3, C_4, C_5, C_6, C_7, C_8, C_9\}$$
$$= \{\{x_1\}, \{x_2\}, \{x_3\}, \{x_4\}, \{x_5\}, \{x_6\}, \{x_7\}, \{x_8\}, \{x_9\}, \}$$

- 1-dimensional 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

Contingency table

	T_1	T_2
C_1	1	0
C_2	1	0
C_3	0	1
C_4	0	1
C_5	1	0
C_6	1	0
C_7	0	1
C_8	0	1
C_9	1	0



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

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

$$prec_2 = \frac{1}{n_2} max\{n_{11}, n_{12}\}$$

$$prec_2 = \frac{1}{1} max\{1, 0\} = \frac{1}{1} \cdot 1 = 1$$

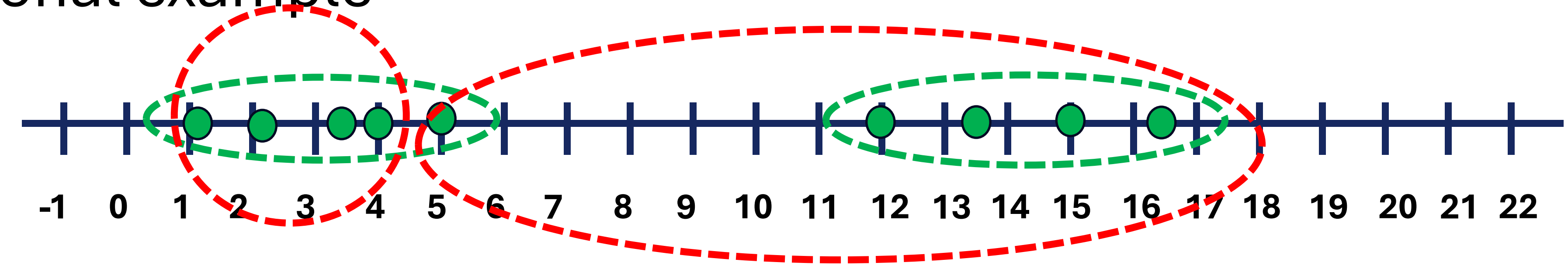
$$n_{ij} = |C_i \cap T_j|$$

Recall

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

- 1-dimensional 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



$T_1 = 5$ $T_2 = 4$

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

Contingency table		
	T_1	T_2
C_1	4	0
C_2	1	4

$$recall_i = \frac{n_{ij}}{|T_{j_i}|} \qquad j_i = argmax_{j=1}^{k}\{n_{ij}\}$$

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

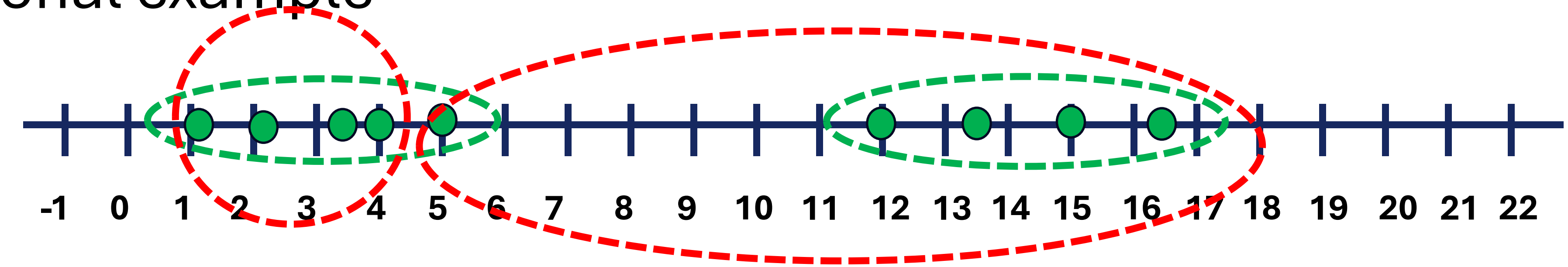
measures the fraction of point in partition T_{ji} shared with cluster C_i .

Recall

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

- 1-dimensional 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



$T_1 = 5$ $T_2 = 4$

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

Contingency table		
	T_1	T_2
C_1	4	0
C_2	1	4

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

$$recall_i = \frac{n_{ij}}{|T_{j_i}|}$$

$$j_i = argmax_{j=1}^{k}\{n_{ij}\}$$

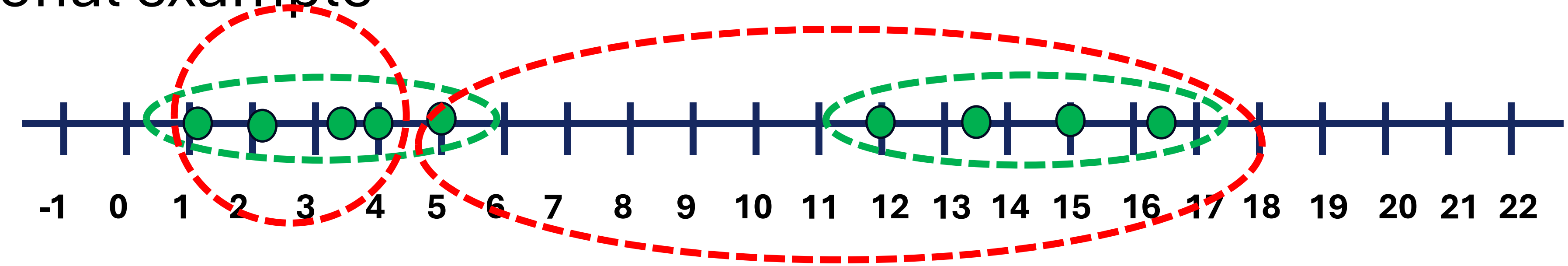
$$j_1 = argmax_{j=1}^{k}\{n_{1j}\} = 1$$

Recall

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

- 1-dimensional 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



$T_1 = 5$ $T_2 = 4$

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

Contingency table		
	T_1	T_2
C_1	4	0
C_2	1	4

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

$$recall_i = \frac{n_{ij}}{|T_{j_i}|}$$

$$j_i = argmax_{j=1}^{k}\{n_{ij}\}$$

$$j_1 = argmax_{j=1}^{k}\{n_{1j}\} = 1$$

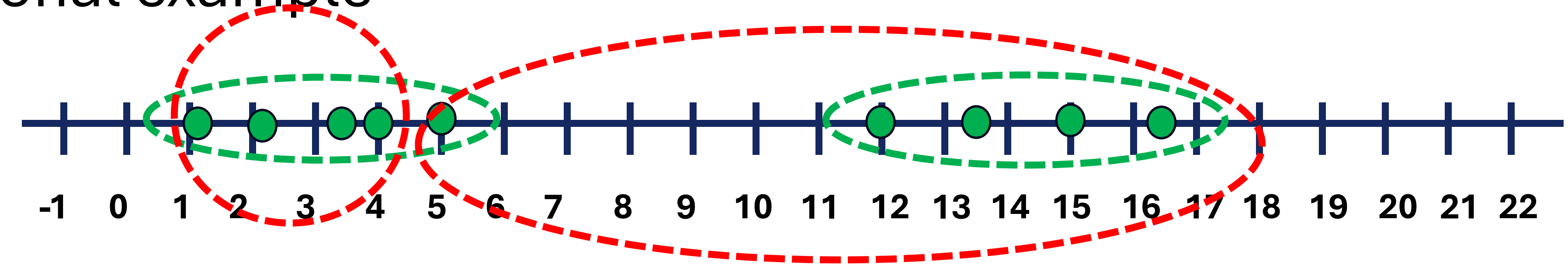
$$recall_1 = \frac{n_{11}}{|T_1|}$$

Recall

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

- 1-dimensional 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



$T_1 = 5$ $T_2 = 4$

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

Contingency table		
	T_1	T_2
C_1	4	0
C_2	1	4

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

$$recall_i = \frac{n_{ij}}{|T_{j_i}|}$$

$$j_i = argmax_{j=1}^{k}\{n_{ij}\}$$

$$j_1 = argmax_{j=1}^{k}\{n_{1j}\} = 1$$

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

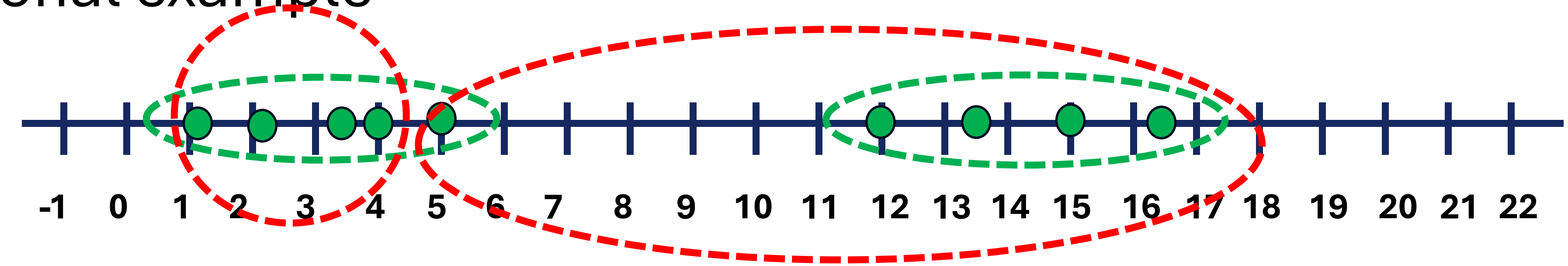
$$recall_1 = \frac{4}{5} = 0.8$$

Recall

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

- 1-dimensional 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



$T_1 = 5$ $T_2 = 4$

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

Contingency table		
	T_1	T_2
C_1	4	0
C_2	1	4

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

$$recall_i = \frac{n_{ij}}{|T_{j_i}|}$$

$$j_i = argmax_{j=1}^{k}\{n_{ij}\}$$

$$j_2 = argmax_{j=1}^{k}\{n_{2j}\} = 2$$

$$recall_2 = \frac{n_{22}}{|T_2|}$$

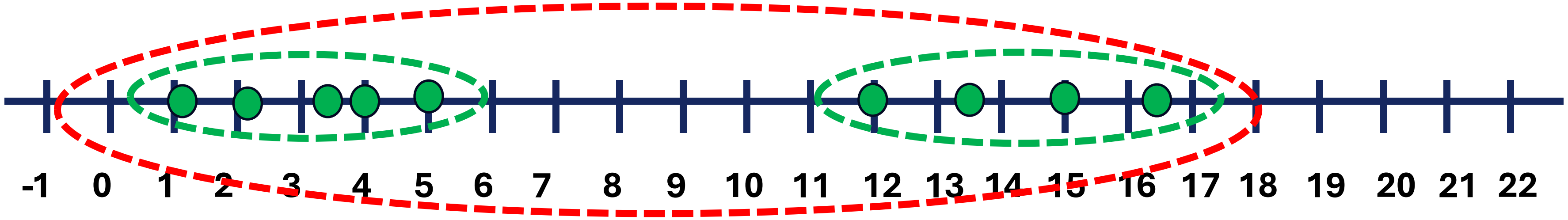
$$recall_2 = \frac{4}{4} = 1$$

Recall

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

- 1-dimensional 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



$T_1 = 5$ $T_2 = 4$

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

Contingency table		
	T_1	T_2
C_1	5	4

$$recall_i = \frac{n_{ij}}{|T_{j_i}|}$$

$$j_i = argmax_{j=1}^{k}\{n_{ij}\}$$

$$j_1 = argmax_{j=1}^{k}\{n_{1j}\} = 1$$

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

$$recall_1 = \frac{5}{5} = 1$$

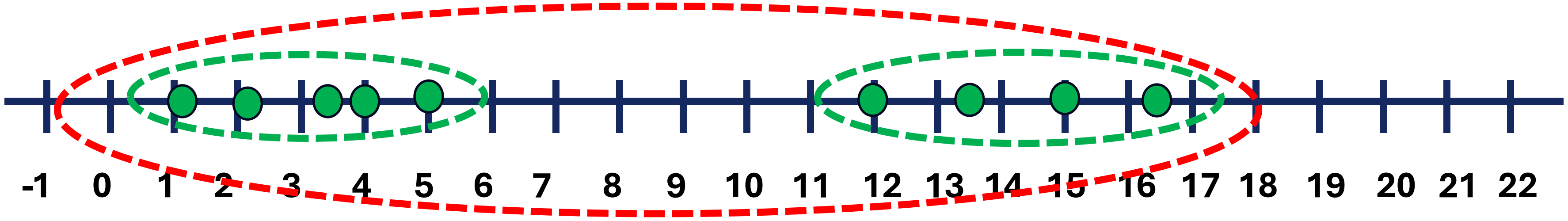
$$n_{ij} = |C_i \cap T_j|$$

Exercise

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

- 1-dimensional 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



$$\mathcal{T} = \{T_1, T_2\} = \{\underbrace{\{x_1, x_2, x_5, x_6, x_9\}}_{T_1 = 5}, \underbrace{\{x_3, x_4, x_7, x_8\}}_{T_2 = 4}\}$$

Contingency table		
	T_1	T_2
C_1	5	4

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

$$recall_i = \frac{n_{ij}}{|T_{j_i}|}$$

$$j_i = argmax_{j=1}^{k}\{n_{ij}\}$$

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

What is the precision of C_1?

Exercise

$$\mathcal{C} = \{C_1, C_2, C_3, C_4, C_5, C_6, C_7, C_8, C_9\}$$
$$= \{\{x_1\}, \{x_2\}, \{x_3\}, \{x_4\}, \{x_5\}, \{x_6\}, \{x_7\}, \{x_8\}, \{x_9\}, \}$$

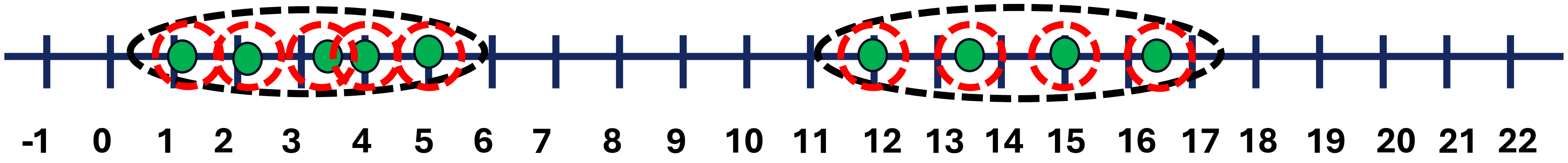
- 1-dimensional example

	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

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

Contingency table

	T_1	T_2
C_1	1	0
C_2	1	0
C_3	0	1
C_4	0	1
C_5	1	0
C_6	1	0
C_7	0	1
C_8	0	1
C_9	1	0



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

$$recall_i = \frac{n_{ij}}{|T_{j_i}|} \quad j_i = argmax_{j=1}^{k}\{n_{ij}\} \quad prec_i = \frac{1}{n_i}\max_{j=1}^{k}\{n_{ij}\}$$

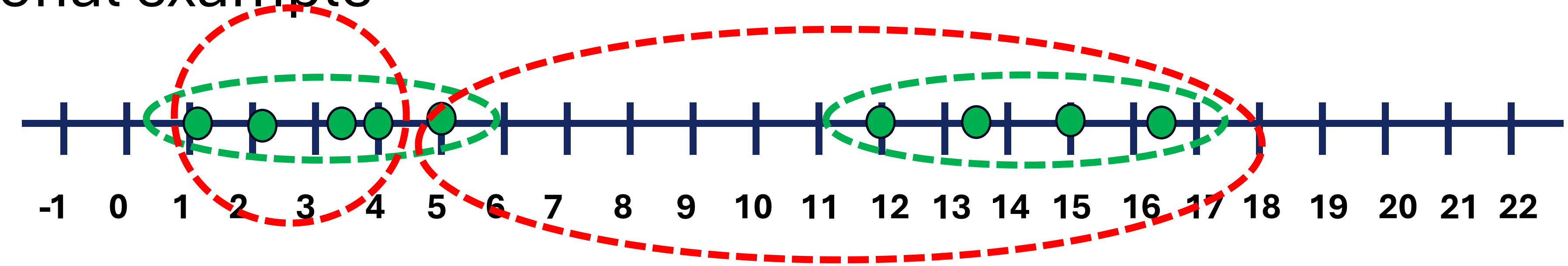
What is the recall of C_1?

F-score

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

- 1-dimensional 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



$|T_1| = 5$ $\quad |T_2| = 4$

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

Contingency table		
	T_1	T_2
C_1	4	0
C_2	1	4

$$recall_i = \frac{n_{ij}}{|T_{j_i}|} \qquad j_i = argmax_{j=1}^{k}\{n_{ij}\}$$

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

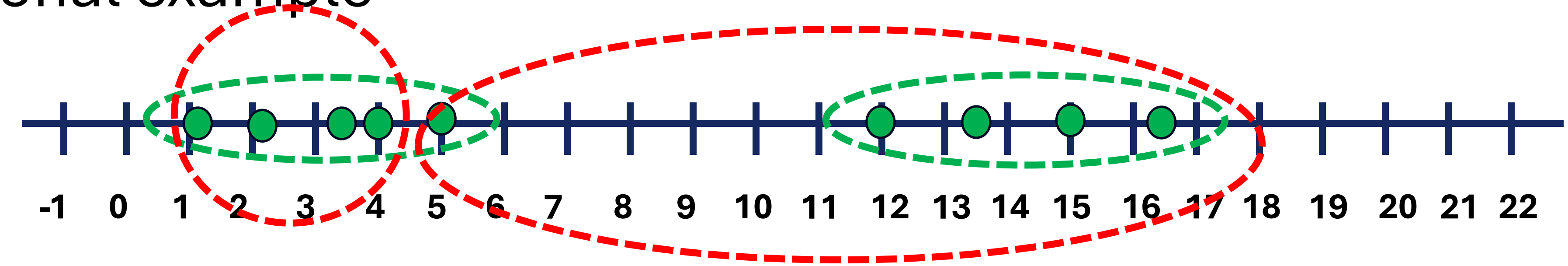
$$n_{ij} = |C_i \cap T_j|$$

F-score

$F - score\ is\ the\ harmonic\ mean\ of\ precision\ and\ recall$

- 1-dimensional example

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



	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

$T_1 = 5$ $\quad T_2 = 4$

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

Contingency table		
	T_1	T_2
C_1	4	0
C_2	1	4

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

$$recall_i = \frac{n_{ij}}{|T_{j_i}|} \qquad j_i = argmax_{j=1}^{k}\{n_{ij}\}$$

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

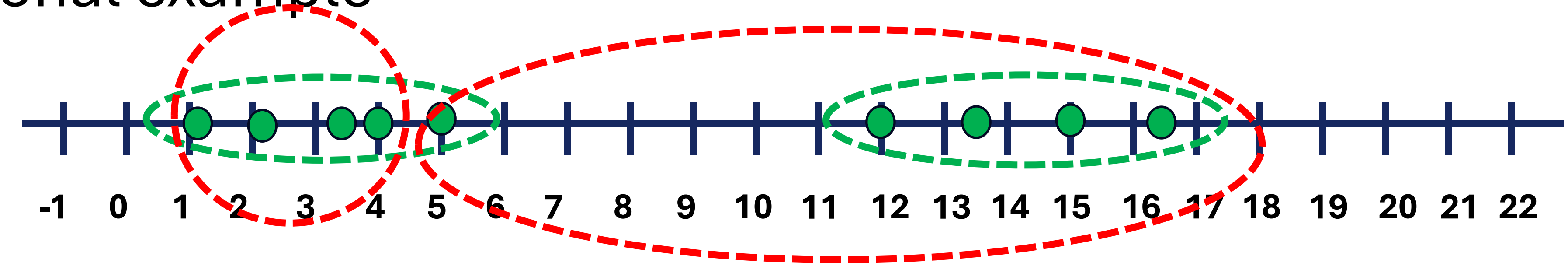
$$F_i = \frac{1}{\frac{1}{2}\left(\frac{1}{prec_i} + \frac{1}{recall_i}\right)}$$

$$F_i = \frac{2(prec_i)(recall_i)}{prec_i + recall_i}$$

F-score

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

- 1-dimensional example



$$prec_1 = \frac{4}{4} = 1$$

$$recall_1 = \frac{4}{5} = 0.8$$

$$F_1 = \frac{2(1)(0.8)}{1 + 0.8} = 0.89$$

$T_1 = 5$ $T_2 = 4$

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

Contingency table

	T_1	T_2
C_1	4	0
C_2	1	4

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

$$\boxed{recall_i = \frac{n_{ij}}{|T_{j_i}|}}$$

$$\boxed{j_i = argmax_{j=1}^{k}\{n_{ij}\}}$$

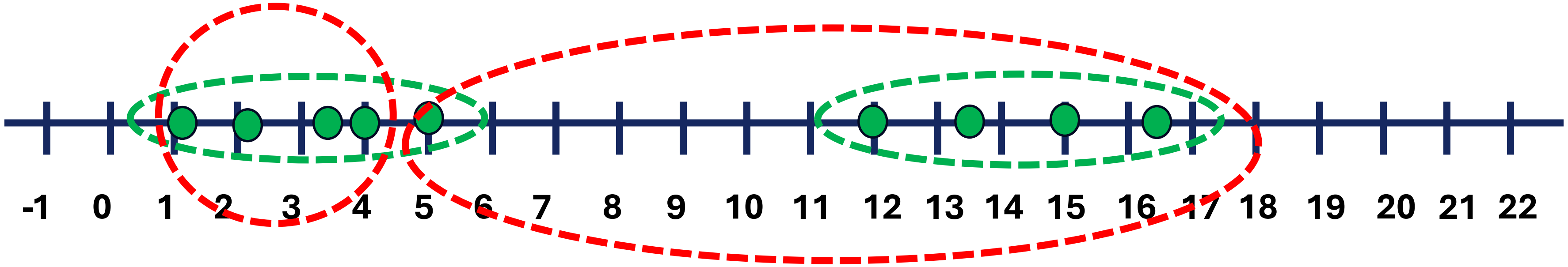
$$\boxed{prec_i = \frac{1}{n}\max_{j=1}^{k}\{n_{ij}\}}$$

$$\boxed{F_i = \frac{2(prec_i)(recall_i)}{prec_i + recall_i}}$$

F-score

- 1-dimensional example

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



$$prec_2 = \frac{4}{5} = 0.8$$

$$recall_2 = \frac{4}{4} = 1$$

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

$$F_2 = \frac{2(0.8)(1)}{0.8 + 1} = 0.89$$

Contingency table

	T_1	T_2
C_1	4	0
C_2	1	4

$$recall_i = \frac{n_{ij}}{|T_{j_i}|} \qquad j_i = argmax_{j=1}^{k}\{n_{ij}\}$$

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

$$F_i = \frac{1}{\frac{1}{2}\left(\frac{1}{prec_i} + \frac{1}{recall_i}\right)}$$

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

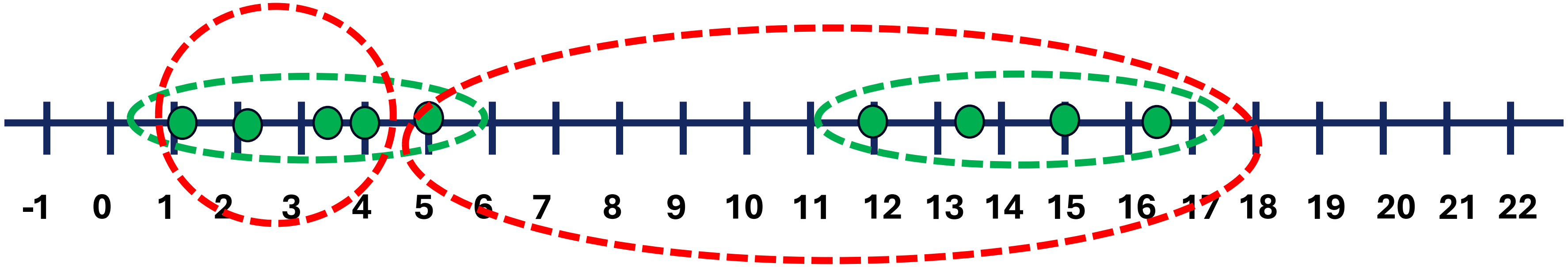
$$F_i = \frac{2(prec_i)(recall_i)}{prec_i + recall_i}$$

F-score

- 1-dimensional example

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

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



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

Contingency table		
	T_1	T_2
C_1	4	0
C_2	1	4

$$recall_i = \frac{n_{ij}}{|T_{j_i}|}$$

$$j_i = argmax_{j=1}^{k}\{n_{ij}\}$$

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

$$F_i = \frac{1}{\frac{1}{2}\left(\frac{1}{prec_i} + \frac{1}{recall_i}\right)}$$

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

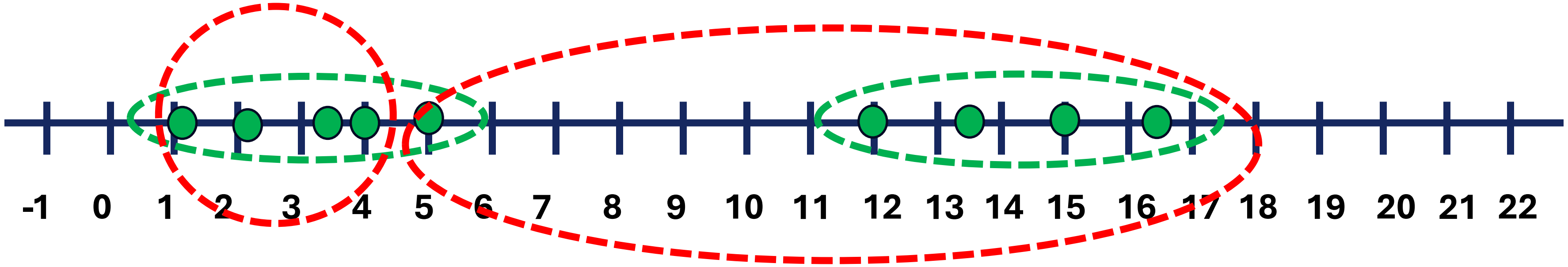
$$F_i = \frac{2(prec_i)(recall_i)}{prec_i + recall_i}$$

F-score

- 1-dimensional example

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

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



Contingency table		
	T_1	T_2
C_1	4	0
C_2	1	4

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

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

$$recall_i = \frac{n_{ij}}{|T_{j_i}|} \qquad j_i = argmax_{j=1}^{k}\{n_{ij}\}$$

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

$$F_i = \frac{1}{\frac{1}{2}\left(\frac{1}{prec_i} + \frac{1}{recall_i}\right)}$$

$$F_i = \frac{2(prec_i)(recall_i)}{prec_i + recall_i}$$

$$F_1 = 0.89$$

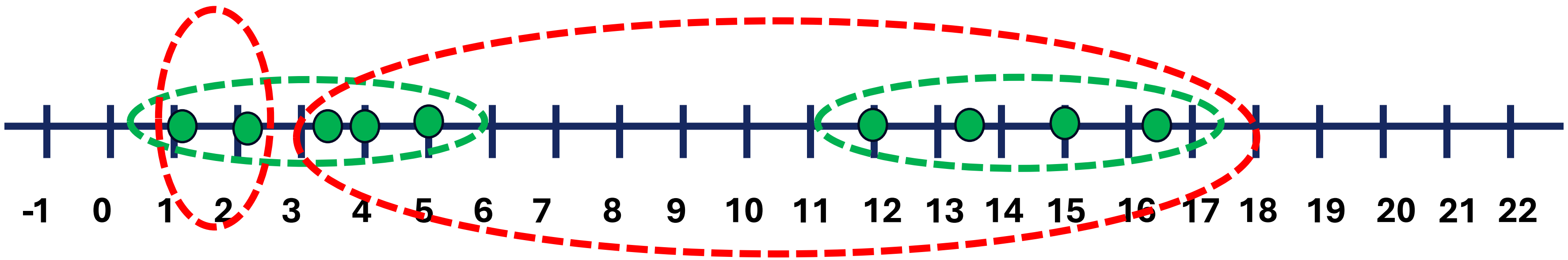
$$F_2 = 0.89$$

$$F = \frac{1}{2}(F_1 + F_2) = 0.89$$

F-score - exercise

- 1-dimensional example

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



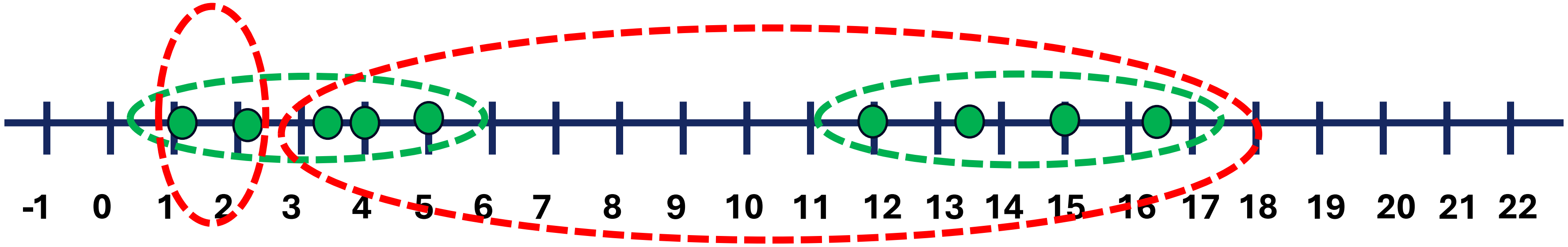
$$\mathcal{T} = \{T_1, T_2\} = \{\{x_1, x_2, x_5, x_6, x_9\}, \{x_3, x_4, x_7, x_8\}\}$$

F-score

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

- 1-dimensional example

Contingency table		
	T_1	T_2
C_1	2	0
C_2	3	4



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

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

$$F_1 = 0.57$$

$$F_2 = 0.73$$

$$F = \frac{1}{2}(F_1 + F_2) = 0.65$$

$$recall_1 = \frac{2}{5}$$

$$prec_1 = 1$$

$$F_1 = \frac{2(1)(0.4)}{1 + 0.4} = 0.57$$

$$recall_2 = \frac{4}{4}$$

$$prec_2 = \frac{4}{7}$$

$$F_2 = \frac{2(0.57)(1)}{0.57 + 1} = 0.73$$