https://www.geeksforgeeks.org/machine-learning/gaussian-mixture-

Assignment - 1

Name: Nusrat Jannat

ID : 20121176145

Topic: Expectation-Maximization Algorithm

Course: Machine Learning

Session: 2019-20

Year: 4th year even semester

Gaussian Mixture Model: It is a probabilistic model that assumes all data points are generated from a mixture of several Guassian distributions with unknown parameters. To fit a Guassian Mixture Model to the data we use the Expectation-Maximization (EM) algorithm which is an iterative method that optimize the parameters of the Gaussian distributions like mean, covariance and mixing coeffécients. It works in two main steps:

(i) Expectation step (E-step): In this step the algorithm calculates the probability that each data points belongs to each cluster based on the current parameter estimates.

(ii) Maximization step (M-step): After estimating the probabilities the algorithm updates the parameters to better fit the data.

These two steps are repeated until the model converges meaning the parameters no longer change significantly between iterations.
Gaussian mixture model includes four steps such as initialization, E-step, M-step and repeat the steps.

Now we will see an example of 4 sequences of 10 coins. The sequences are following:

1. HTTHTTTHTT → 3 heads, 7 tails
2. THTTHTTTHH → 4 heads, 6 tails
3. HHTHHTHTHH → 7 heads, 3 tails
4. HHHTHHHHHH → 9 heads, 1 tail

So we have the number of heads and tails for each of the four sequences.

Now the steps for EM algorithm:-

(i) At first, we will guess a random estimate ($\hat{\theta}_A^{(1)}$ and $\hat{\theta}_B^{(1)}$) for this bias.

(ii) Then, we have to compute the likelihood of each sequence under each of the two coins

(iii) After that, we can convert the likelihoods into probabilities with Bayes' rule

$$P(A \mid x_i, \hat{\theta}^{(t)})$$

$$= \frac{P(x_i \mid \hat{\theta}^{(t)}) \times P(A)}{P(x_i \mid \hat{\theta}^{(t)}) \times P(A) + P(x_i \mid \hat{\theta}_B^{(t)}) \times P(B)}$$

$$= \frac{P(x_i \mid \hat{\theta}_A^{(t)}) \times \frac{1}{2}}{P(x_i \mid \hat{\theta}^{(t)}) \times \frac{1}{2} + P(x_i \mid \hat{\theta}_B^{(t)}) \times \frac{1}{2}} = \frac{L(A)}{L(A)+L(B)}$$

(iv) Then, checking the probability we will attribute the heads/tails to coin A and B according to the probabilities.

(v) Now, we have to figure out how many heads and tails to coin A and Coin B, by adding the columns.

(vi) Then we will get next estimates of the parameters

(vii) Then we will repeat these steps until we have reached a local optimum and the estimates stop changing so much.

In the video tutorial total 4 estimates out of 10 estimates have been shown. So we will only find out the other ~~es~~ six remaining estimates and demonstrate the calculation from $\hat{\theta}_A^{(4)}$ and $\hat{\theta}_B^{(4)}$.

we have at last got $\hat{\theta}_A^{(3)} = 0.785$ and $\hat{\theta}_B^{(3)} = 0.361$.

Heads	Tails	Likelihood with Coin A, L(A)	Likelihood with Coin B, L(B)	Prob (coin A), P(A)	Prob (Coin B), P(B)	Heads/Tails Attributed to Coin A	Heads/Tails Attributed to Coin B
H	T	$(\hat{\theta}_A^t)^H \times (1-\hat{\theta}_A^t)^T$	$(\hat{\theta}_B^t)^H \times (1-\hat{\theta}_B^t)^T$	$\frac{L(A)}{L(A)+L(B)}$	$\frac{L(B)}{L(A)+L(B)}$	$H \times P(A)$ $T \times P(A)$	$H \times P(B)$ $T \times P(B)$
3	7	0.00001	0.002	0.0049	0.995	0.0147H 0.0343 T	2.985H 6.965T
4	6	0.000038	0.0015	0.024	0.97	0.096H 0.144T	3.88H 5.82T
7	3	0.0018	0.0002	0.9	0.1	6.3H 2.7T	0.7H 0.3T
9	1	0.024	0.000067	0.99	0.0028	8.91H 0.99T	0.0252H 0.0028T
Total						15.3207H 3.8683T	7.59H 13.0878T

So, $\hat{\theta}_A^{(4)} = \dfrac{15.3207}{15.3207 + 3.8683} = 0.798$

$$\hat{\theta}_B^{(4)} = \frac{7.59}{7.59 + 13.0878} = 0.367$$

Now we will find out our new estimate $\hat{\theta}_A^{(5)}$ and $\hat{\theta}_B^{(5)}$

Here for calculating $\hat{\theta}_A^t$ and $\hat{\theta}_B^t$, we use the formula,

$$\hat{\theta}_A^t = \frac{\text{Heads Attributed to coin A}}{\text{Heads attributed to coin A} + \text{Tails attributed to coin A}}$$

$$\hat{\theta}_B^t = \frac{\text{Heads attributed to coin B}}{\text{Heads attributed to coin B} + \text{Tails attributed to coin B}}$$

Heads	Tails	Likelihood with coin A, L(A)	Likelihood with coin B, L(B)	Prob (coin A), P(A)	Prob (Coin B), P(B)	Heads/Tails Attributed to coin A	Heads/Tails Attributed to Coin B
H	T	$(\hat{\theta}_A^{t})^H \times (1-\hat{\theta}_A^{t})^T$	$(\hat{\theta}_B^{t})^H \times (1-\hat{\theta}_B^{t})^T$	$\frac{L(A)}{L(A)+L(B)}$	$\frac{L(B)}{L(A)+L(B)}$	$H \times P(A)$ $T \times P(A)$	$H \times P(B)$ $T \times P(B)$
3	7	0.0000069	0.002	0.003	0.996	0.0092 H 0.0214 T	2.988 H 6.972 T
4	6	0.000027	0.001	0.026	0.973	0.1 H 0.156 T	3.892 H 5.838 T
7	3	0.0017	0.00022	0.885	0.114	6.195 H 2.655 T	0.798 H 0.342 T
9	1	0.02	0.000076	0.00026	0.0038	0.002 H 0.00026 T	0.0342 H 0.0038 T
Total						6.306 H 2.83 T	7.712 H 13.155 T

So, $\theta_A^{(5)} = \frac{6.306}{6.306+2.83} = 0.69$

$$\theta_B^{(5)} = \frac{7.712}{7.712+13.155} = 0.38$$

Now, we will calculate $\theta_A^{(6)}$ and $\theta_B^{(6)}$.

Heads	Tails	Likelihood with coin A, L(A)	Likelihood with coin B, L(B)	Prob (Coin A), P(A)	Prob (Coin B), P(B)	Heads/Tails Attributed to coin A	Heads/Tails Attributed to coin B
H	T	$(\hat{\theta}_A^*)^H \times (1-\hat{\theta}_A^*)^T$	$(\hat{\theta}_B^*)^H \times (1-\hat{\theta}_B^*)^T$	$\frac{L(A)}{L(A)+L(B)}$	$\frac{L(B)}{L(A)+L(B)}$	$H \times P(A)$ $T \times P(A)$	$H \times P(B)$ $T \times P(B)$
3	7	0.00009	0.0019	0.045	0.95	0.135 H 0.315 T	2.85 H 6.65 T
4	6	0.0002	0.001	0.166	0.83	0.664 H 0.996 T	3.32 H 4.98 T
7	3	0.0022	0.00027	0.89	0.109	6.23 H 2.67 T	0.763 H 0.327 T
9	1	0.01	0.0001	0.99	0.0099	8.91 H 0.99 T	0.0891 H 0.0099 T
Total						15.939 H 4.971 T	7.0221 H 11.9669 T

So,
$$\hat{\theta}_A^{(6)} = \frac{15.939}{15.939 + 4.971} = 0.762$$

$$\hat{\theta}_B^{(6)} = \frac{7.0221}{7.0221 + 11.9669} = 0.369$$

Our next target is to find out $\hat{\theta}_A^{(7)}$ and $\hat{\theta}_B^{(7)}$.

Heads	Tails	Likelihood with coin A, L(A)	Likelihood with coin B, L(B)	Prob (coin A), P(A)	Prob (Coin B), P(B)	Heads/Tails Attributed to coin A	Heads/Tails Attributed to coin B
H	T	$(\hat{\theta}_A^{t})^H \times (1-\hat{\theta}_A^{t})^T$	$(\hat{\theta}_B^{t})^H \times (1-\hat{\theta}_B^{t})^T$	$\frac{L(A)}{L(A)+L(B)}$	$\frac{L(B)}{L(A)+L(B)}$	$H \times P(A)$ $T \times P(A)$	$H \times P(B)$ $T \times P(B)$
3	7	0.000019	0.002	0.0094	0.99	0.0282H 0.0658T	2.97H 6.93T
4	6	0.00006	0.001	0.0566	0.94	0.2264H 0.3396T	3.76H 5.64T
7	3	0.002	0.00023	0.896	0.1	6.272H 2.688T	0.7H 0.3T
9	1	0.02	0.00008	0.996~~15~~	0.038	8.964H 0.996T	0.342H 0.038T
Total						15.4906H 4.0894T	7.772H 12.908T

So, $\hat{\theta}_A^{(7)} = \frac{15.4906}{15.4906+4.0894} = 0.791$

$\hat{\theta}_B^{(7)} = \frac{7.772}{7.772+12.908} = 0.376$

Again, we will re-estimate $\hat{\theta}_A^{(8)}$ and $\hat{\theta}_B^{(8)}$.

Heads	Tails	Likelihood with coin A, L(A)	Likelihood with coin B, L(B)	Prob (coin A), P(A)	Prob (coin B), P(B)	Heads/Tails Attributed to coin A	Heads/Tails Attributed to coin B
H	T	$(\hat{\theta}_A^*)^H \times (1-\hat{\theta}_A^*)^T$	$(\hat{\theta}_B^*)^H \times (1-\hat{\theta}_B^*)^T$	$\frac{L(A)}{L(A)+L(B)}$	$\frac{L(B)}{L(A)+L(B)}$	$H \times P(A)$ $T \times P(A)$	$H \times P(B)$ $T \times P(B)$
3	7	0.0000086	0.00196	0.0044	0.996	0.0132H 0.0308T	2.988H 6.972T
4	6	0.000033	0.0012	0.027	0.97	0.108H 0.162T	3.88H 5.82T
7	3	0.0018	0.00026	0.87	0.126	6.09H 2.61T	0.882H 0.378T
9	1	0.025	0.000094	0.996	0.0037	8.964H 0.996T	0.033H 0.996T
Total						15.1752H 3.7988T	7.783H 14.166T

So, $\hat{\theta}^{8}_{(A)} = \frac{15.1752}{15.1752 + 3.7988} = 0.799$

$\hat{\theta}^{8}_{(B)} = \frac{7.783}{7.783 + 14.166} = 0.354$

At last, we will calculate the estimation value $\hat{\theta}^{(9)}_{(A)}$ and $\hat{\theta}^{(9)}_{B}$.

Heads	Tails	Likelihood with the coin A, L(A)	Likelihood with coin B, L(B)	Prob (coin A), P(A)	Prob (coin B), P(B)	Heads/Tails Attributed to coin A	Heads/Tails Attributed to coin B
H	T	$(\hat{\theta}_A^t)^H \times (1-\hat{\theta}_A^t)^T$	$(\hat{\theta}_B^t)^H \times (1-\hat{\theta}_B^t)^T$	$\frac{L(A)}{L(A)+L(B)}$	$\frac{L(B)}{L(A)+L(B)}$	$H \times P(A)$ $T \times P(A)$	$H \times P(B)$ $T \times P(B)$
3	7	0.0000068	0.002	0.003	0.997	0.009 H 0.021 T	2.991 H 6.979 T
4	6	0.000027	0.001	0.026	0.974	0.1 H 0.156 T	3.896 H 5.844 T
7	3	0.0017	0.00019	0.899	0.1	6.293 H 2.697 T	0.7 H 0.3 T
9	1	0.027	0.000056	0.997	0.002	8.973 H 0.997 T	0.018 H 0.002 T
Total						15.375 H 3.871 T	7.605 H 13.125 T

So, $\hat{\theta}_A^{(0)} = \frac{15.375}{15.375+3.871} = 0.798$

$\hat{\theta}_B^{(0)} = \frac{7.605}{7.605+13.125} = 0.367$

From the video, we get $\hat{\theta}_A^{(10)} = 0.801$, $\hat{\theta}_B^{(10)} = 0.3715$

Eventually, the answers will stop changing very much. We reach a local optimum.

That's the EM algorithm as we take an expectation which accounts for the uncertainty ~~and~~ in E-step and in M-step we will choose the $\hat{\theta}_A$ and $\hat{\theta}_B$ that maximize the likelihood of observing our data.