CSCI 567 Homework 4

Qilin Ye & Furong (Flora) Jia

November 17, 2022

Problem 1

The numbers corresponding to each leaf and each metric are shown below.

$\mathcal{T}_1$ left leaf	entropy	$-(1/4\log(1/4)+3/4\log(3/4))\approx 0.562$
	Gini impurity	$1/4\cdot 3/4+3/4\cdot 1/4=0.375$
	cl. error	0.25
$\mathcal{T}_1$ right leaf	entropy	same as left leaf
	Gini impurity	same as left leaf
	cl. error	same as left leaf
$\mathcal{T}_2$ left leaf	entropy	0
	Gini impurity	0
	cl. error	0
$\mathcal{T}_2$ right leaf	entropy	$-(1/3\log(1/3)+2/3\log(2/3))\approx 0.6365$
	Gini impurity	$1/3\cdot 2/3+1/3\cdot 2/3\approx 0.4444$
	cl. error	≈ 0.3333

The conditional entropy, weighted Gini impurity, and total classification error are shown below:

$\mathcal{T}_1$	entropy	0.562	Gini	0.375	cl.	0.25
$\mathcal{T}_2$	entropy	0.4774	Gini	0.3333	cl.	0.25

From these we see that $\mathcal{T}_2$ outperforms or at least is equal to $\mathcal{T}_1$ under all three metrics; that is, $\mathcal{T}_2$ appears to be a better split.
Also, in this case, assuming $\mathcal{T}_2$ is the better split, conditional entropy appears to be the most suitable choice as it gives the most definitive answer by having the largest difference between scores for $\mathcal{T}_1$ and $\mathcal{T}_2$.

Problem 2

First note that in the expression

$$\operatorname*{argmax}_{\pi_j,\mu_j,\Sigma_j}\sum_{i,j}\gamma_{i,j}\log\pi_j+\sum_{i,j}\gamma_{i,j}\log\mathcal{N}(x_i\mid\mu_j,\sigma_j),$$

π_j depends only on the first sum and μ_j,Σ_j only on the second. According to the hint,

$$\operatorname*{argmax}_{\pi_j}\sum_{i,j}\gamma_{i,j}\log\pi_j \qquad \text{is given by} \qquad \pi_j=\frac{\sum_i\gamma_{i,j}}{\sum_{i,j}\gamma_{i,j}}=\frac{\sum_i\gamma_{i,j}}{n} \tag{1}$$

since by definition

$$\sum_{i,j} \gamma_{i,j} = \sum_{i,j} \mathbb{P}(z_i = j \mid x_i) = \sum_i \sum_j \mathbb{P}(z_i = j \mid x_i) = \sum_i 1 = n.$$

Now, the log-likelihood $\log \mathcal{N}(x_i \mid \mu_j, \Sigma_j)$ is

$$\log \frac{1}{(2\pi)^{d/2}|\Sigma_j|^{1/2}} \exp(-(x_i - \mu_j)^T \Sigma_j^{-1}(x_i - \mu_j)/2) = \text{Constant} - \frac{\log|\Sigma_j|}{2} - \frac{(x_i - \mu_j)^T \Sigma_j^{-1}(x_i - \mu_j)}{2}.$$

Taking the derivative of $\sum_i \gamma_{i,j} \log \mathcal{N}(x_i \mid \mu_j, \Sigma_j)$ w.r.t. μ_j and setting it to 0 gives

$$\text{Constant} \cdot \sum_i \gamma_{i,j}(x_i - \mu_j) = 0 \implies \mu_j \sum_i \gamma_{i,j} = \sum_i \gamma_{i,j} x_i.$$

This implies

$$\underset{\mu_j}{\operatorname{argmax}} \sum_i \gamma_{i,j} \log \mathcal{N}(x_i \mid \mu_j, \Sigma_j) \qquad \text{is given by} \qquad \mu_j = \frac{\sum_i \gamma_{i,j} x_i}{\sum_i \gamma_{i,j}}. \tag{2}$$

Finally, the derivative of $\sum_i \gamma_{i,j} \log \mathcal{N}(x_i \mid \mu_j, \Sigma_j)$ w.r.t. Σ_j^{-1} is

$$\begin{aligned}
\frac{\partial}{\partial \Sigma_j^{-1}} \sum_i \gamma_{i,j} \log \mathcal{N}(x_i \mid \mu_j, \Sigma_j) &= \frac{\partial}{\partial \Sigma_j^{-1}} \left[-\sum_i \frac{\gamma_{i,j} \log|\Sigma_j|}{2} - \sum_i \frac{(x_i - \mu_j)^T (x_i - \mu_j) \Sigma^{-1}}{2} \right] \\
&\sim \frac{\partial}{\partial \Sigma_j^{-1}} \left[\sum_i \gamma_{i,j} \log|\Sigma_j^{-1}| - \sum_i (x_i - \mu_j)^T (x_i - \mu_j) \Sigma_j^{-1} \right] \\
&\sim \Sigma_j \sum_i \gamma_{i,j} - \sum_i (x_i - \mu_j)^T (x_i - \mu_j).
\end{aligned}$$

Setting this quantity to zero, we obtain $\Sigma_j = \sum_i \gamma_{i,j} \|x_i - \mu_j\| / \sum_i \gamma_{i,j}$; that is,

$$\underset{\Sigma_j}{\operatorname{argmax}} \sum_i \gamma_{i,j} \log \mathcal{N}(x_i \mid \mu_j, \Sigma_j) = \frac{1}{\sum_i \gamma_{i,j}} \sum_i \gamma_{i,j} \|x_i - \mu_j\|. \tag{3}$$

We can view k-means as a limiting result of GMM by letting $\sigma_j \to 0$. To see this, consider the $\gamma_{i,j}$ in GMM, where $\Sigma_j = \sigma_j^2 I$. Exponential decay implies

$$\lim_{\sigma_j \to 0} \gamma_{i,j} = \lim_{\sigma_j \to \infty} \frac{\pi_j \exp(-\|x_i - \mu_j\|^2/(2\sigma_j))}{\sum_k \pi_k \exp(-\|x_i - \mu_k\|^2/(2\sigma_k))} = \begin{cases} 1 & \text{if } j = \operatorname{argmin}_k \|x_i - \mu_k\|^2 \\ 0 & \text{otherwise.} \end{cases}$$