Práctica 2. Clasificador Bayesiano

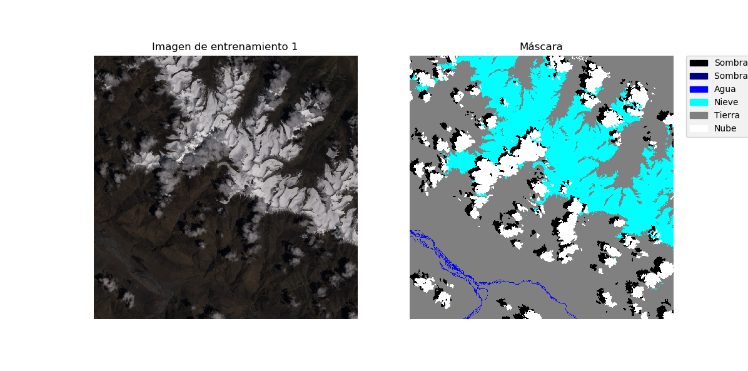
Andrés González Flores, Facultad de Ingeniería, UNAM.

—————————— ◆ ——————————

# 1 Objetivos

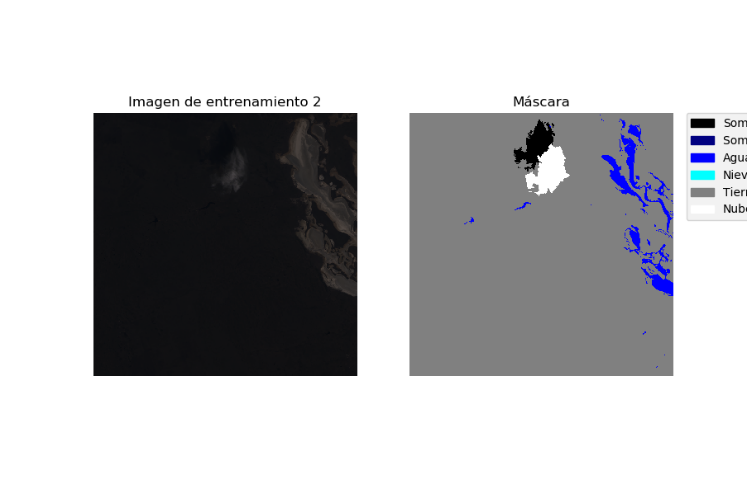
Clasificar imágenes con 2, 3 o 4 regiones utilizando el clasificador de Bayes

# 2 Introducción

El clasificador Bayesiano es un método de clasificación supervisada que consiste en asignar a un objeto descrito por un conjunto de atributos o características, , a una de *m* clases posibles, , tal que la probabilidad de la clase dados los atributos se maximiza:



La formulación de este clasificador se basa en utilizar la regla de Bayes para calcular la probabilidad a posteriori de la clase dados los atributos



Dado que el denominador no varía para las diferentes clases, se puede considerar como una constante si lo que interesa es maximizar la probabilidad de la clase.



El clasificador Bayesiano ingenuo (naive Bayes) se basa en la suposición de que todos los atributos son independientes dada la clase, esto es, cada atributo  es condicionalmente independiente de los demás atributos dada la clase

Bajo estas consideraciones, la ecuación 2 se puede escribir como:

Donde . se puede considerar como una constante de normalización.



# 3 Desarrollo

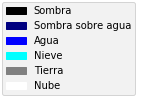
1. Seleccionar imágenes con zonas aledañas a clasificar.

Escogí trabajar con los datos de validación de SPARCS de M. Joseph Huges de la Universidad del Estado de Oregon. Estos datos se usaron para la detección automática de nubes usando redes neuronales.

Los datos consisten en 80 subconjuntos de escenas del

Landsat 8 de 1000x1000px en formato .tiff. Cada escena tiene 10 bandas. Estos datos ya venían con sus respectivas máscaras de segmentación. Las máscaras venían segmentadas en 6 clases que se interpretan como sigue:

Valor / Interpretación

0 Sombra

1 Sombra Sobre Agua

2 Agua

3 Nieve

4 Tierra

5 Nube

6 Inundado

Para que la carga de cómputo no resultara muy pesada, seleccioné sólo 5 bandas (de la 1 a la 5) de 3 escenas. Esto resultó en datos de tamaño 3x1000x1000x5 (un total de 15,000,000‬ de datos simples).

Las imágenes RGB se encuentran en las bandas 2-3, siendo B la banda 2, G la banda 3 y R la banda 4. A continuación, muestro las imágenes de entrenamiento.

Con los datos de las 5 bandas de cada escena, calculé las probabilidades a priori de cada una de las regiones. Para estas imágenes, sólo tomé en cuenta las clases 0 a 5.



Luego, calculé los vectores de medias de cada clase.



Luego, se calcularon las matrices de covarianzas Σk con la siguiente fórmula.



# 4 Resultados

IEEE

# 5 Código

# 6 Conclusiones

**Referencias**

1. U.S. Geological Survey, 2016. L8 SPARCS Cloud Validation Masks. U.S. Geological Survey data release. doi:10.5066/F7FB5146.
2. Hughes M.J. & Hayes, D.J. (2014). Automated detection of cloud and cloud shadow in single-date Landsat imagery using neural networks and spatial post-processing. Remote Sensing, 6(6), 4907–4926. doi:10.3390/rs6064907.W.-K. Chen, *Linear Networks and Systems.* Belmont, Calif.: Wadsworth, pp. 123-135, 1993. (Book style)
3. Sucar, Luis Enrique. *Clasificadores Bayesianos: de Datos a Conceptos.*