

INIT: CLUSTER CENTERS μ_k - RANDOM SUBSET OF DATA POINTS

SLOD (E): $\| \text{DIST} \|^2$ BETWEEN EVERY DATA POINT & EVERY CENTER

BATCH ↗

ON-LINE ↘

STOCHASTIC APPROX. ALG.

$\forall x_n$

THE NEAREST PROTOTYPE UPDATE.

$$\mu_k^{\text{new}} = \mu_k^{\text{old}} + \eta_n (x_n - \mu_k^{\text{old}})$$

GENERALIZATION: EUCLIDEAN DISTANCE \rightarrow OTHER MEASURES

$V(x, x')$ DISSIMILAR MEASURE

$$\tilde{J} = \sum \sum w_{nk} V(x_n, \mu_k)$$

(K-MEANS ALGORITHM)

GENERALIZATION:

$x_n \rightarrow$ HARD ASSIGNMENT (ONLY 1 CLASS) 0/1

\rightarrow SOFT - i.e. - (\forall CLASS + WEIGHT)

(PROBABILISTIC)

$$w_i \geq 0, \sum_i w_i = 1$$

K-MEANS AS DATA REDUCTION:

(DATA COMPRESSION)

LOSSLESS

LOSSY

$x_n \rightarrow$ ONLY K & $\{\mu_k\}$

NEW x' \rightarrow CLASS μ_k

VECTOR-QUANTIZATION

$\{\mu_k\}$ CODE-BOOK VECTORS