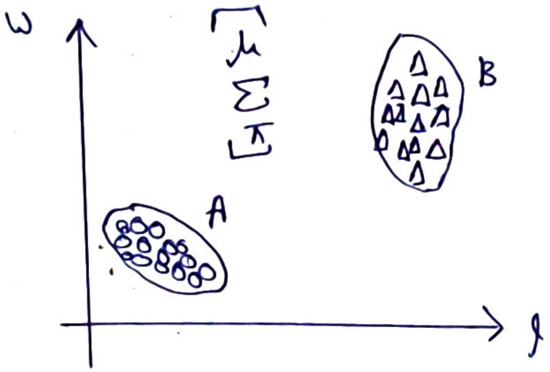
GMM

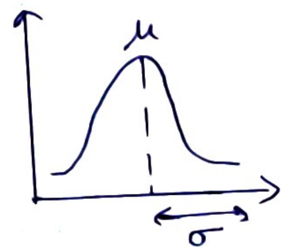


Model assumes classes are distributed in a Gaussian distribution.

1D Gaussian defined by:

→ μ - Mean

→ σ - std dev / how spread distribution is



Multivariate gaussian defined by:

→ μ - Mean

→ Σ - covariance matrix or (size $n \times n$) $n \equiv$ dimension
tells us shape of distribution

→ π - probability of being in either class

Let the distributions be A and B.

$A = N(x \mid \mu_A \, \Sigma_A) \Rightarrow$ If I know something is A, we know it is distributed by μ_A and Σ_A.

$P(A) = \pi_A$

$x = \begin{pmatrix} l_i \\ w_i \end{pmatrix}$

We want to maximize $P(x) \equiv$ probability of seeing x in any of the distributions.

$$P(x) = \pi_A \underbrace{N(x \mid \mu_A \Sigma_A)}_{\text{prob of seeing } x \text{ in } A} + \pi_B N(x \mid \mu_B \Sigma_B)$$

How to pick the best π, μ, Σ?

We want to maximise $P(X \mid \pi, \mu, \Sigma)$

data points $\longrightarrow$ ($N \times$ dim) size

$$P(X \mid \pi, \mu, \Sigma) = \prod_{n=1}^{N} \left[\sum_{k=1}^{K} \pi_k N(x_n \mid \mu_k, \Sigma_k) \right]$$

Multiply all points probabilities

classes

Reason why values update and algo is not cyclic, acts as gradient in the NN

To maximize it we take its derivative wrt ~~each of the~~ π_k, μ_k, Σ_k with which we get new value for those 3.

$\gamma(z_{nk}) = P(z_{nk} = 1 \mid x_n)$ ≡ probability observation is in class k given our observation.

$$z_{nk} = \begin{cases} 1, & x_n \text{ in } k \\ 0, & \text{if not} \end{cases}$$

called responsibility

= prob any obv in k × prob obv given its in k

above thing but for all classes

$$\pi_k \longrightarrow = \frac{P(z_k = 1)\, P(x_n \mid z_{nk} = 1)}{\sum_{j=1}^{K} P(z_j = 1)\, P(x_n \mid z_{nj} = 1)}$$

$P(x_n \mid z_{nk} = 1) \rightarrow N(x_n \mid \mu_k, \Sigma_k)$

⇒ μ_k, Σ_k, π_k depend on $\gamma(z_{nk})$ and vice versa.

<u>Expectation - Maximization</u> (EM) algorithm:

1) Initialize μ_k, π_k and Σ_k.
2) Compute $\gamma(z_{nk})$'s (E)
3) Recompute μ_k, π_k and Σ_k (M), repeat 2.
4) Stop cycle based on criteria and hopefully get best values of μ_k, π_k, Σ_k.

The optimal number of clusters for the process are chosen using the BIC algorithm.

To calculate clusters a threshold is used and the responsibilities are extracted ~~from~~ by using GMM, using which soft clustering is performed.