Clustering and competitive learning

Sebastian Seung
Competitive learning

\[ y_a = \begin{cases} 
1, & a = \arg\min_b |x - w_b| \\
0, & \text{otherwise} 
\end{cases} \]

\[ \Delta w_a = \eta y_a (x - w_a) \]
Move the closest weight vector to the input vector
Average velocity approximation

\[ \Delta w_a \approx \eta \left( \langle y_a x \rangle - \langle y_a \rangle w_a \right) \]

- steady state

\[ w_a \approx \frac{\langle y_a x \rangle}{\langle y_a \rangle} \]
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

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

- Data structure for cluster memberships.
$k$-means algorithm

• Alternate between computing means and computing assignments.

\[
W_\alpha = \frac{\sum_{a=1}^{m} x_a Y_{a\alpha}}{\sum_{b=1}^{m} Y_{b\alpha}}
\]

\[
Y_{a\alpha} = 1 \text{ for } a = \arg \min_{\beta} |x_a - w_\beta|
\]
Objective function for $k$-means

$$E(Y,w) = \frac{1}{2} \sum_{a=1}^{m} \sum_{\alpha=1}^{k} Y_{a\alpha} \left| x_{a} - w_{\alpha} \right|^2$$
Avoiding local minima

• Good initialization
• Splitting
• Annealing
Model selection

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