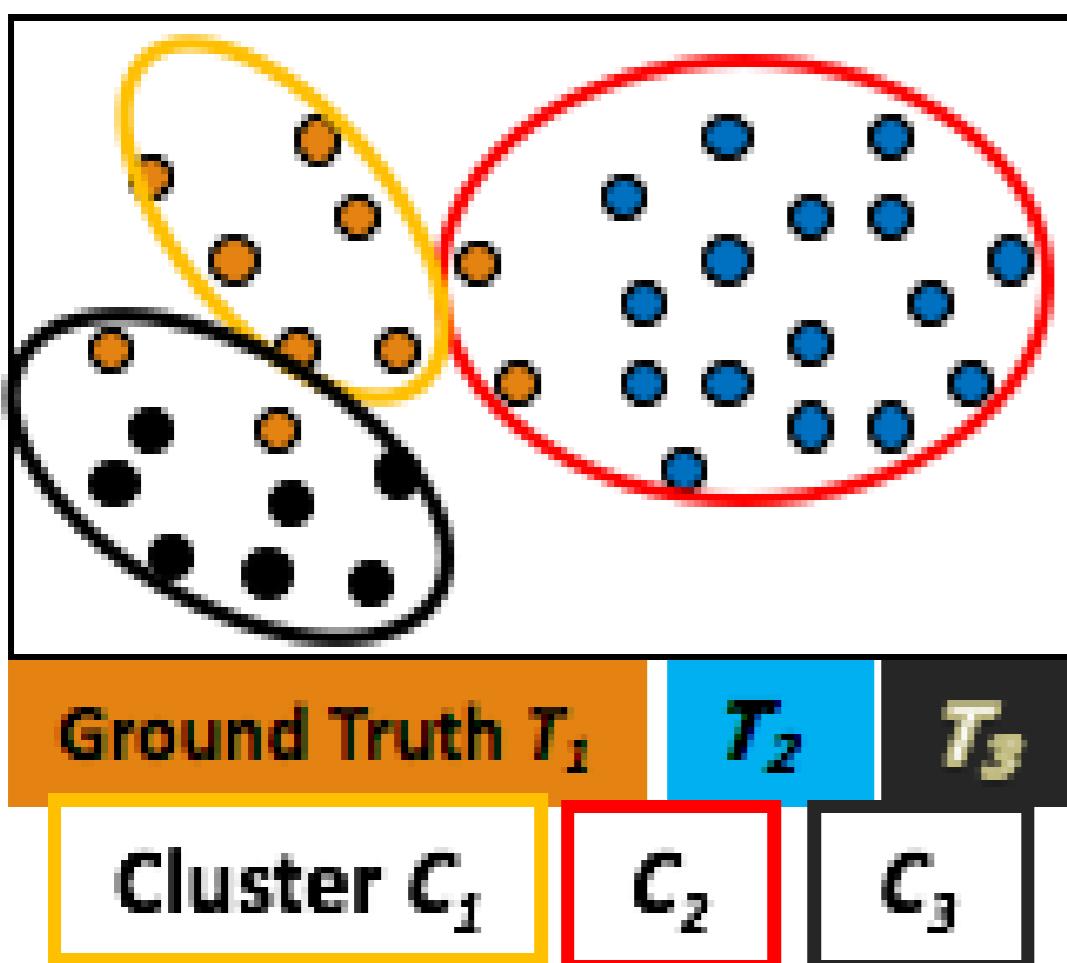


CS4641B Machine Learning

# Focus video: Clustering evaluation – entropy based measures

Rodrigo Borela ▶ [rborelav@gatech.edu](mailto:rborelav@gatech.edu)

# Entropy-based measures: example



C = cluster

T = partition

	$k = 1$	$k = 2$	$k = 3$	$n_r$
$r = 1$	$n_{11} = 6$	$n_{12} = 0$	$n_{13} = 0$	6
$r = 2$	$n_{21} = 2$	$n_{22} = 16$	$n_{23} = 0$	18
$r = 3$	$n_{31} = 2$	$n_{32} = 0$	$n_{33} = 7$	9
$m_k$	10	16	7	33

$n_r$     $k$

Cluster   Ground-truth  
(C)        (T)

# Joint probability distribution

$$p(C = r, T = k)$$

T = partition

C = cluster

	$k = 1$	$k = 2$	$k = 3$	$p_{C_r}$
$r = 1$	$\frac{6}{33}$	$\frac{0}{33}$	0	$\frac{6}{33}$
$r = 2$	$\frac{2}{33}$	$\frac{16}{33}$	0	$\frac{18}{33}$
$r = 3$	$\frac{2}{33}$	$\frac{0}{33}$	$\frac{7}{33}$	$\frac{9}{33}$
$p_{T_k}$	$\frac{10}{33}$	$\frac{16}{33}$	$\frac{7}{33}$	1

Diagram illustrating the joint probability distribution:

- An orange arrow points from the cell  $p(k=2|r=3)$  to the label  $p(k=2|r=3)$ .
- A green arrow points from the cell  $p(r=2)$  to the label  $p(r=2)$ .

# Joint probability distribution

$$p(C = r, T = k)$$

T = partition

		$k = 1$	$k = 2$	$k = 3$	$p_{C_r}$
		0.182	0.0	0.0	0.182
<u>C = cluster</u>	$r = 1$	0.182	0.0	0.0	0.182
	$r = 2$	0.060	0.485	0.0	0.545
	$r = 3$	0.060	0.0	0.212	0.272
$p_{T_k}$		0.303	0.485	0.212	1

# Entropy

- For clusters (result from clustering)

$$H(\mathcal{C}) = - \sum_{r=1}^R p_{C_r} \log_2 p_{C_r} = p_{C_1} \log_2 p_{C_1} + p_{C_2} \log_2 p_{C_2} + p_{C_3} \log_2 p_{C_3}$$

$$H(\mathcal{C}) = -0.182 \log_2 0.182 - 0.545 \log_2 0.545 - 0.272 \log_2 0.272$$

$$H(\mathcal{C}) = 1.435 \text{ bits}$$

- For partitions (ground-truth)

$$H(\mathcal{T}) = - \sum_{r=1}^R p_{T_k} \log_2 p_{T_k} = p_{T_1} \log_2 p_{T_1} + p_{T_2} \log_2 p_{T_2} + p_{T_3} \log_2 p_{T_3}$$

$$H(\mathcal{T}) = -0.303 \log_2 0.303 - 0.485 \log_2 0.485 - 0.212 \log_2 0.212$$

$$H(\mathcal{T}) = 1.503 \text{ bits}$$

# Conditional probability distribution

$$p(T = k | C = r)$$

T = partition

C = cluster

	$k = 1$	$k = 2$	$k = 3$
$r = 1$	$\frac{6}{6}$	$\frac{0}{6}$	$\frac{0}{6}$
$r = 2$	$\frac{2}{18}$	$\frac{16}{18}$	$\frac{0}{18}$
$r = 3$	$\frac{2}{9}$	$\frac{0}{9}$	$\frac{7}{9}$

Arrows point from the highlighted cells to the corresponding conditional probability expressions:

- An orange arrow points from the cell  $\frac{0}{9}$  to the expression  $p(k = 2 | r = 3)$ .
- A green arrow points from the cell  $\frac{0}{18}$  to the expression  $p(k = 3 | r = 2)$ .

# Conditional probability distribution

$$p(T = k | C = r)$$

T = partition

		$k = 1$	$k = 2$	$k = 3$
		1.0	0	0
<u>C = cluster</u>	$r = 1$	1.0	0	0
	$r = 2$	0.111	0.889	0
	$r = 3$	0.222	0	0.778

# Conditional entropy

- For each cluster:

$$H(\mathcal{T}|C_1) = -1.0 \log_2 1.0 - 0.0 \log_2 0.0 - 0.0 \log_2 0.0 = 0.0 \text{ bits}$$

$$H(\mathcal{T}|C_2) = -0.111 \log_2 0.111 - 0.889 \log_2 0.889 - 0.0 \log_2 0.0 = 0.503 \text{ bits}$$

$$H(\mathcal{T}|C_3) = -0.222 \log_2 0.222 - 0.0 \log_2 0.0 - 0.778 \log_2 0.778 = 0.764 \text{ bits}$$

- Conditional entropy

$$H(\mathcal{T}|\mathcal{C}) = \sum_R P(C_r) \times H(\mathcal{T}|C_r) = p_{C_1} \times H(\mathcal{T}|C_1) + p_{C_2} \times H(\mathcal{T}|C_2) + p_{C_3} \times H(\mathcal{T}|C_3)$$

$$H(\mathcal{T}|\mathcal{C}) = 0.182 \times 0.0 + 0.545 \times 0.503 + 0.272 \times 0.764 = 0.482 \text{ bits}$$

# Normalized mutual information

- Mutual information

$$I(\mathcal{C}, \mathcal{T}) = H(\mathcal{T}) - H(\mathcal{T}|\mathcal{C}) = 1.503 - 0.482 = 1.021 \text{ bits}$$

- Normalized mutual information

$$NMI(\mathcal{C}, \mathcal{T}) = \frac{I(\mathcal{C}, \mathcal{T})}{\sqrt{H(\mathcal{C}) \times H(\mathcal{T})}} = \frac{1.021}{\sqrt{1.503 \times 1.435}} = 0.695 \text{ bits}$$

