Tarea 03 - Reconocimiento de patrones
Giovanni Gamaliel López Padilla

Problema 1

Usemos la distancia de Kullback-Leibler para T-SNE. En el caso discreto, la definición de la distancia de Kullback-Leibler es:

$$d(P^1, P^2) = \sum_i P_i^1 \log\left(\frac{P_i^1}{P_i^2}\right)$$

- Calcula $d(P^1, P^2)$ si $P^1 \sim Bern(\theta_1)$ y $P^2 \sim Bern(\theta_2)$.
- Grafica $d(P^1, P^2)$ como función de θ_2 con θ_1 fija, y verifica que efectivamente mide de alguna manera la disimilitud entre P^1 y P^2 .

Se tiene que

$$P(\theta) = \theta^x(1-\theta)^{1-x}$$

Calculando $\log\left(\frac{P_i^1}{P_i^2}\right)$, se obtiene que:

$$\begin{aligned}
\log\left(\frac{P_i^1}{P_i^2}\right) &= \log\left(\frac{\theta_1^x(1-\theta_1)^{1-x}}{\theta_2^x(1-\theta_2)^{1-x}}\right) \\
&= \log\left(\theta_1^x(1-\theta_1)^{1-x}\right) - \log\left(\theta_2^x(1-\theta_2)^{1-x}\right) \\
&= x\log(\theta_1) + (1-x)\log(1-\theta_1) - x\log(\theta_2) - (1-x)\log(1-\theta_2) \\
&= x\left(\log(\theta_1) - \log(\theta_2)\right) + (1-x)\left(\log(1-\theta_1) - \log(1-\theta_2)\right)
\end{aligned}$$

Entonces, la distancia de Kullback-Leibler para dos distribuciones de Bernoulli es la siguiente:

$$\begin{aligned}
d(P^1, P^2) &= \sum_{i=0}^{1} \theta_1^i(1-\theta_1)^{1-i}\left[i\left(\log(\theta_1) - \log(\theta_2)\right) + (1-i)\left(\log(1-\theta_1) - \log(1-\theta_2)\right)\right] \\
d(P^1, P^2) &= (1-\theta_1)\left(\log(1-\theta_1) - \log(1-\theta_2)\right) + \theta_1\left(\log(\theta_1) - \log(\theta_2)\right)
\end{aligned}$$

Por lo tanto, la distancia de Kullback-Leibler para dos distribuciones de Bernoulli se puede calcular como en la ecuación 1.

$$d(P^1, P^2) = (1-\theta_1)\left(\log(1-\theta_1) - \log(1-\theta_2)\right) + \theta_1\left(\log(\theta_1) - \log(\theta_2)\right) \tag{1}$$

En la figura se visualiza la ecuación 1 para diferentes valores de θ_1. En esta se ve que la función alcanza un valor mínimo cuando θ_1 y θ_2 son muy semejantes. Por otro lado esta función tiende a infinito cuando los valores tienen una diferencia mayor.

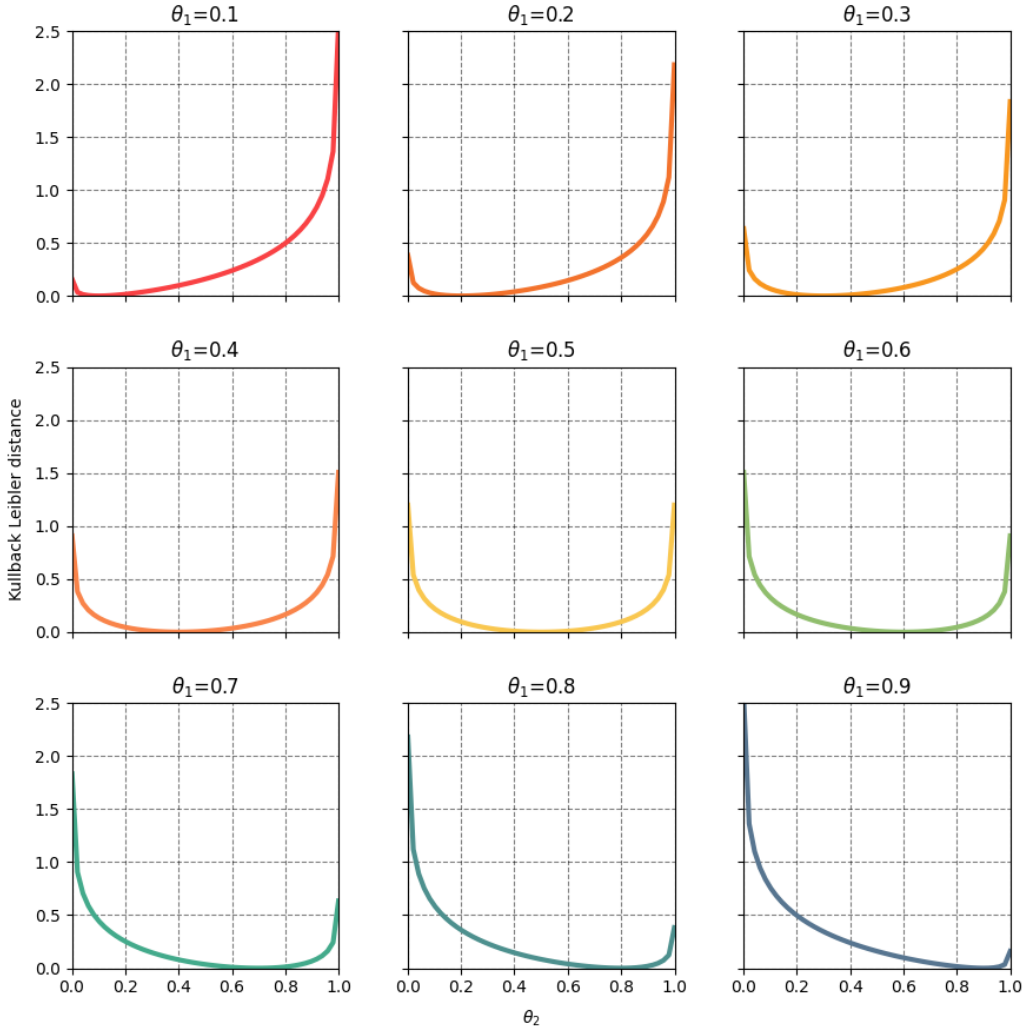
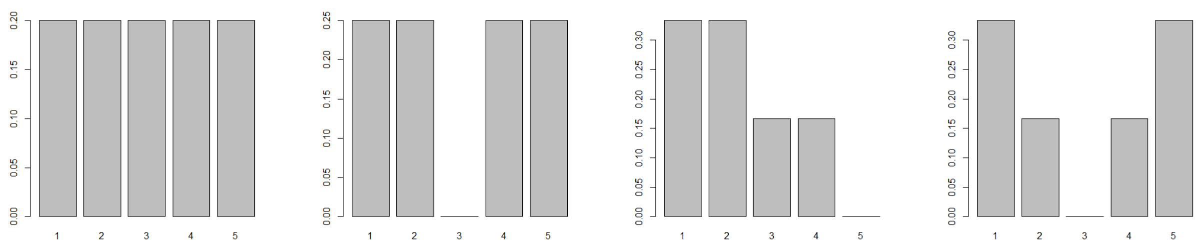


Figura 1: Resultados para diferentes valores de θ_1 en la ecuación 1.

Problema 02

Ordena las siguientes distribuciones en base de su entropía de Shannon (de menor a mayor). Motiva tu respuesta.



(a) Distribución 1 (b) Distribución 2 (c) Distribución 3 (d) Distribución 4

Figura 2: Diferentes distribuciones.

En la tabla 1 se encuentran los alores de las distribuciones de la figura 2.

Distribución	1	2	3	4	5
1	0.20	0.20	0.20	0.20	0.20
2	0.25	0.25	0	0.25	0.25
3	0.34	0.34	0.16	0.16	0
4	0.34	0.16	0	0.16	0.34

Tabla 1: Valores de las distribuciones mostradas en la figura 2

Basandonos en los datos, se puede decir que la distribución con mayor entropía es la distribución 1. Esto debido a que todos sus valores contiene la misma probabilidad. Otra afirmación que se puede realizar es que las distribuciones 3 y 4 tienen la misma entropía debido a que tienen una semejanza en su distribución de probabilidades. El cambio es que probabilidad esta asociada a cada valor. Con la información de la tabla 1 se puede calcular la entropía de Shannon para comprobar lo antes dicho. Los resultados se muestran en la tabla 2.

Distribución	1	2	3	4	5	Entropía
1	0.20	0.20	0.20	0.20	0.20	0.6989
2	0.25	0.25	0	0.25	0.25	0.6026
3	0.34	0.34	0.16	0.16	0	0.5732
4	0.34	0.16	0	0.16	0.34	0.5732

Tabla 2: Valores de las distribuciones mostradas en la figura 2 con sus respectivos valores de entropía.

Con esto podemos afirmar que el orden de menor a mayo entropía es distribución 3 y 4, después la distribución 2 y por último la distribución 1.

Problema 03

Muestra que el incremento $\mathbf{I_{ij}}$ en la suma de los errores al cuadrado, sum of squared errors (SSE), del método de mínima varianza de Ward (Ward's minimum variance method), i.e.,

$$
\begin{aligned}
I_{ij} &= SSE_{ij} - (SSE_i + SSE_j) \\
SSE_i &= \sum_{x \in C_i} ||x_i - \mu_i||^2 \\
SSE_j &= \sum_{x \in C_i} ||x_i - \mu_j||^2 \\
SSE_{ij} &= \sum_{x \in C_i} ||x_i - \mu_{ij}||^2
\end{aligned}
$$

donde $\$mu_i$, μ_j y μ_{ij} son los centroides de los clusters C_i , C_j y $C_i \cup Cj$ respectivamente, se puede escribir como sigue.

$$I_{ij} = \frac{|C_i||C_j|}{|C_i| + |C_j|}||\mu_i - \mu_j||^2$$

donde $|C_i|$ y $|C_j|$ representan la cardinalidad de los clusters C_i y C_j respectivamente.
Calculando SSE_{ij} se obtiene lo siguiente:

$$\begin{aligned}
SSE_{ij} &= \sum_{i \in C_{ij}} ||x_i - \mu_{ij}||^2 \\
&= \sum_{i \in C_i} ||x_i - \mu_{ij}||^2 + \sum_{i \in C_j} ||x_i - \mu_{ij}||^2
\end{aligned}$$

Calculando $\sum_{i \in C_i} ||x_i - \mu_{ij}||^2$ se obtiene lo siguiente:

$$\begin{aligned}
\sum_{i \in C_i} ||x_i - \mu_{ij}||^2 &= \sum_{i \in C_i} ||x_i - \left(\mu_i - \frac{|C_j|}{|C_i| + |C_j|}(\mu_i - \mu_j)\right)||^2 \\
&= \sum_{i \in C_i} ||(x_i - \mu_i) + \frac{|C_j|}{|C_i| + |C_j|}(\mu_i - \mu_j)||^2 \\
&= \sum_{i \in C_i} ||x_i - \mu_i||^2 + 2\sum_{i \in C_i}(x_i - \mu_i)\left(\frac{|C_j|}{|C_i| + |C_j|}\right)(\mu_i - \mu_j) + \sum_{i \in C_i}\left(\frac{|C_j|}{|C_i| + |C_j|}\right)||\mu_i - \mu_j||^2 \\
&= \sum_{i \in C_i} ||x_i - \mu_i||^2 + |C_i|\left(\frac{|C_j|^2}{(|C_i| + |C_j|)^2}\right)||\mu_i - \mu_j||^2
\end{aligned}$$

De forma semejante se puede calcular $\sum_{i \in C_j} ||x_i - \mu_{ij}||^2$:

$$\begin{aligned}
\sum_{i \in C_j} ||x_i - \mu_{ij}||^2 &= \sum_{i \in C_j} ||x_i - \left(\mu_j - \frac{|C_i|}{|C_i| + |C_j|}(\mu_i - \mu_j)\right)||^2 \\
&= \sum_{i \in C_j} ||(x_i - \mu_j) + \frac{|C_i|}{|C_i| + |C_j|}(\mu_i - \mu_j)||^2 \\
&= \sum_{i \in C_j} ||x_i - \mu_j||^2 + 2\sum_{i \in C_j}(x_i - \mu_j)\left(\frac{|C_i|}{|C_i| + |C_j|}\right)(\mu_i - \mu_j) + \sum_{i \in C_j}\left(\frac{|C_i|}{|C_i| + |C_j|}\right)||\mu_i - \mu_j||^2 \\
&= \sum_{i \in C_j} ||x_i - \mu_j||^2 + |C_j|\left(\frac{|C_i|^2}{(|C_i| + |C_j|)^2}\right)||\mu_j - \mu_j||^2
\end{aligned}$$

Entonces SSE_{ij} puede ser escrito de la siguiente manera:

$$
\begin{aligned}
SSE_{ij} &= \sum_{x \in C_i} ||x_i - \mu_{ij}||^2 \\
&= \sum_{i \in C_i} ||x_i - \mu_{ij}||^2 + \sum_{i \in C_j} ||x_i - \mu_{ij}||^2 \\
&= \sum_{i \in C_i} ||x_i - \mu_i||^2 + \sum_{i \in C_j} ||x_i - \mu_j||^2 + \frac{|C_j||C_i|^2 + |C_j|^2|C_i|}{(|C_i| + |C_j|)^2} ||\mu_i - \mu_j||^2 \\
&= SSE_i + SSE_j + \frac{|C_j||C_i|}{|C_i| + |C_j|} ||\mu_i - \mu_j||^2
\end{aligned}
$$

por lo tanto, I_{ij} es igual a:

$$
\begin{aligned}
I_{ij} &= SSE_{ij} - (SSE_i + SSE_j) \\
&= SSE_{ij} - SSE_i - SSE_j \\
&= SSE_i + SSE_J + \frac{|C_j||C_i|}{|C_i| + |C_j|} ||\mu_i - \mu_j||^2 - SSE_i - SSE_j \\
I_{ij} &= \frac{|C_j||C_i|}{|C_i| + |C_j|} ||\mu_i - \mu_j||^2
\end{aligned}
$$