Clustering and competitive learning

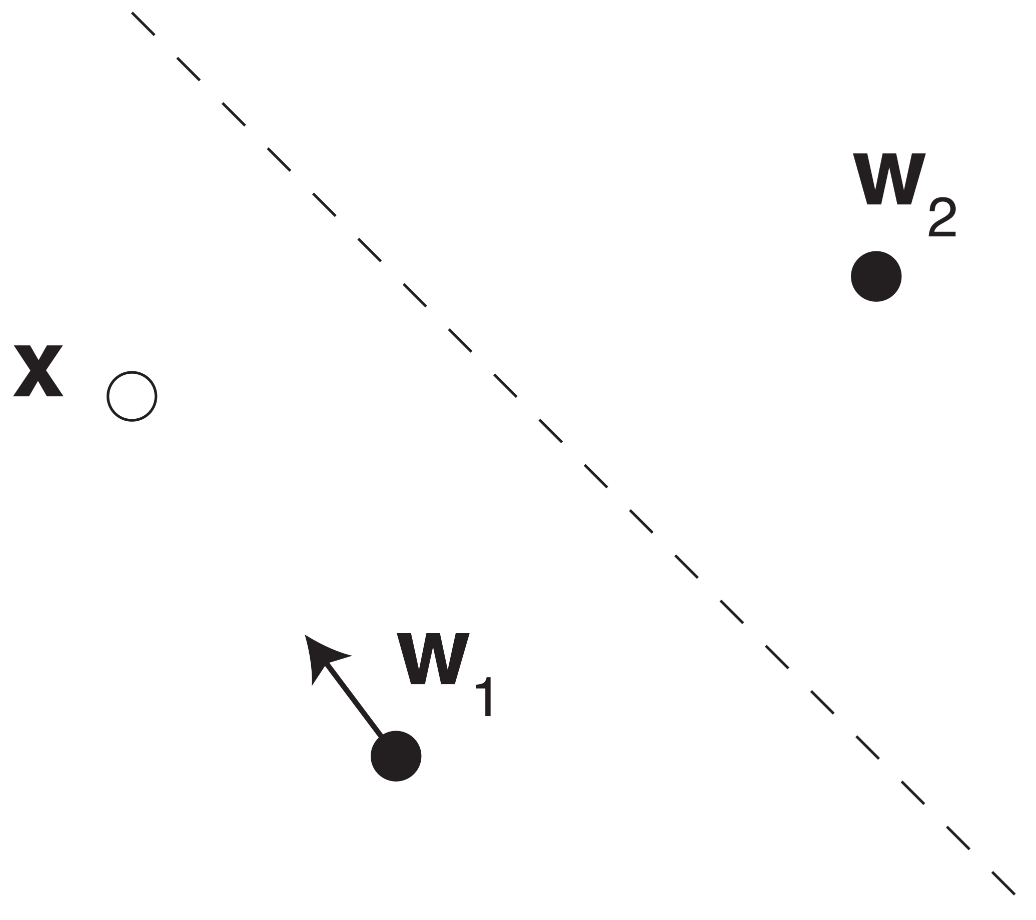
Sebastian Seung

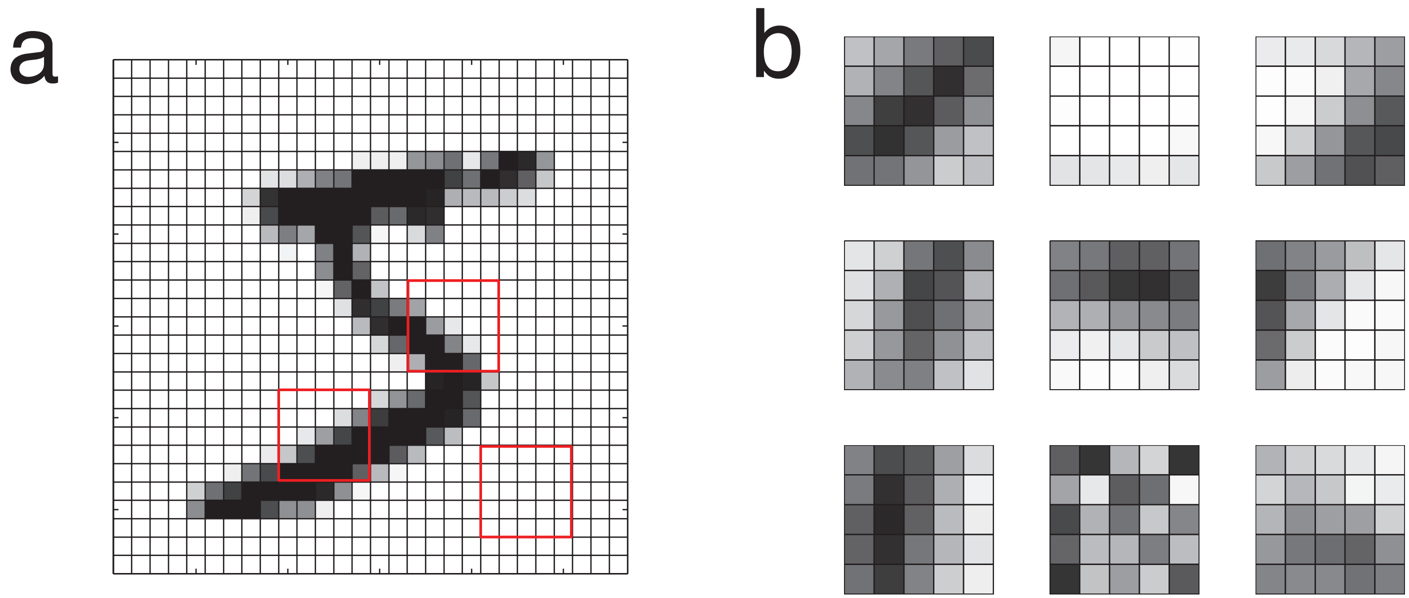
Competitive learning

$$y_a = \begin{cases} 1, & a = \mathrm{argmin}_b \left|\mathbf{x} - \mathbf{w}_b\right| \\ 0, & \text{otherwise} \end{cases}$$

$$\Delta \mathbf{w}_a = \eta y_a (\mathbf{x} - \mathbf{w}_a)$$

Move the closest weight vector to the input vector




a
b

Average velocity approximation

$$\Delta \mathbf{w}_a \approx \eta(\langle y_a \mathbf{x} \rangle - \langle y_a \rangle \mathbf{w}_a)$$

- steady state

$$\mathbf{w}_a \approx \frac{\langle y_a \mathbf{x} \rangle}{\langle y_a \rangle}$$

Clustering

- Divide data vectors into clusters
- Summarize each cluster by a single prototype.

A single prototype

- Summarize all data with the sample mean.

$$\mu = \frac{1}{m}\sum_{a=1}^{m} x_a$$

Multiple prototypes

- Each prototype is the mean of a subset of the data.
- Divide data into *k* clusters.
  - One prototype for each cluster.

Vector quantization

- Many telecom applications
- Codebook of prototypes
- Send index of prototype rather than whole vector
- Lossy encoding

Assignment matrix

$$\begin{array}{cc} & \text{cluster } \alpha \\ \begin{array}{c}\text{data}\\ \text{vector}\\ a\end{array} & \begin{pmatrix}0&1&0\\1&0&0\\0&1&0\\0&0&1\\0&1&0\end{pmatrix}\end{array}$$

$$Y_{a\alpha} = \begin{cases} 1, & x_a \in \text{ cluster } \alpha \\ 0, & \text{otherwise} \end{cases}$$

- Data structure for cluster memberships.

k-means algorithm

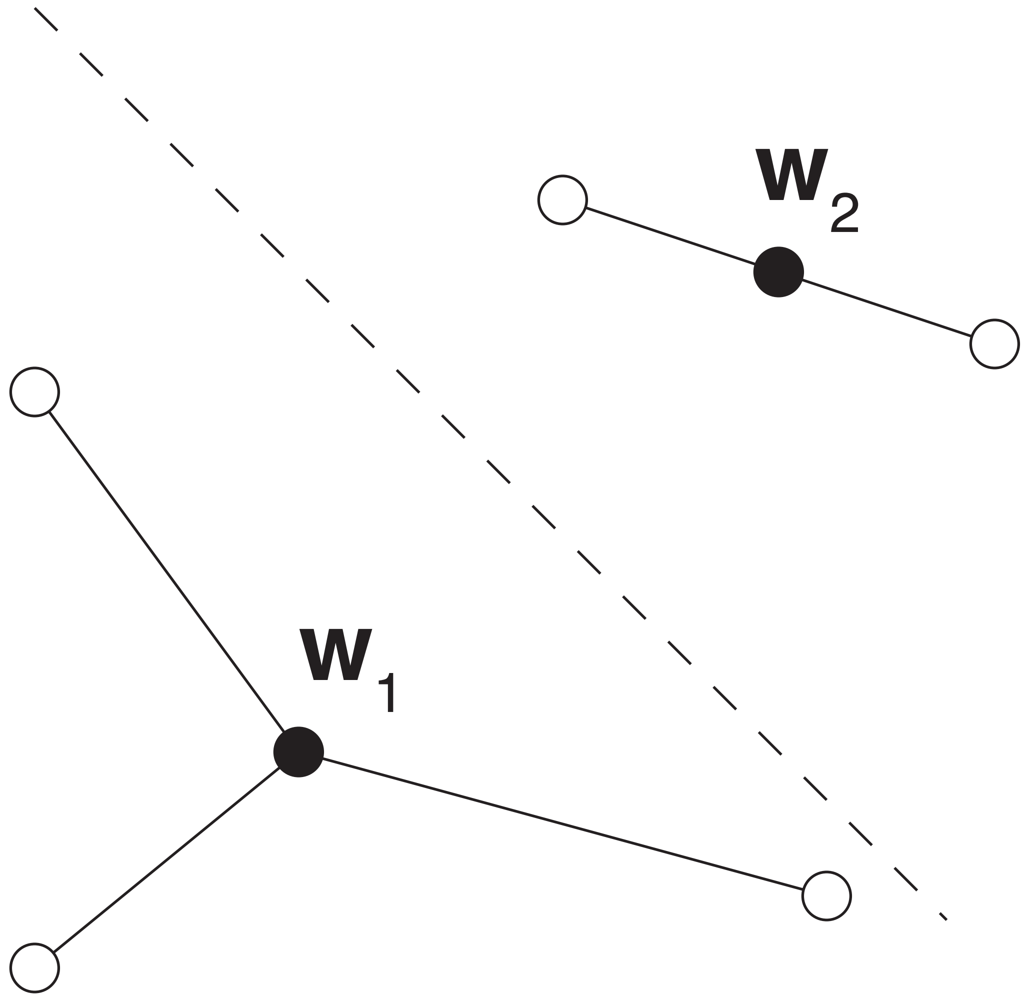
- Alternate between computing means and computing assignments.

$$\mathbf{w}_\alpha = \frac{\sum_{a=1}^{m} \mathbf{x}_a Y_{a\alpha}}{\sum_{b=1}^{m} Y_{b\alpha}}$$

$$Y_{a\alpha} = 1 \text{ for}$$
$$a = \arg\min_\beta \left| \mathbf{x}_a - \mathbf{w}_\beta \right|$$

Objective function for *k*-means

$$E(Y,w) = \frac{1}{2}\sum_{a=1}^{m}\sum_{\alpha=1}^{k} Y_{a\alpha}\left|\mathbf{x}_a - \mathbf{w}_\alpha\right|^2$$



w2
w1

Avoiding local minima

- Good initialization
- Splitting
- Annealing

Model selection

- How to choose the number of clusters?
- Tradeoff between model complexity and objective function.