

1 Diagrama de Voronói y Clustering

Datos de entrada: En “Personas_de_villa_laminera.txt” se muestran dos variables de estado, $X_1 = \text{“nivel de estrés”}$ y $X_2 = \text{“afición a los dulces”}$, para un conjunto de $A = \{a_i, \{X_j\}_{j=1}^2\}_{i=1}^{1500}$ de 1500 personas de la población ficticia “Villa Laminera”.

Determina el número ideal de franjas de Villa Laminera (sistema A) a partir del número óptimo de *clusters* o vecindades de Voronói. Para ello, utiliza el coeficiente de Silhouette (\bar{s}), que **puede emplearse directamente desde la librería sklearn**:

- i) Obtén el coeficiente \bar{s} de A para diferente número de vecindades $k \in \{2, 3, \dots, 15\}$ usando el algoritmo KMeans. Muestra en una gráfica el valor de \bar{s} en función de k y decide con ello cuál es el número óptimo de vecindades. En una segunda gráfica, muestra la clasificación (*clusters*) resultante con diferentes colores y representa el *diagrama de Voronói* en esa misma gráfica.
- ii) Obtén el coeficiente \bar{s} para el mismo sistema A usando ahora el algoritmo DBSCAN con la métrica ‘euclidean’ y luego con ‘manhattan’. En este caso, el parámetro que debemos explorar es el *umbral de distancia* $\epsilon \in (0.1, 0.4)$, fijando el número de elementos mínimo en $n_0 = 10$. Compara gráficamente ambos resultados.
- iii) ¿De qué franja del plan diríamos que nos separa con coordenadas $a := (1/2, 0)$ y $b := (0, -3)$? Comprueba tu respuesta con la función `kmeans.predict`.