Vereinfachung von Ward's Methode:

Abstand zwischen zwei Clustern A,B (zu minimieren):

$$W(A,B)=\sum_{i\in A\cup B} d(p_i, m_{A\cup B})^2 - \sum_{i\in A} d(p_i, m_A)^2 - \sum_{i\in B} d(p_i, m_B)^2$$

$p_i := (x_i, y_i)$

$m_C := (m_C^x, m_C^y)$ $\qquad n_C := |C|$ $\qquad A\cup B$ = Fusion von A, B

$$d(p_i,p_j)^2 = \left(\sqrt{(x_i-x_j)^2+(y_i-y_j)^2}\right)^2 = (x_i-x_j)^2+(y_i-y_j)^2 = \underbrace{(p_i-p_j)^2}_{\text{Vektordiff.}}$$

---

im folg. nur x-Anteile gerechnet, m_C kurz für m_C^x:

$$W(A,B) = \sum_{A\cup B} (x_i - m_{A\cup B})^2 - \sum_A (x_i - m_A)^2 - \sum_B (x_i - m_B)^2$$

$\leftarrow (v-w)^2 = v^2 - 2vw + w^2$ und einzeln addiert

$$= \cancel{\sum_{A\cup B} x_i^2 - \sum_A x_i^2 - \sum_B x_i^2} \; 0$$

$$- \sum_{A\cup B} 2x_i m_{A\cup B} + \sum_A 2x_i m_A + \sum_B 2x_i m_B$$

$$+ \sum_{A\cup B} m_{A\cup B}^2 - \sum_A m_A^2 - \sum_B m_B^2$$

Forts. unten

$$m_A = \frac{\sum_A x_i}{n_A} \qquad m_B \text{ analog} \qquad m_{A\cup B} = \frac{\sum_{A\cup B} x_i}{n_A+n_B} = \frac{\sum_A x_i}{n_A}\cdot\frac{n_A}{n_A+n_B} + \frac{\sum_B x_i}{n_B}\cdot\frac{n_B}{n_A+n_B}$$

$$= m_A \cdot \frac{n_A}{n_A+n_B} + m_B \cdot \frac{n_B}{n_A+n_B} \qquad (1)$$

$$\sum_C 2x_i \cdot m_C = 2\cdot m_C \cdot \sum_C x_i = 2\cdot m_C \cdot (m_C \cdot n_C) = 2\, n_C\, m_C^2 \qquad (2)$$

$$\sum_C m_C^2 = n_C \cdot m_C^2 \qquad (3)$$

Forts. $\Rightarrow$

$$W(A,B) \underset{(2)(3)}{=} -2\cdot(n_A+n_B)\cdot(m_{A\cup B})^2 + 2\cdot n_A m_A^2 + 2\cdot n_B\cdot m_B^2$$
$$+ (n_A+n_B)(m_{A\cup B})^2 - n_A\cdot m_A^2 - n_B\cdot m_B^2$$
$$= -(n_A+n_B)(m_{A\cup B})^2 + n_A\cdot m_A^2 + n_B\, m_B^2$$
$$\underset{(1)}{=} -\cancel{(n_A+n_B)}\cdot\frac{(m_A n_A + m_B n_B)^2}{(n_A+n_B)^{\cancel{2}}} + \frac{(n_A+n_B) n_A m_A^2}{(n_A+n_B)} + \frac{(n_A+n_B)\cdot n_B\cdot m_B^2}{(n_A+n_B)}$$
$$= \frac{-(m_A^2 n_A^2 + 2 m_A n_A m_B n_B + m_B^2 n_B^2) + (n_A+n_B) n_A m_A^2 + (n_A+n_B) n_B m_B^2}{(n_A+n_B)}$$
$$= \frac{m_A^2\cdot(\cancel{-n_A^2} \cancel{+n_A^2} + n_A\cdot n_B) + m_B^2(\cancel{-n_B^2} + n_A n_B \cancel{+n_B^2}) - 2 m_A n_A m_B n_B}{(n_A+n_B)}$$
$$= \frac{n_A n_B (m_A - m_B)^2}{(n_A+n_B)}$$

$\Rightarrow$ Ward's Distanz $W(A,B) = \frac{n_A\cdot n_B}{(n_A+n_B)} \cdot d(m_A, m_B)^2 = \frac{d(m_A,m_B)^2}{\frac{1}{n_A}+\frac{1}{n_B}}$!

[anfangs (für Einpunkte-Cluster): $W(A,B) = \frac{1\cdot 1}{1+1}\cdot d(\text{Punkte})^2 = \frac{1}{2}\cdot d(\text{Punkte})^2$]