# ML HW2

Jadon Co

October 2023

# 1 KMeans Clustering and DBScan

## 1.1 Programming K-Means

# 2 EM Algorithm

## 2.1 Write the marginal probability of x, (p(x))

Sum Rule:

$$\text{p(x)} = \text{P}(\text{x}_n|z=0) * P(z=0) + P(x_n|z=1) * P(z=1) =$$

Given: $\text{P(z=0)} = \alpha, P(z=1) = 1-\alpha$

$$\text{P}(\text{x}_n|z=0) = N(6v + 5\omega^2)$$
$$\text{P}(\text{x}_n|z=1) = N(3v + 7\omega^2)$$

Therefore,

$$\text{p(x)} = \alpha(N(6v + 5\omega^2)) + (1-\alpha)(N(3v + 7\omega^2))$$

## 2.2 E-Step: Compute the posterior probabilities, $\mathbf{p}(\mathbf{z}_0|x), p(z_1|x)$

Bayes Rule:

$$\mathrm{P}(\mathrm{z}_0|x) = \frac{P(x|z_0)*P(z_0)}{P(x)} = T_0$$

$$\mathrm{P}(\mathrm{z}_1|x) = \frac{P(x|z_1)*P(z_1)}{P(x)} = T_1$$

$$\mathrm{T}_0 = \frac{\alpha*(N(6v+5\omega^2))}{\alpha(N(6v+5\omega^2))+(1-\alpha)(N(3v+7\omega^2))}$$

$$\mathrm{T}_1 = \frac{(1-\alpha)(N(3v+7\omega^2))}{\alpha(N(6v+5\omega^2))+(1-\alpha)(N(3v+7\omega^2))}$$

### 2.3 M-Step: Compute the updated value of $\omega^2$

$$l(\theta|x) = \sum^N \sum^Z p(z_k|x_n, \theta_{old}) ln[p(x_n, z_k|\theta)]$$

$$l(v,\ \omega^2, \alpha|x) = \sum^N \sum^Z p(z_k|x_n, \theta_{old}) ln[p(x_n, z_k|\mu_k, \sigma_k^2, \alpha)]$$

$$=$$

$$\sum^N \sum^Z p(z_k|x_n, \theta_{old}) ln[p(z_k|\alpha) p(x_n|z_k, v, \omega^2]$$

$$\mathrm{p}(\mathrm{z}_0|\alpha) = \alpha, p(z_1|\alpha) = 1-\alpha$$

$$\mathrm{T}_0 = p(z_0|x_n, \theta_{old}) = \frac{\alpha*(N(6v+5\omega^2))}{\alpha(N(6v+5\omega^2))+(1-\alpha)(N(3v+7\omega^2))}$$

$$\mathrm{T}_1 = p(z_1|x_n, \theta_{old}) = \frac{(1-\alpha)(N(3v+7\omega^2))}{\alpha(N(6v+5\omega^2))+(1-\alpha)(N(3v+7\omega^2))}$$

$$p(x_n|z_0,\theta) = N(x_n|6v{+}5\omega^2), p(x_n|z_1,\theta) = N(x_n|3v{+}7\omega^2)$$

$$= \Sigma^N(T_0 * ln(\alpha * N(x_n|6v+5\omega^2)) + T_1 * ln((1-\alpha) * N(x_n|3v+7\omega^2)))$$

first part of summation $= T_0 * ln(\alpha * N(x_n|6v+5\omega^2))$

second part of summation =
$T_1 * ln((1-\alpha) * N(x_n|3v+7\omega^2))$

LN properties:
$\ln(a^b) = b * ln(a)$
ln(ab) = ln(a) + ln(b)
$\ln(e^x) = x$
ln(a/b) = ln(a) - ln(b)

simplification of first part of summation:

$$T_0 * ln(\alpha * (\frac{1}{\sqrt{2*\pi*\sigma^2}} * e^{\frac{(x_n-\mu)^2}{-2*\sigma^2}})$$

$$T_0 * ln(\alpha * (\frac{1}{\sqrt{2*\pi*5\omega^2}} * e^{\frac{(x_n-6v)^2}{-2*5\omega^2}})$$

$$T_0 * (ln(a) + ln(\frac{1}{\sqrt{2*\pi*5\omega^2}}) + ln(e^{\frac{(x_n-6v)^2}{-2*5\omega^2}}))$$

$$T_0 * (ln(a) + ln(\frac{1}{\sqrt{10\pi\omega^2}}) + \frac{(x_n-6v)^2}{-10\omega^2})$$

$$T_0 * (ln(a) + (-0.5)ln(10\pi\omega^2) - \frac{(x_n-6v)^2}{10\omega^2})$$

$$\text{df/d}\omega^2 \rightarrow \textbf{differentiate w/ respect to } \omega^2$$
$$\text{df/d}\omega^2(ln(\alpha)) = 0$$
$$\text{df/d}\omega^2((-0.5)ln(10\pi\omega^2)) = 0.5 * 10\pi * \frac{1}{10\pi\omega^2}$$
$$\text{df/d}\omega^2(\frac{(x_n-6v)^2}{-10\omega^2}) = \frac{(x_n-6v)^2}{10\omega^4}$$
$$\text{T}_0 * (-0.5 * 10\pi * \frac{1}{10\pi\omega^2} + \frac{(x_n-6v)^2}{10\omega^4})$$
$$\text{T}_0 * (\frac{-1}{2\omega^2} + \frac{(x_n-6v)^2}{10\omega^4})$$

simplification of second part of summation:

$$\text{T}_1 * ln((1-\alpha) * (\frac{1}{\sqrt{2*\pi*\sigma^2}} * e^{\frac{(x_n-\mu)^2}{-2*\sigma^2}})$$
$$\text{T}_1 * ln((1-\alpha) * (\frac{1}{\sqrt{2\pi*7\omega^2}} * e^{\frac{(x_n-3v)^2}{-2*7\omega^2}})$$
$$\text{T}_1 * (ln(1-a) + ln(\frac{1}{\sqrt{2\pi*7\omega^2}}) + ln(e^{\frac{(x_n-3v)^2}{-2*7\omega^2}}))$$
$$\text{T}_1 * (ln(1-a) + ln(\frac{1}{\sqrt{14\pi\omega^2}}) + \frac{(x_n-3v)^2}{-14\omega^2})$$
$$\text{T}_1 * (ln(1-a) + (-0.5)ln(14\pi\omega^2) - \frac{(x_n-3v)^2}{14\omega^2})$$
$$\text{df/d}\omega^2 \rightarrow \textbf{differentiate w/ respect to } \omega^2$$
$$\text{df/d}\omega^2(ln(1-\alpha)) = 0$$
$$\text{df/d}\omega^2((-0.5)ln(14\pi\omega^2)) = 0.5 * 14\pi * \frac{1}{14\pi\omega^2}$$
$$\text{df/d}\omega^2(\frac{(x_n-6v)^2}{-14\omega^2}) = \frac{(x_n-3v)^2}{14\omega^4}$$
$$\text{T}_1 * (-0.5 * 14\pi * \frac{1}{14\pi\omega^2} + \frac{(x_n-3v)^2}{14\omega^4})$$

$$\mathrm{T}_1 * (\frac{-1}{2\omega^2} + \frac{(x_n-3v)^2}{14\omega^4})$$

$$\partial l(\partial|x)/\partial\omega^2 =$$
$$\Sigma^N[T_0(\frac{-1}{2\omega^2} + \frac{(x_n-6v)^2}{10\omega^4}) + T_1 * (\frac{-1}{2\omega^2} + \frac{(x_n-3v)^2}{14\omega^4})] = 0$$

Solve For $\omega^2$ :

$$\tau_0 = \alpha * N(6v + 5\omega^2),$$
$$\tau_1 = (1-\alpha) * N(3v + 7\omega^2)$$

$$\partial l(\partial|x)/\partial\omega^2 =$$

$$\Sigma^N[T_0(\frac{-1}{2\omega^2} + \frac{(x_n-6v)^2}{10\omega^4}) + T_1 * (\frac{-1}{2\omega^2} + \frac{(x_n-3v)^2}{14\omega^4})] = 0$$

$$\Sigma_{n=1}^N c = c + c + c... = cN$$

$$\Sigma^N(T_0(\frac{-1}{2\omega^2})+T_1(\frac{-1}{2\omega^2}))+\Sigma^N(T_0(\frac{(x_n-6v)^2}{10\omega^4})+T_1(\frac{(x_n-3v)^2}{14\omega^4})) = 0$$

$$\mathrm{N*T}_0 * \frac{-1}{2\omega^2} + N * T_1 * \frac{-1}{2\omega^2} + \Sigma^N(T_0(\frac{(x_n-6v)^2}{10\omega^4}) + T_1(\frac{(x_n-3v)^2}{14\omega^4}))$$

$$\mathrm{NT}_0 * \frac{-1}{2\omega^2} + NT_1 * \frac{-1}{2\omega^2} + \frac{T_0}{10\omega^4}\Sigma^N(x_n-6v)^2 +$$
$$\frac{T_1}{14\omega^4}\Sigma^N(x_n-3v)^2 = 0$$

multiply by $\omega^4$

$$\text{-}(\omega^2)/2*(T_0+T_1)+\frac{T_0}{10}\Sigma^N(x_n-6v)^2+\frac{T_1}{14}\Sigma^N(x_n-3v)^2 = 0$$

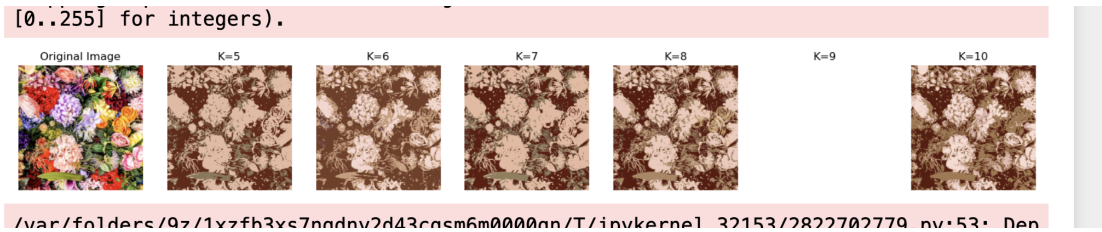
$$(\text{-}\omega^2)/2 * N * (T_0 + T_1) =$$
$$-(\frac{T_0}{10}\Sigma^N(x_n-6v)^2 + \frac{T_1}{14}\Sigma^N(x_n-3v)^2)$$

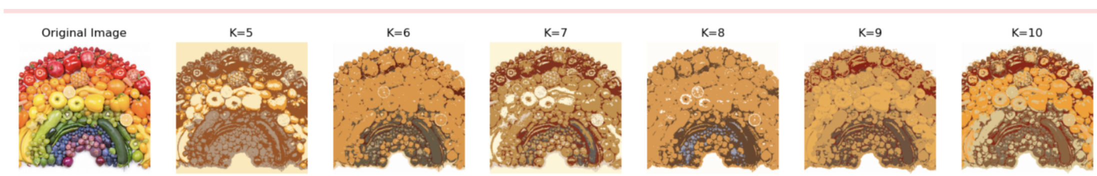
$$\omega^2 = \frac{1}{5}\Sigma^N(x_n-6v)^2 + \frac{1}{7}\Sigma^N(x_n-3v)^2)$$

$$\omega^2 = \frac{\tau_0}{5N}\Sigma^N(x_n - 6v)^2 + \frac{\tau_1}{7N}\Sigma^N(x_n - 3v)^2)$$

$$\omega = \sqrt{\frac{\tau_0}{5N}\Sigma^N(x_n - 6v)^2 + \frac{\tau_1}{7N}\Sigma^N(x_n - 3v)^2)}$$

### 2.4 Question 3



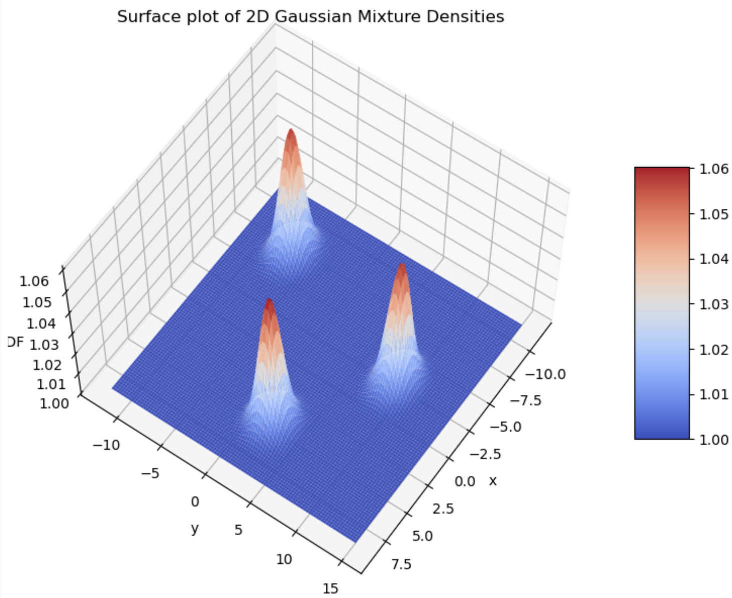




**Written Questions [2pts]:**

1. Why should we add the maximum for each row of $logit$ to **logsumexp()** function?
   *-- Use a simple input like* $logit \in \mathbb{R}^{1\times 3}$ *and work through a mathematical example.*
   *-- Let N=1, D=3,* $logit = \{logit_{11}, logit * 12, logit * 13\}$ *and* $max = logit_{13}$ *is the maximum for this row.*
   -- _Start by subtracting the max of the row from each element in $s1 = \log\left(\sum_{j=1}^{D} \exp(logit * 1, j)\right)$ *

**Answer:** We add the maximum for each row of logit to logsumexp() because we are working with exponential functions. By subtracting the max of the row from each element in s1, we see that the numbers become ridciously small, effectively reducing the results of the function. Thus, by adding the max for each row, we will ideally prevent underflow issues, as the value will be fairly large enough to extract data from the results.

```python
ax.view_init(60, 35)
```




In [150]:
```python
################################
### DO NOT CHANGE THIS CELL ###
################################

fig = plt.figure(figsize=(13, 7), dpi=100)
ax = plt.axes(projection="3d")
w = ax.plot_wireframe(xx, yy, densities)
ax.set_xlabel("x")
ax.set_ylabel("y")
ax.set_zlabel("PDF")
ax.set_title("Wireframe plot of 2D Gaussian Mixture")
```