Training set \( S = \{x^{(1)}, x^{(2)}, \ldots, x^{(n)}\} \).

Clustering: Group data in the training set into clusters.

Training set \( S = \{x^{(1)}, \ldots, x^{(n)}\} \).

Goal: Discover the distribution of the training data.

Distribution of the training data is most conveniently described using "latent"/hidden variables.

"Mixture of Gaussians": \( x \in \mathbb{R}^d \)

- Pick \( z \in \mathcal{Z}, \ldots, k \) is distributed according to multinomial distribution \( \phi \).

\[ \phi_z = P[z = i] \]

- \( x \sim N(M_z, \Sigma_z) \)

\[ p(x | z) = \frac{1}{(2\pi)^{d/2} \Sigma_z^{1/2}} e^{-\frac{1}{2} (x - M_z)^T \Sigma_z^{-1} (x - M_z)} \]

GDA

Given a training set \( S \)

Goal: Discover the parameters that define mixture of Gaussians distribution from which \( S \) is drawn.

\( (\phi, \frac{1}{k} M_z, \Sigma_z, \frac{1}{k}, z, \phi_z) \)
\[
S = \sum x_{(i)} \cdot x_{(i)}^2 .
\]

The pdf is

\[
\frac{\text{pdf}}{L(S; \phi, M, \Sigma_1, \ldots, M_k, \Sigma_k)} = \prod_{i=1}^{n} p(x_{(i)}; \phi, M, \Sigma)
\]

\[
L(S; \phi, M_j, \Sigma_j) = \sum_{i=1}^{n} \log p(x_{(i)}; \phi, M_j, \Sigma_j)
\]

\[
= \sum_{i=1}^{n} \log \prod_{j=1}^{k} p(x_{(i)}; \phi, M_j, \Sigma_j)
\]

\[
= \sum_{i=1}^{n} \log \prod_{j=1}^{k} p(z; \phi, M_j, \Sigma_j)
\]

\[
\rightarrow \text{Multinomial (}\phi) \sim \mathcal{N}(M_j, \Sigma_j)
\]

Suppose \( z_{(i)} \) is known (\( z_{(i)} \) is the distribution from which \( x_{(i)} \) is drawn), then finding the parameters that maximize likelihood is easy.

\[
p(z_{(i)} = j) = \phi_j = \frac{\sum_{i=1}^{n} 1[z_{(i)} = j]}{\sum_{i=1}^{n} 1[z_{(i)} = j]}
\]

\[
M_j = \frac{\sum_{i=1}^{n} 1[z_{(i)} = j] x_{(i)}}{\sum_{i=1}^{n} 1[z_{(i)} = j]}
\]

\[
\Sigma_j = \frac{\sum_{i=1}^{n} 1[z_{(i)} = j] (x - M_j)(x - M_j)^T}{\sum_{i=1}^{n} 1[z_{(i)} = j]}
\]

\[
w_{(i)}_{(j)} = p(z_{(i)} = j | x_{(i)}; \phi \ldots)
\]

\[
= \frac{p(z_{(i)} = j) p(x_{(i)} | z_{(i)} = j; \phi \ldots)}{p(x_{(i)}; \phi \ldots)}
\]

\[
= \phi_j \frac{p(x_{(i)} | z_{(i)} = j; \phi \ldots)}{p(x_{(i)}; \phi \ldots)}
\]

\[
= \phi_j \frac{p_{N(M_j, \Sigma_j)}(x_{(i)})}{\sum_{j=1}^{k} p_{N(M_j, \Sigma_j)}(x_{(i)})}
\]
Expectation - Maximization for Mixture of Gaussians:

Initialize $\phi, \mu_1, \Sigma_1, \ldots, \mu_k, \Sigma_k$

Repeat

\[ E\text{-step}: \]
\[ w^{(i)}_j = f(z = j \mid x^{(i)}; \phi, \mu_1, \Sigma_1, \ldots, \mu_k, \Sigma_k) \]

\[ M\text{-step}: \]
\[ \phi_j = \frac{\sum_{i=1}^{n} w^{(i)}_j}{n} \]
\[ \mu_j = \frac{\sum_{i=1}^{n} w^{(i)}_j x^{(i)}}{\sum_{i=1}^{n} w^{(i)}_j} \]
\[ \Sigma_j = \ldots \]

Until convergence.

\[ \text{EM algorithm finds a local optimum for the parameters.} \]

\[ \text{EM algorithms is a more general approach to maximizing likelihood when the likelihood has the form} \]
\[ \sum \log \pi \]

...