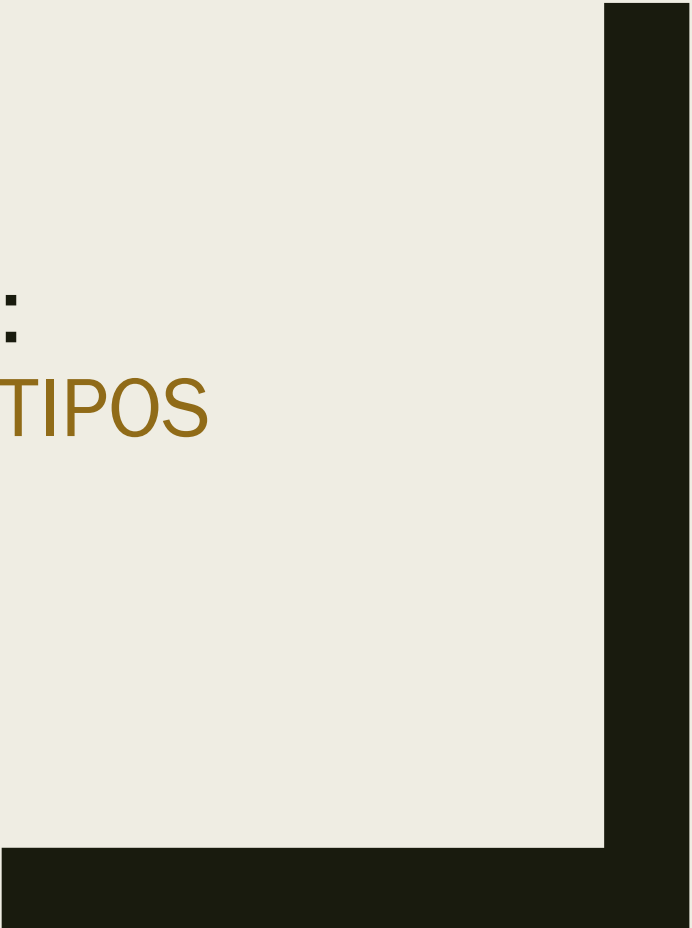




ANALÍTICA AVANZADA DE DATOS: CLASIFICADORES BASADOS EN MÉTRICAS

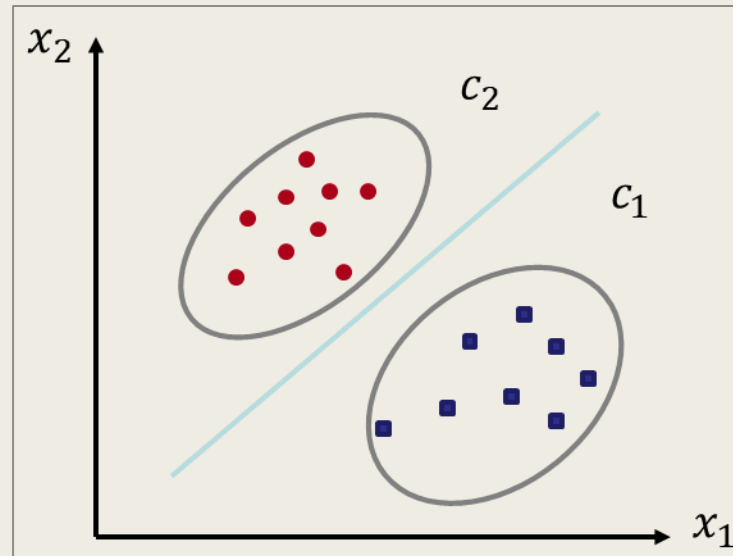
A. Alejandra Sánchez Manilla
asanchezm.q@gmail.com



Clasificadores basados en métricas

Clasificador Euclidiano

- En la tarea de clasificación la meta principal es dividir el espacio de características en regiones asignadas a cada clase a clasificar.
- Si un vector de características cae en una de estas regiones, entonces es asignada a la clase correspondiente de dicha región



Clasificador Euclidiano

- Para poder clasificar los patrones es necesario realizar una comparación entre ellos o con un estándar.
- Cuando se va a clasificar un nuevo patrón, la proximidad o similitud de este con los patrones en el conjunto de entrenamiento debe ser encontrada.
- Existen varias medidas de distancia o similitud que pueden ser utilizadas, siendo una de las más sencilla la **distancia euclidiana**.

Clasificador Euclidiano

- El clasificador Euclidiano se basa en la métrica de la distancia euclidiana para realizar la clasificación
- Cada clase es representada por un patrón, el cual se calcula como el vector medio de cada clase.
- Cuando se va a clasificar un nuevo patrón se calcula la distancia euclidiana con respecto al vector medio de cada clase.
- El nuevo vector es asignado a la clase del vector medio más cercano.

Clasificadores basados en métricas

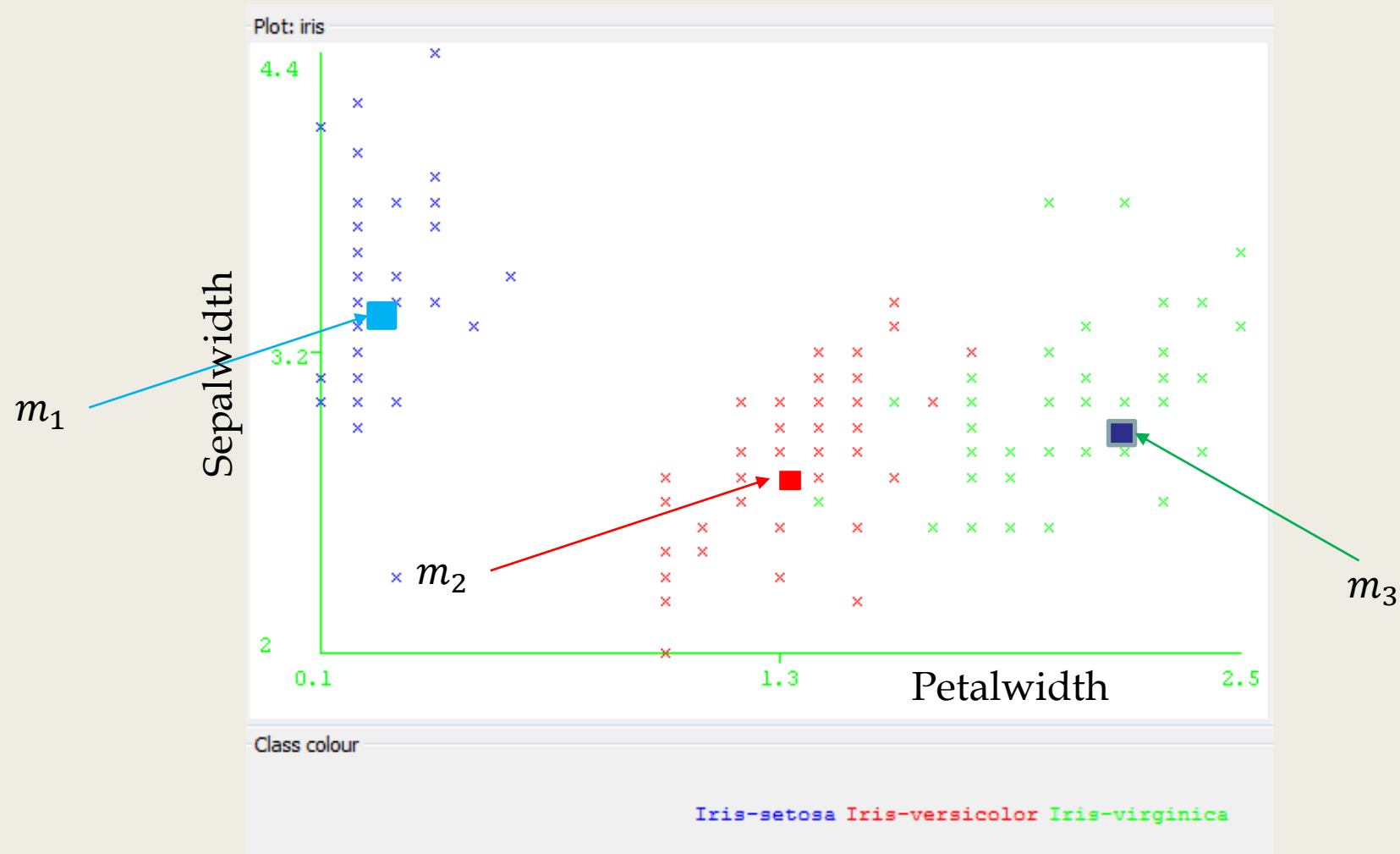
Clasificador Euclidiano (Fase de aprendizaje)

Se asume que se particionó el Banco de Datos X (empleando alguno de los métodos de validación en R.P.) en los conjuntos E y P , donde $E \cap P = \emptyset$ y $E \cup P = X$.

- For $k = 1$ hasta c
- Calcular $m_k^i = \frac{\sum_{j=1}^{N_E(\omega_k)} x_j^i}{N_E(\omega_k)}$ donde $N_E(\omega_k)$ es el número de patrones de la clase k perteneciente a E , x_j^i es la i – ésima componente del patrón j .
- End

Clasificadores basados en métricas

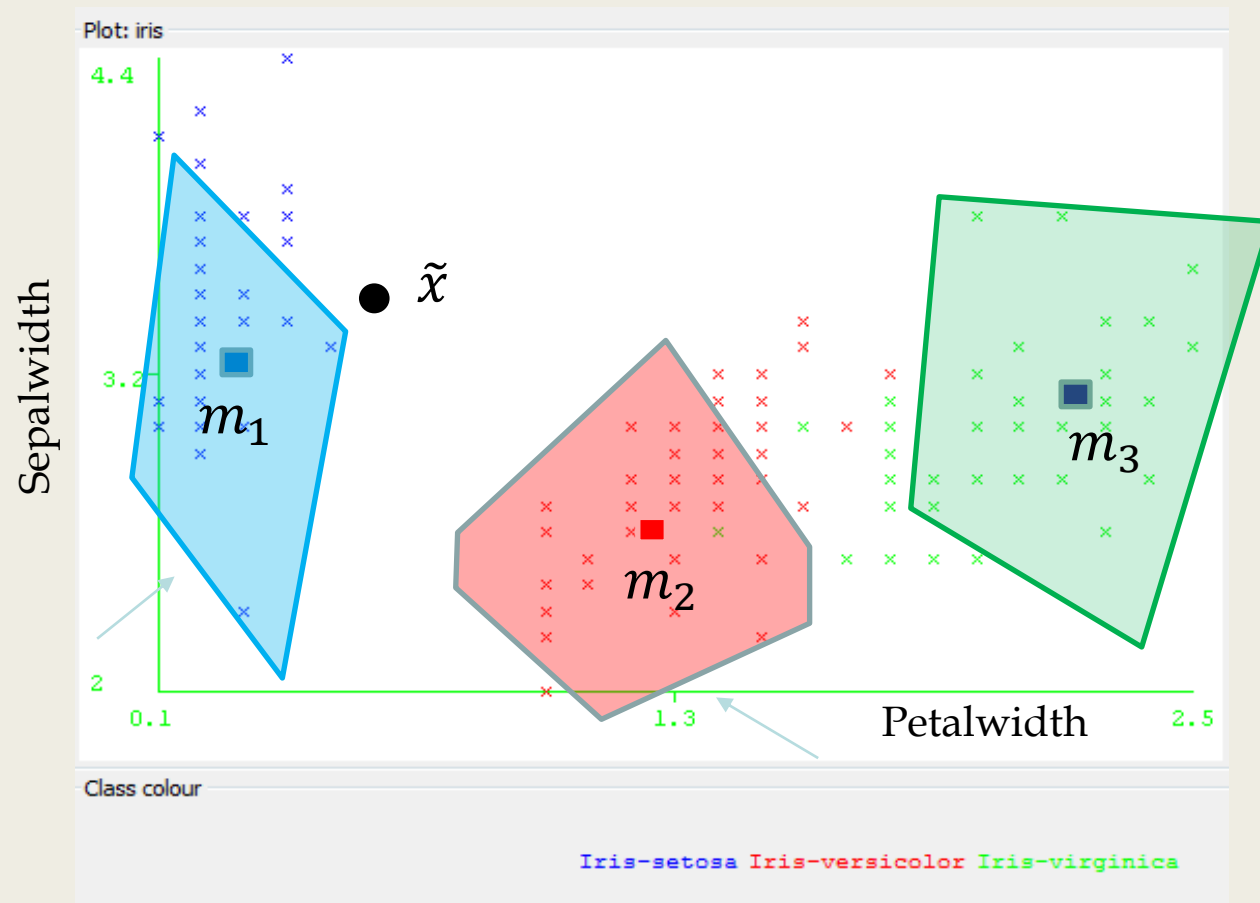
Clasificador Euclidiano



Clasificadores basados en métricas

Clasificador Euclidiano (Fase de clasificación)

1. Sea $\tilde{x} \in P$ un patrón cuya clase se desconoce.



Clasificadores basados en métricas

Clasificador Euclidiano (Fase de clasificación)

2. Sea \tilde{x} calcula la distancia entre \tilde{x} y cada uno de los m_k , la distancia la representaremos por d_k .

La distancia euclidiana entre 2 puntos en un espacio euclidiano de n dimensiones se calcula usando la siguiente ecuación:

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Clasificadores basados en métricas

Clasificador Euclidiano (Fase de clasificación)

3. Se busca el menor valor entre las distancias d_k .
4. \tilde{x} se asigna a la clase ω_k que corresponde con la d_k de menor valor.
5. Se verifica que la clase asignada a \tilde{x} sea la correcta.

Referencias

- [1] Leondes, C.T. (2018). *Image Processing and Pattern Recognition*. California: Academic Press.
- [2] Duda, R.O., Hart, P.E. & Stork, D.G. (2001). *Pattern Classification*. 2nd edition. Wiley-Interscience.
- [3] Marques de Sá, J:P. (2001). *Pattern Recognition: Concepts, Methods and Applications*. Berlin: Springer-Verlag.
- [4] Kuncheva, L. (2014). *Combining Pattern Classifiers: Methods and Algorithms*. 2nd edition. USA: Wiley.
- [5] Witten, I.H., Frank, E. & Hall, M.A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*. 3rd edition. USA: Elsevier.
- [6] Murty, N.M. & Devi, V.S. (2011). *Pattern Recognition: An Algorithmic Approach*. Springer.
- [7] Zaki, M.J. & Meira, W. (2014). *Data Mining and Analysis: Fundamental Concepts and Algorithms*. Cambridge University Press.
- [8] Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H. & Bing, G. (2017). Learning from class-imbalanced data: Review of methods and applications. *Expert Systems With Applications*, 73, 220-239.