11. 6. 2014

K-Means-Clustering

Step 1: Determine the number of clusters k.

Step 2: Random choice of k different vectors $\underline{y}_1, \underline{y}_2, \ldots \underline{y}_k$ and set those equal to the initial cluster centers $\underline{w}_1, \underline{w}_2, \ldots \underline{w}_k$

Step 3: Choose a vector $\underline{y}(k)$ and calculate distance between $\underline{y}(k)$ and vectors $\underline{w}_1, \dots, \underline{w}_K$. Assign $\underline{y}(k)$ to cluster m^* with minimum distance and store $\underline{y}(k)$ in a list L_{m^*} for this cluster

Step 4: Set $k = k + 1$ and go back to step 3 until all vectors have been processed. The K lists $L_1 - L_K$ remain.

Step 5: Calculate new reference vectors $\underline{w}_1 - \underline{w}_k$ for each cluster by calculating the mean vector from each list.

Step 6: Set $k = 0$ and go back to step 3 to start a new iteration. Altogether, 20 – 50 iterations are carried out.

Labelling of reference vectors

$\underline{v}_1 \dots \underline{v}_k$ according to the available classes

- Collect a class-labeled database
- perform another k-means iteration with those vectors and assign corresponding lists $L_1 \dots L_k$

Compute matrix P :

$$P = \begin{bmatrix} P(\underline{\psi}_1|\Omega_1) & \cdots & P(\underline{\psi}_k|\Omega_1)\\ P(\underline{\psi}_1|\Omega_2) & & P(\underline{\psi}_k|\Omega_2)\\ \vdots & & \vdots \\ P(\underline{\psi}_1|\Omega_M) & & P(\underline{\psi}_k|\Omega_M)\end{bmatrix}$$

$$\underbrace{\qquad\qquad\qquad}_{\max\ \Omega_1} \quad \cdots \quad \underbrace{\qquad\qquad\qquad}_{\max\ \Omega_i}$$

Advantages:

- not chosen arbitrarily but by self-organizing principles
- they serve well for application of the KNN classification procedure
- they could modified for other procedures, e.g. NN classification

3.2.3.2 Determination of parameters for decision functions

Recall procedure for decision functions.

$$\underbrace{\begin{bmatrix} \underline{w}_1^T \\ \underline{w}_2^T \\ \vdots \\ \underline{w}_M^T \end{bmatrix}}_{W} \cdot \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}}_{\underline{y}} = \underbrace{\begin{bmatrix} d_1 \\ \vdots \\ d_M \end{bmatrix}}_{\underline{d}} \to \max$$

$$W \cdot \underline{y} = \underline{d}$$

parameters   unknown vector   decisions

Data Collection:

$$y(1),\ y(2)\ \cdots\ y(k)\ \cdots\ y(K)$$

$$\Omega_2 \qquad \Omega_{10} \qquad \Omega_5 \qquad \Omega_1$$

$$\begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \qquad \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \begin{matrix} \\ \\ \\ \#5 \\ \\ \\ \end{matrix} \quad \underline{\hat{d}}(k)$$

Classification rule for training vector

$\underline{y}(k)$:

$$\underline{d}(k) = W \cdot \underline{y}(k)$$

Classification error for k-th vector

$\underline{y}(k)$:

$$E(k) = \frac{1}{2} \sum_{i=1}^{m} \left[\vec{d}_i(w) - d_i(k) \right]^2$$

For all presentations (entire data collection) :

$$E = \sum_{k=1}^{K} E(k) = \frac{1}{2} \sum_{k=1}^{K} \sum_{i=1}^{M} \left[d_i(k) - \hat{d}_i(k) \right]^2$$

with $$\hat{d}_i(k) = \sum_{j=1}^{N} w_{ij}\, y_j(k)$$

Derivative of error E with respect to a weight w_{ij}:

$$\frac{\partial \mathcal{E}}{\partial w_{ij}} = -\sum_{k=1}^{K} \left[\hat{d}_i(k) - d_i(k) \right] \underbrace{\frac{\partial \hat{d}_i(k)}{\partial w_{ij}}}_{= y_j(k)} = 0$$

$$\rightarrow \sum_{k=1}^{K} \left[\hat{d}_i(k) - d_i(k) \right] \cdot y_j(k) = 0$$

$$\sum_{k=1}^{K} \hat{d}_i(k) \cdot y_j(k) = \sum_{k=1}^{K} d_i(k) \cdot y_j(k)$$

inserting $d_j(k)$ results into

$$\sum_{k=1}^{K} \hat{d}_i(k) \cdot y_j(k) = \sum_{k=1}^{K} \left[\sum_{n=1}^{N} w_{in} y_n(k) \right] \cdot y_j(k)$$

$$= \sum_{n=1}^{N} w_{in} \cdot \sum_{k=1}^{K} y_n(k) \cdot y_j(k)$$

$\Rightarrow$ On left side: $x_{ij} = E\{d_i \cdot y_j\}$

Right side: $c_{nj} = E\{y_n \cdot y_j\}$

$$\Rightarrow \cancel{K} \cdot x_{ij} = \cancel{K} \cdot \sum_{n=1}^{N} w_{in} \cdot c_{nj}$$

Rewrite this equation in matrix notation:

$$X = W \cdot C$$

with $$\hat{X} = E\{\hat{\underline{d}} \cdot \underline{y}^T\} = \frac{1}{K}\sum_{k=1}^{K} \hat{\underline{d}}(k) \cdot \underline{y}^T(k)$$

$$C = E\{\underline{y} \cdot \underline{y}^T\}$$

$$= \frac{1}{K}\sum_{k=1}^{K} \underline{y}(k) \cdot \underline{y}_k^T$$

$$\Rightarrow h = X^T \cdot C^{-1}$$