EM ALGORITHM

- In the real-world applications of machine learning, it is very common that there are many relevant features available for learning but only a small subset of them are observable.
- The ***Expectation-Maximization algorithm*** can be used for the latent variables (variables that are not directly observable and are actually inferred from the values of the other observed variables).
- This algorithm is actually the base for many unsupervised clustering algorithms in the field of machine learning.

EM ALGORITHM

Let us understand the EM algorithm in detail.

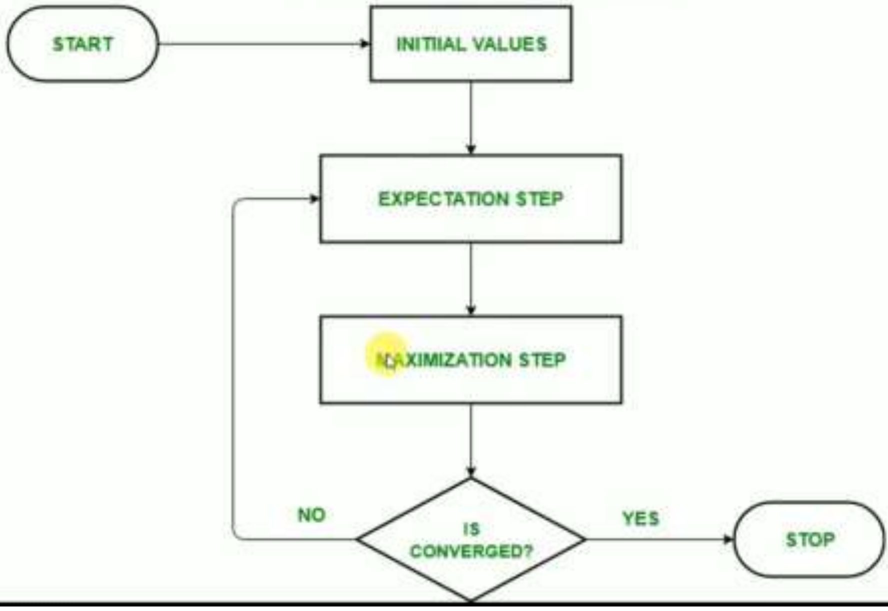
- Initially, a set of initial values of the parameters are considered. A set of incomplete observed data is given to the system with the assumption that the observed data comes from a specific model.
- The next step is known as "Expectation" – step or *E-step*. In this step, we use the observed data in order to estimate or guess the values of the missing or incomplete data. It is basically used to update the variables.
- The next step is known as "Maximization"-step or *M-step*. In this step, we use the complete data generated in the preceding "Expectation" – step in order to update the values of the parameters. It is basically used to update the hypothesis.
- Now, in the fourth step, it is checked whether the values are converging or not, if yes, then stop otherwise repeat *step-2* and *step-3* i.e. "Expectation" – step and "Maximization" – step until the convergence occurs.

EM ALGORITHM

Algorithm:

1. Given a set of incomplete data, consider a set of starting parameters.
2. **Expectation step (E – step):** Using the observed available data of the dataset, estimate (guess) the values of the missing data.
3. **Maximization step (M – step):** Complete data generated after the expectation (E) step is used in order to update the parameters.
4. Repeat step 2 and step 3 until convergence.

EM ALGORITHM



EM ALGORITHM

Usage of EM algorithm –

- It can be used to fill the missing data in a sample.
- It can be used as the basis of unsupervised learning of clusters.
- It can be used for the purpose of estimating the parameters of Hidden Markov Model (HMM).
- It can be used for discovering the values of latent variables.

EM ALGORITHM

Advantages of EM algorithm –

- It is always guaranteed that likelihood will increase with each iteration.
- The E-step and M-step are often pretty easy for many problems in terms of implementation.
- Solutions to the M-steps often exist in the closed form.

EM ALGORITHM

Disadvantages of EM algorithm –

- It has slow convergence.
- It makes convergence to the local optima only.
- It requires both the probabilities, forward and backward (numerical optimization requires only forward probability).

Expectation-Maximization – EM Algorithm Solved Example

- Expectation-Maximization (EM) – a very popular technique for estimating parameters of probabilistic models.
- Many popular algorithms like Hidden Markov Models, Gaussian Mixtures, Kalman Filters, and others uses EM technique.
- It is beneficial **when working with data that is incomplete, has missing data points, or has unobserved latent variables**.

Expectation-Maximization – EM Algorithm Solved Example

- Assume that we have two coins, C1 and C2
- Assume the bias of C_1 is θ_1 (i.e., probability of getting heads with C_1)
- Assume the bias of C_2 is θ_2 (i.e., probability of getting heads with C2)
- We want to find θ_1, θ_2 by performing a number of trials (i.e., coin tosses)

Expectation-Maximization – EM Algorithm Solved Example

First experiment

- We choose 5 times one of the coins.
- We toss the chosen coin 10 times

H T T T H H T H T H

H H H H T H H H H H

H T H H H H H T H H

H T H T T T H H T T

T H H H T H H H T H

$$\theta_1 = \frac{number\ of\ heads\ using\ C1}{total\ number\ of\ flips\ using\ C1}$$

$$\theta_2 = \frac{number\ of\ heads\ using\ C2}{total\ number\ of\ flips\ using\ C2}$$

Expectation-Maximization – EM Algorithm Solved Example

B H T T T H H T H T H

A H H H H T H H H H H

A H T H H H H H T H H

B H T H T T T H H T T

A T H H H T H H H T H

Coin A	Coin B
	5 H, 5 T
9 H, 1 T	
8 H, 2 T	
	4 H, 6 T
7 H, 3 T	
24 H, 6 T	9 H, 11 T

$$\theta_1 = \frac{24}{24+6} = 0.8$$

$$\theta_2 = \frac{9}{9+11} = 0.45$$

Expectation-Maximization – EM Algorithm Solved Example

H T T T H H T H T H

H H H H T H H H H H

H T H H H H H T H H

H T H T T T H H T T

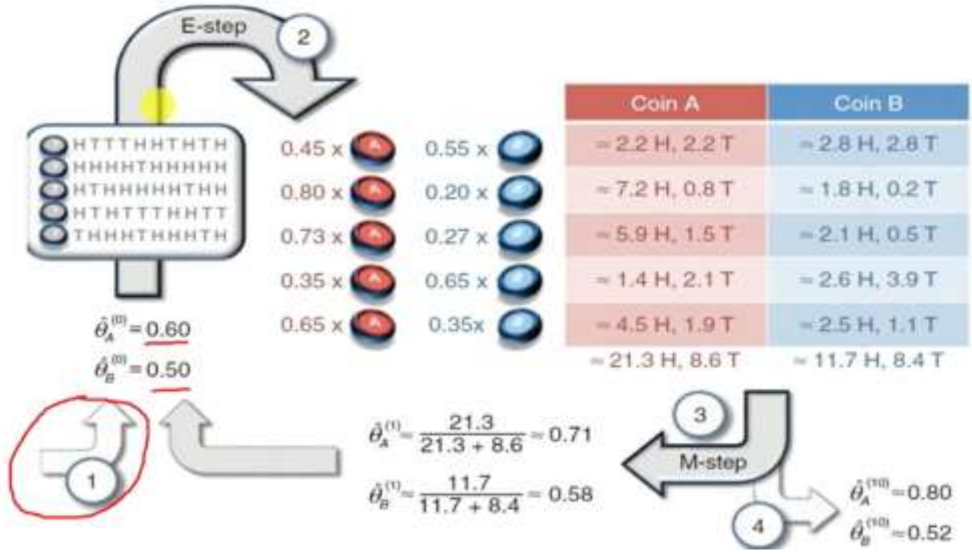
T H H H T H H H T H

Coin A	Coin B
	5 H, 5 T
9 H, 1 T	
8 H, 2 T	
	4 H, 6 T
7 H, 3 T	
24 H, 6 T	9 H, 11 T

$$\theta_1 = \frac{24}{24+6} = 0.8$$

$$\theta_2 = \frac{9}{9+11} = 0.45$$

Expectation-Maximization – EM Algorithm Solved Example



Expectation-Maximization – EM Algorithm Solved Example

$$P(E \mid Z_A) = P(HHHHHHHHHT \mid \text{A chosen}) = \binom{n}{x} \theta_A^x (1-\theta_A)^n$$

$$P(E \mid Z_B) = P(HHHHHHHHHT \mid \text{B chosen}) = \binom{n}{x} \theta_B^x (1-\theta_B)^n$$

$$P(E|Z_A) = \binom{9}{1} * (0.6)^9 * (0.4)^1 = 0.036$$

$$P(E|Z_B) = \binom{9}{1} * (0.5)^9 * (0.5)^1 = 0.009$$

Expectation-Maximization – EM Algorithm Solved Example