NPTEL Week-9 Live Session

on Machine Learning and Deep Learning - Fundamentals and Applications (noc24_ee146)

A course offered by: Prof. Manas Kamal Bhuyan, IIT Guwahati

NPTEL Quiz Solution: Week-8 → GMM
→ K-means clustering.

$D = \{x, y\}$ → supervised
Feature, label

$D = \{x\}$ → unsupervised case.

Powered by:

PMRF
Prime Minister's Research Fellows
Ministry of Education
Government of India

By

Arka Roy
NPTEL PMRF TA

Prime Minister's Research Fellow
Department of Electrical Engineering, IIT Patna
Web: https://sites.google.com/view/arka-roy/home

1) The EM algorithm is based on the principle of

- ✓ Maximum likelihood estimation
- Bayesian theory
- Feature selection
- Regularization

GMM = linear Combination of different gaussian distribution

$= \sum_{i=1}^{k} \pi_k \, N(x \mid \mu_k, \Sigma_k)$ $\quad x \in \mathbb{R}^{d \times 1}$

Multivariate Gaussian Distribution

MLE :-

$f(x \mid \theta)$

$\ln f(x \mid \theta)$

$\frac{\partial}{\partial \theta} \ln f(x \mid \theta) = 0$

$\theta = \{\mu, \sigma^2\}$.

$GMM = \sum_k \pi_k N(x \mid \mu_k, \Sigma_k)$

* If we have to evaluate these parameters using MLE then, we have to do partial differentiation for all such k no of μ_k and Σ_k.

* Therefore solving such k no of differential solutions is difficult as for analytical solution is concerned

* Therefore we go by computational method of solving through EM algorithm

2) Each component in a GMM is characterized by

- Mean, ~~median, and mode~~
- Mean and covariance matrix
- Variance, ~~skewness, and kurtosis~~
- Standard deviation and ~~correlation coefficient~~

Mean = 1st order statistical moment

Variance = 2nd order statistical moment

In case of Gaussian or normal dist = $N(\mu, \sigma^2)$

$$\rightarrow GMM = \sum_{k=1}^{K} \pi_k N(x \mid \mu_k, \Sigma_k)$$

$x \in R^{d \times 1}$

π_k → weighting / mixing coefficient

Normal distribution → multi variate normal distribution.

Parameterised by → Mean, & Covariance matrix

GMM → weighting factor, μ_k, Σ_k of multivariate Gaussian distribution

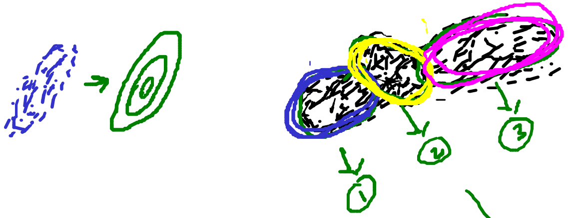
$$CC = \frac{\text{Covariance}(x, d)}{\text{Var}(x d) \cdot \text{Var}(x d)}$$

3) Which of the following statements is true about GMMs?

- GMMs are only used for ~~regression tasks~~ clustering Task.
- GMMs can only model data with a single cluster
- GMMs are sensitive to the initial placement of the Gaussian components
- GMMs are a type of neural network

Black box → heuristic approach → learning based

statistical algorithm process



$$\sum_k \pi_k N(x \mid \Sigma_k, \mu_k)$$

GMM → used for clustering problems → unsupervised learning approach

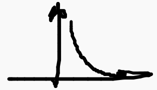
4) What does each component in a GMM represent?

- A data point in the dataset
- A principal component
- A cluster center (μ_k)
- A Gaussian distribution with its mean and covariance

$$GMM = \sum_{k} \pi_k \, N(x \mid \mu_k, \Sigma_k)$$

5) Let there be 10 data points: $\{-1, 27, 31, 2, 59, 3, 61, 34, 0, 12\}$. Use K-means clustering for 3 iterations to cluster them into 3 clusters. The initial labels are $\{1, 1, 1, 2, 2, 2, 3, 3, 3, 1\}$. List out the labels after 3 iterations.

$\{1, 2, 2, 1, 3 \ldots\}$



- ○ $\{1, 2, 2, 1, 3, 1, 3, 2, 1, 1\}$
- ○ $\{1, 1, 2, 1, 3, 1, 3, 2, 1, 2\}$
- ○ $\{1, 2, 3, 1, 3, 1, 3, 2, 3, 1\}$
- ○ $\{1, 1, 2, 1, 3, 1, 3, 3, 1, 1\}$

$D = \{-1, 27, 31, 2, 59, 3, 61, 34, 0, 12\}$

$\mathcal{L} = \{1, 1, 1, 2, 2, 2, 3, 3, 3, 1\}$

(C_1)

$D_1 = \{-1, 27, 31, 12\} \in$ Class 1 $\Rightarrow$ Centroid of class 1 data cloud $= \frac{-1+27+31+12}{4} = 17.25$

Centroid-2 $= (2+59+3)/3 = 21.33$ (C_2)

Centroid-3 $= (0+61+34)/3 = 31.66$ (C_3)

(ϕ)

Distance $= \sqrt{(x_1 - x_2)^2}$ (ℓ^2 dist)

$= |x_1 - x_2|$ (ℓ^1 dist)

It-1:-

$\phi(P_1, C_1) = |-1 - 17.25| = 18.25$
$\phi(P_1, C_2) = 22.33$
$\phi(P_1, C_3) = 32.66$
$\rightarrow P \in \omega_1$

$\phi(P_2, C_1) = 9.75$
$\phi(P_2, C_2) = 5.67$
$\phi(P_2, C_3) = 4.66$
$\rightarrow P_2 \in \omega_3$

$\phi(P_3, C_1) = 13.75$
$\phi(P_3, C_2) = 9.67$
$\phi(P_3, C_3) = 0.67$
$\Rightarrow P_3 \in \omega_3$

$\phi(P_4, C_1) = 15.25$, $\phi(P_4, C_2) = 19.33$, $\phi(P_4, C_3) = 29.67$

$P_4 \in \omega_1$

$\phi(P_5, C_1) = 41.75$
$\phi(P_5, C_2) = 37.65$
$\phi(P_5, C_3) = 27.33$
$\rightarrow P_5 \in \omega_3$

$\phi(P_6, C_1) = 14.25$
$\phi(P_6, C_2) = 18.33$
$\phi(P_6, C_3) = 28.67$
$P_6 \in \omega_1$

5) Let there be 10 data points: $\{-1, 27, 31, 2, 59, 3, 61, 34, 0, 12\}$. Use K-means clustering for 3 iterations to cluster them into 3 clusters. The initial labels are $\{1, 1, 1, 2, 2, 2, 3, 3, 3, 1\}$. List out the labels after 3 iterations.

- $\{1,2,2,1,3,1,3,2,1,1\}$
- $\{1,1,2,1,3,1,3,2,1,2\}$
- $\{1,2,3,1,3,1,3,2,3,1\}$
- $\{1,1,2,1,3,1,3,3,1,1\}$

$D = \{-1, 27, 31, 2, 59, 3, 61, 34, 0, 12\}$

$l_0 = \{1, 1, 1, 2, 2, 2, 3, 3, 3, 1\}$

$\left.\begin{array}{l}\phi(P_7, c_1) = 43.75 \\ \phi(P_7, c_2) = 39.67 \\ \phi(P_7, c_3) = 29.33\end{array}\right\} \Rightarrow P_7 \in \omega_3.$

$\left.\begin{array}{l}\phi(P_8, c_1) = 16.75 \\ \phi(P_8, c_2) = 12.67 \\ \phi(P_8, c_3) = 2.33\end{array}\right\} \Rightarrow P_8 \in \omega_3$

$\left.\begin{array}{l}\phi(P_9, c_1) = 17.25 \\ \phi(P_9, c_2) = 21.33 \\ \phi(P_9, c_3) = 31.67\end{array}\right\}$

$P_9 \in \omega_1$

$\left.\begin{array}{l}\phi(P_{10}, c_1) = 5.25 \\ \phi(P_{10}, c_2) = 9.33 \\ \phi(P_{10}, c_3) = 19.67\end{array}\right\} \Rightarrow P_{10} \in \omega_1$

Therefore after iteration 1

$l_1 = \{1, 3, 3, 1, 3, 1, 3, 3, 1, 1\}$

After It-1:-

$\left.\begin{array}{l}\text{Centroid-1} = -1 + 2 + 3 + 0 + 12 / 5 = 3.2 \\ \text{Centroid-2} = 21.33 \\ \text{Centroid-3} = 27 + 31 + 59 + 61 + 34 / 5 = 42.4\end{array}\right\} \rightarrow \text{start iteration 2}$

5) Let there be 10 data points: $\{-1, 27, 31, 2, 59, 3, 61, 34, 0, 12\}$. Use K-means clustering for 3 iterations to cluster them into 3 clusters. The initial labels are $\{1,1,1,2,2,2,3,3,3,1\}$. List out the labels after 3 iterations.

- $\{1,2,2,1,3,1,3,2,1,1\}$
- $\{1,1,2,1,3,1,3,2,1,2\}$
- $\{1,2,3,1,3,1,3,2,3,1\}$
- $\{1,1,2,1,3,1,3,3,1,1\}$

$D = \{-1, 27, 31, 2, 59, 3, 61, 34, 0, 12\}$

$L_1 = \{1, 3, 3, 1, 3, 1, 3, 3, 1, 1\}$

$C_1 = 3.2, \quad C_2 = 21.33, \quad C_3 = 42.4$

$d(P_1 C_1) = 4.2$
$d(P_1 C_2) = 22.33$
$d(P_1 C_3) = 43.4$
$\Rightarrow P_1 \in W_1$

$d(P_2 C_1) = 24.2$
$d(P_2 C_2) = 6.67$
$d(P_2 C_3) = 15.4$
$\Rightarrow P_2 \in W_2$

$d(P_3 C_1) = 27.8$
$d(P_3 C_2) = 9.67$
$d(P_3 C_3) = 11.4$
$\Rightarrow P_3 \in W_2$

$P_4 \in W_1$
$P_5 \in W_3$
$P_6 \in W_1$

$P_7 \in W_3$
$P_8 \in W_3$
$P_9 \in W_1$
$P_{10} \in W_1$

$|12 - 3.2| = 8.8$
$|12 - 21.33| = 9.33$
$|12 - 42.4|$
$\Rightarrow P_{10} \in W_1$

$L_2 = \{1, 2, 2, 1, 3, 1, 3, 3, 1, 1\}$

$d(P_8 C_1) = 30.8$
$d(P_8 C_2) = 12.67$
$d(P_8 C_3) = 8.4$

$\text{Centroid-1} = \frac{-1+2+3+0+12}{5} = 3.2$

$\text{Centroid-2} = \frac{27+31}{2} = 29$

$\text{Centroid-3} = \frac{59+61+34}{3} = 51.33$

5) Let there be 10 data points: $\{-1, 27, 31, 2, 59, 3, 61, 34, 0, 12\}$. Use K-means clustering for 3 iterations to cluster them into 3 clusters. The initial labels are $\{1, 1, 1, 2, 2, 2, 3, 3, 3, 1\}$. List out the labels after 3 iterations.

- [x] $\{1, 2, 2, 1, 3, 1, 3, 2, 1, 1\}$
- [ ] $\{1, 1, 2, 1, 3, 1, 3, 2, 1, 2\}$
- [ ] $\{1, 2, 3, 1, 3, 1, 3, 2, 3, 1\}$
- [ ] $\{1, 1, 2, 1, 3, 1, 3, 3, 1, 1\}$

$D = \{-1, 27, 31, 2, 59, 3, 61, 34, 0, 12\}$

$C_1 = 3.2$

$C_2 = 29$

$C_3 = 51.33$

$L_2 = \{1, 2, 2, 1, 3, 1, 3, 2, 1, 1\}$

Third iteration:-

$P_1 \in \omega_1$

$P_2 \in \omega_2$

$P_3 \in \omega_2$

$P_4 \in \omega_1$

$P_5 \in \omega_3$

$P_6 \in \omega_1$

$P_7 \in \omega_3$

$P_8 \in \omega_2$

$P_9 \in \omega_1$

$P_{10} \in \omega_1$

$$\boxed{L_3 = \{1, 2, 2, 1, 3, 1, 3, 2, 1, 1\}}$$

Centroid-1 $= -1+2+3+0+12/5 = 3.2$

Centroid-2 $= 27+31+34/3 = 30.66$

Centroid 3 $= \dfrac{61+59}{2} = 60$

Centroid $= \{3.2, 30.66, 60\}$

6) For the above question, find the centroids after 3 iterations.

$D = \{-1, 27, 31, 2, 59, 3, 61, 34, 0, 12\}$

$\mathcal{L}_3 = \{1, 2, 2, 1, 3, 1, 3, 2, 1, 1\}$.

- $\{5, 16.5, 56\}$
- $\{3, 21.33, 42.4\}$
- $3.2, 30.67, 60$
- $3.2, 29.51.33$

Centroid-1 = $-1+2+3+0+12/5 = 3.2$

Centroid-2 = $27+31+34/3 = 30.66$

Centroid 3 = $\frac{61+59}{2} = 60$

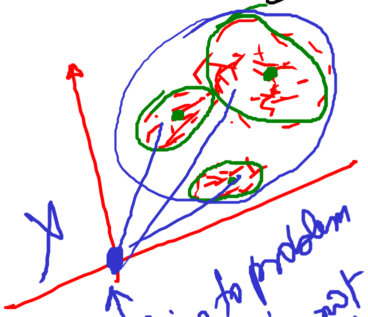
Centroid = $\{3.2, 30.66, 60\}$

7) How does K-Means initialize cluster centroids?

$L \rightarrow$ Random

- At the maximum data point values
- Based on the mean of the data points
- At the origin (0,0)
- Randomly

$$w_{n+1} = w_n - \eta \frac{\partial E}{\partial w} \rightarrow \text{gradient}$$

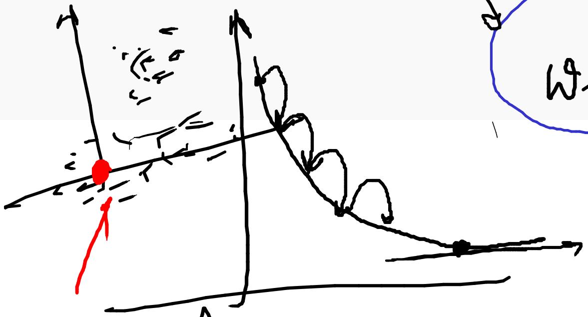


bias to problem solution is not generic for arbitrary data cloud dist

$K = 1, 2, 3$

$K = 1, 2$

$\frac{1+2}{2} = \frac{3}{2} = 1.5$



$w(0) \rightarrow$ Random initialization

$w_{n+1} = w_n$

$x_{n+1} = x_n - \eta\, \text{grad}(f)$

8) What is the best way to determine the optimal number of clusters (K) in K-Means?

- Trial and error
- Using the Elbow Method
- Setting K equal to the number of data points
- Using the Silhouette score

Sum of square error loss.

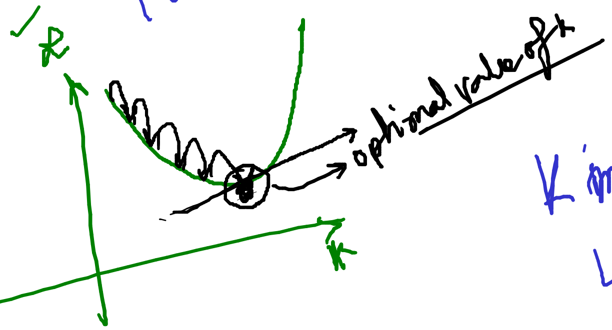
$L = \sum (y - \hat{y})^2$

True output

Tangent output from K-mean clustering.

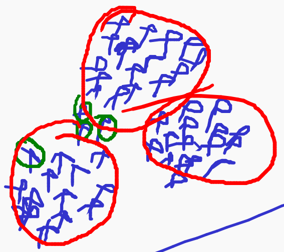
Quadratic equation (Parabola)

Convex



$\hat{y} = f(K)$

100 groups of pen



100 data points

Pen → different colour
→ different length

K may vary
→ Problem specific
→ heuristically chosen variable