

$$w_j'' = \frac{(w'_{j+1} - w'_{j-1})}{2} = \left(\frac{(w_{j+2} - w_j)}{2} - \frac{(w_j - w_{j-2})}{2} \right)$$

$$= \frac{w_{j+2} - 2w_j + w_{j-2}}{4}$$

→ Unsupervised process since labels were only used to evaluate in the end.

→ We could in the end try & figure out our Purity by comparing results with others

Purity

let N_{ij} be no. of instances (Supervised) label j within cluster C_i

$$\text{Purity}(C_i) = \frac{1}{|C_i|} \max_j N_{ij}$$

$$\text{avg Clustering purity} = \frac{1}{N} \sum_{i=1}^n |C_i| \text{Purity}(C_i)$$