Machine Learning I: Fractal 2

Rajendra Nagar

Assistant Professor
Department of Electircal Engineering
Indian Institute of Technology Jodhpur

These slides are prepared from the following books:

Shalev-Shwartz, Shai, and Shai Ben-David. Understanding machine learning: From theory to algorithms. Cambridge university press, 2014.

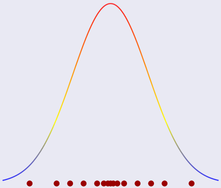
Bishop, C. M. (2006). Pattern recognition and machine learning. springer.

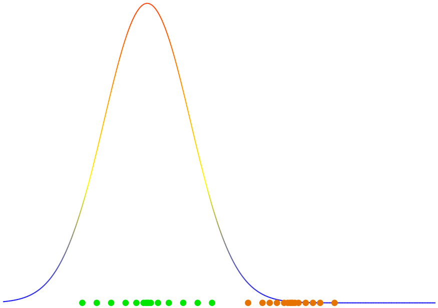
Maximum Likelihood Estimator

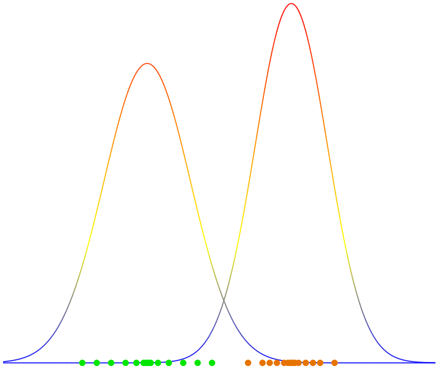
- Given an IID training set $\mathcal{S} = (x_1, \ldots, x_m)$ sampled according to a density distribution $\mathcal{P}_\theta$, we define the likelihood of $\mathcal{S}$ given θ as

$$L(\mathcal{S};\theta) = \log\left(\prod_{i=1}^{m} \mathcal{P}_\theta(x_i)\right) \text{ and solve } \hat{\theta} \in \arg\max_{\theta} L(\mathcal{S};\theta).$$

- For $\mathcal{P}_\theta(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, we obtain the maximum likelihood estimates for μ and σ as: $\hat{\mu} = \frac{1}{m}\sum_{i=1}^{m} x_i$ and $\hat{\sigma}^2 = \frac{1}{m}\sum_{i=1}^{m}(x_i - \hat{\mu})^2$.







Gaussian Mixture Models

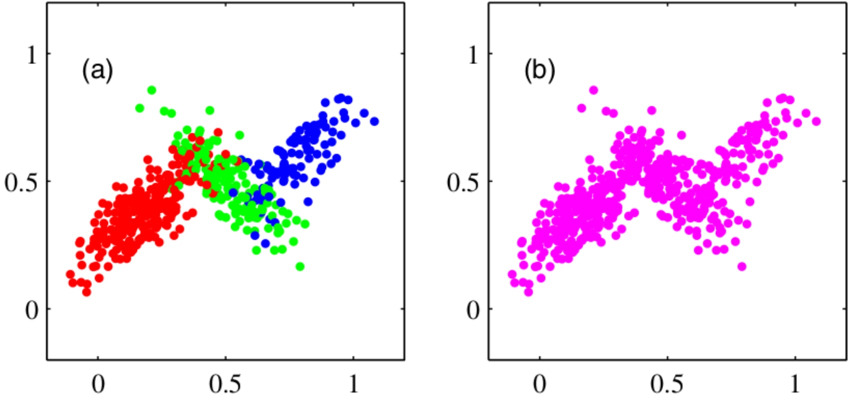


Image Source: Bishop, C. M. (2006). Pattern recognition and machine learning. springer.

Algorithm 1 Expectation Maximization

1: **Input**: $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \ldots, \mathbf{x}_n \end{bmatrix} \in \mathbb{R}^{d\times n}$ and $p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, \pi_k)$.

2: **Initialize:** $\boldsymbol{\mu}_k$, $\boldsymbol{\Sigma}_k$, and π_k.

3: **E** step. Evaluate the responsibilities using the current parameter values.

$$\gamma(z_{nk}) \leftarrow \frac{\pi_k \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum\limits_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}.$$

4: **M** step. Re-estimate the parameters using the current responsibilities.

$$\begin{aligned} \boldsymbol{\mu}_k^{\text{new}} &\leftarrow \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) \mathbf{x}_n \\ \boldsymbol{\Sigma}_k^{\text{new}} &\leftarrow \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) (\mathbf{x}_n - \boldsymbol{\mu}_k^{\text{new}})(\mathbf{x}_n - \boldsymbol{\mu}_k^{\text{new}})^\top \\ \pi_k^{\text{new}} &\leftarrow \frac{N_k}{N}. \end{aligned}$$

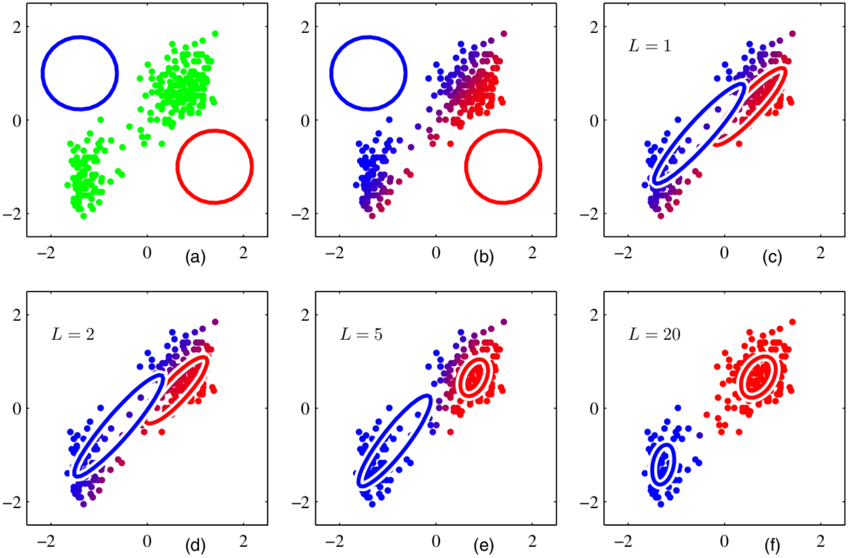


Image Source: Bishop, C. M. (2006). Pattern recognition and machine learning. springer.