

Tarea 8: Ajuste de Mezclas Gaussianas

Optimización I

Maestría en Computación

Centro de Investigación en Matemáticas

Esteban Reyes Saldaña
esteban.reyes@citmat.mx

19 de abril de 2021

Resumen

En esta tarea se realizó el ajuste de muestras Gaussianas de un histograma mediante la minimización con descenso gradiente. Dada la forma de la función a optimizar no se utilizaron métodos que necesitaran el Hessiano, en su lugar se usó búsqueda en línea con backtracking. Se presenta a continuación una descripción general, así como el pseudocódigo de los métodos implementados. Finalmente se incluyen conclusiones observadas a partir de la experimentación.

1 Introducción

1.1 Descenso Gradiente

El objetivo general es encontrar los mínimos locales de una función $f : \mathbb{R}^n \rightarrow \mathbb{R}$ tal que $f(x) \in \mathcal{C}^1$ de manera iterativa. Un resultado muy conocido sobre el gradiente es

Teorema 1.1. Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una función continuamente diferenciable. Entonces la dirección donde $f(x)$ crece más rápido es $\nabla f(x)$.

Corolario 1.1. Bajo las condiciones del Teorema (1.1), $f(x)$ decrece más rápido en la dirección $-\nabla f(x)$.

Definición 1.1. Una **dirección de descenso** $d \in \mathbb{R}^n$ para $f \in \mathcal{C}^1$ es un vector tal que

$$f(x + td) < f(x)$$

para $t \in (0, T)$. Es decir, permite que el punto x más cerca al mínimo local x^* de la función objetivo $f : \mathbb{R}^n \rightarrow \mathbb{R}$.

Teorema 1.2. Si $g(x)^T d < 0$ entonces d es una dirección de descenso.

1.2 Búsqueda en línea

Este método usa la dirección $d_k = -g(x_k)$ para moverse en cada iteración. El esquema general es

1. Dé un vector inicial x_0 .
2. Haga $k = 0$ y $g_0 = \nabla f(x_0)$
3. Mientras $\|g_k\| \neq 0$,
 - (a) $\alpha_k = \arg \min_{\alpha > 0} f(x_k - \alpha g_k)$
 - (b) $x_{k+1} = x_k - \alpha_k g_k$
 - (c) $g_{k+1} = \nabla f(x_{k+1})$
 - (d) $k \rightarrow k + 1$

Se debe seleccionar un tamaño de paso α_k tal que reduce suficientemente a $f(x)$ y al mismo tiempo se requiere que sea eficiente.

1.3 Condición de Armijo

Un tamaño de decrecimiento suficiente se puede medir por la siguiente ecuación

$$f(x_k + \alpha d_k) \leq f(x_k) + c_1 \alpha \nabla f_k^T d_k,$$

para alguna constante $c_1 \in (0, 1)$.

2 Método

2.1 Descripción del Problema

Se utilizó el método de descenso de gradiente con búsqueda en línea **backtracking** para resolver el problema de ajustar una mezcla de gaussianas a un histograma 3D. La función objetivo de este problema viene dada por $\min_{\alpha^j, \mu^j} g(\alpha^j, \mu^j)$ donde

$$g(\alpha^j, \mu^j) = \sum_{c \in \Omega} \left[h^j(c) - \sum_{i=1}^n \alpha_i^j \exp \left(-\frac{\|c - \mu_i^j\|_2^2}{2\sigma^2} \right) \right], \quad (1)$$

donde

- $j \in \{1, 2\}$ corresponden a fondo y objeto.
- $c = [r, g, b]^T$, con $r \in B_1 = \{1, \dots, b_1\}$, $g \in B_2 = \{1, \dots, b_2\}$, $b \in B_3 = \{1, \dots, b_3\}$ y b_1, b_2, b_3 son el número de bins en cada canal *RGB*.
- $\Omega = B_1 \times B_2 \times B_3$, donde $X \times Y := \{(x, y) : x \in X, y \in Y\}$ representa el producto cartesiano.
- $h^j(r, g, b)$ es el histograma 3D de la clase j .
- n es el número de elementos en la base radial.
- $\alpha^j = [\alpha_1^j, \alpha_2^j, \dots, \alpha_n^j]$ son los pesos de la combinación lineal de la base radial.
- $\mu^j = [\mu_1^j, \mu_2^j, \dots, \mu_n^j]$ son las medias o posiciones de los elementos de la base radial.
- σ es un parámetro conocido.

El problema (1) se puede resolver de manera alternada hasta convergencia, es decir,

$$\begin{aligned} \alpha_{k+1}^j &= \arg \min_{\alpha^j \in \mathbb{R}^n} g(\alpha^j, \mu_k^j) \\ \mu_{k+1}^j &= \arg \min_{\mu^j \in \mathbb{R}^n} g(\alpha_{k+1}^j, \mu^j) \end{aligned}$$

Para obtener los histogramas de clase se utilizó el programa *histograma_3D_class.py* proporcionado por el Profesor.

2.2 Gradiente de la función Objetivo

Para implementar descenso con backtracking de manera iterativa se necesita calcular los gradientes de $g(\alpha^j, \mu^j)$ respecto a α^j y μ^j . Para manejar fácilmente las expresiones, sea

$$w_k^j = \frac{\|c - \mu_k^j\|_2^2}{2\sigma^2}$$

2.2.1 Gradiente respecto a α^j

$$\begin{aligned} \frac{\partial}{\partial \alpha_k^j} g(\alpha^j, \mu^j) &= 2 \sum_{c \in \Omega} \left[h^j(c) - \sum_{i=1}^n \alpha_i^j \exp(w_i^j) \right] \\ &\quad \left[-\frac{\partial}{\partial \alpha_k^j} \sum_{i=1}^m \alpha_i^j \exp(-w_i^j) \right] \\ &= -2 \sum_{c \in \Omega} \left[h^j(c) - \sum_{i=1}^n \alpha_i^j \exp(-w_i^j) \right] \\ &\quad \exp(-w_k^j) \end{aligned}$$

de donde

$$\frac{\partial}{\partial \alpha_k^j} g(\alpha^j, \mu^j) = -2 \exp(-w_k^j) \sum_{c \in \Omega} \left[h^j(c) - \sum_{i=1}^n \alpha_i^j \exp(-w_i^j) \right] \quad (2)$$

2.2.2 Gradiente respecto a μ^j

$$\begin{aligned} \frac{\partial}{\partial \mu_k^j} g(\alpha^j, \mu^j) &= 2 \sum_{c \in \Omega} \left[h^j(c) - \sum_{i=1}^n \alpha_i^j \exp(w_i^j) \right] \\ &\quad \left[-\frac{\partial}{\partial \mu_k^j} \sum_{i=1}^m \alpha_i^j \exp(-w_i^j) \right] \end{aligned}$$

de donde

$$\frac{\partial}{\partial \mu_k^j} g(\alpha^j, \mu^j) = -\frac{2}{\sigma^2} \sum_{c \in \Omega} \left[h^j(c) - \sum_{i=1}^n \alpha_i^j \exp(-w_i^j) \right] \left[\alpha_k^j \exp(w_k^j) (c - m_k^j) \right]$$

2.3 Pseudocódigo

2.3.1 Búsqueda con BackTracking

Input: $\hat{\alpha}, \rho \in (0, 1), c_1 \in (0, 1)$

Output: Tamaño de paso α_k

- 1: Haga $\alpha = \hat{\alpha}$
- 2: $inum = 0$
- 3: **while** $f(x_k + \alpha d_k) > f(x_k) + c_1 \alpha \nabla f_k^T d_k$ **do**
- 4: $\alpha \rightarrow \rho \alpha$
- 5: **end while**
- 6: Regresa $\alpha_k = \alpha$

3 Resultados

Para el método de **descenso gradiente** con búsqueda en línea se usaron los parámetros

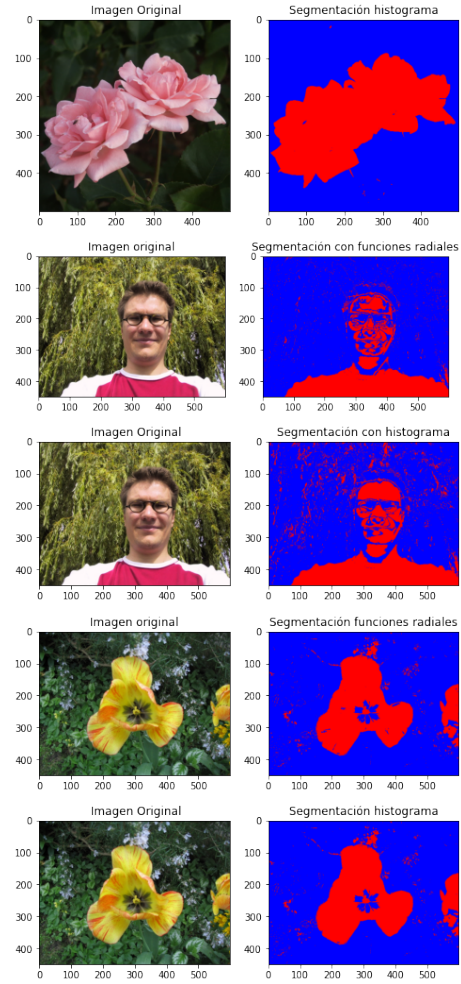
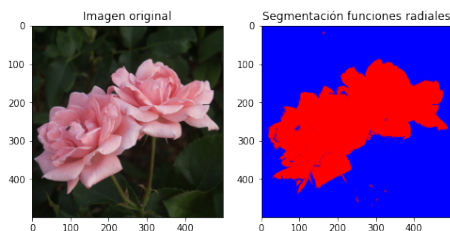
Parámetro	Valor
α	0.9
ρ	0.5
c_1	10^{-4}

Parámetro	max_{iter}	τ_x	τ_f	τ_{grad}
Valor	100	10^{-8}	10^{-8}	10^{-8}

Los valores para los vectores α^j y μ^j se generaron de manera aleatoria uniforme sobre $[0, 1]$. Para realizar el ajuste gaussiano se eligieron

Parámetro	$nbins$	$radial_{size}$	σ	max_{iter}
Valor	3	10	0,5	500

Se probó con tres imágenes del conjunto sugerido. A continuación se muestra la segmentación obtenida.



4 Conclusiones

El problema de segmentación con bases Gaussianas se pudo resolver como un problema de optimización

iterativa. De la experimentación de tareas previas se sabía que, comparado con algoritmos que utilizan el Hessiano de una función dada, suelen dar mejores resultados en cuando a tiempo de ejecución. Por lo anterior se le dieron solo 100 iteraciones al descenso gradiente.

Se observó también que la convergencia depende fuertemente de los valores iniciales para α^j , μ^j y σ .

Además, la convergencia mejoró respecto al tiempo de ejecución para valores de σ cada vez más pequeños. Esto refleja la importancia de los valores w_k^j , es decir, entre más pequeño el valor de σ , más grande es la aportación de w_k^j . Lo anterior también se refleja en los gradientes, por ejemplo, para el gradiente con respecto a μ^j , entre más pequeño es σ , más grande es la norma del gradiente.