

Herramientas de Teledetección Cuantitativa

Clase 6

Francisco Nemiña



Esquema de presentación

Introducción

Nociones básicas

Matriz de confusión

Índice kappa

Muestreo

Práctica

Práctica

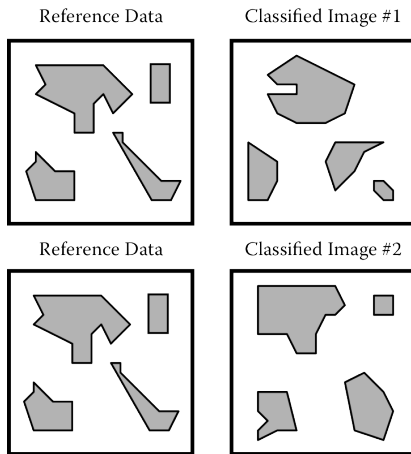


Objetivo de la validación

Lo que esperamos es asignarle a nuestro mapa temático un cierto grado de confianza a partir de datos medidos en el terreno.



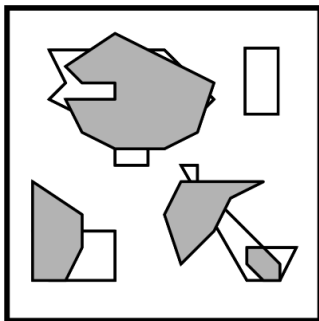
Nociones básicas



Ejemplo de datos de referencia contra un mapa temático.¹

¹Russell G Congalton y Kass Green. *Assessing the accuracy of remotely sensed data: principles and practices*. CRC press, 2008.

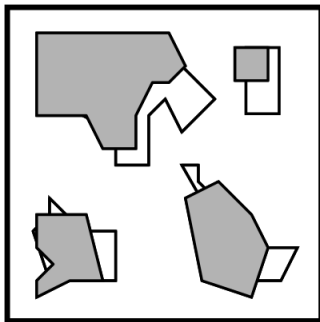




Comparación de área total.²

²Russell G Congalton y Kass Green. *Assessing the accuracy of remotely sensed data: principles and practices*. CRC press, 2008.





Comparacion de espacial.³

³Russell G Congalton y Kass Green. *Assessing the accuracy of remotely sensed data: principles and practices*. CRC press, 2008.



Definición

Lo que esperamos es asignarle a nuestro mapa temático un cierto grado de precisión a partir de datos medidos en el terreno.



Matriz de confusión

Definición

$$\begin{bmatrix} & 1 & 2 & \dots & k & n_{i+} \\ 1 & n_{11} & n_{12} & \dots & n_{1k} & n_{1+} \\ 2 & n_{21} & n_{22} & \dots & n_{2k} & n_{2+} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots \\ k & n_{k1} & n_{k2} & \dots & n_{kk} & n_{k+} \\ n_{+j} & n_{+1} & n_{+2} & \dots & n_{+k} & N \end{bmatrix}$$



Definición

Donde

$$n_{i+} = \sum_j n_{ij}$$

$$n_{+j} = \sum_i n_{ij}$$

y donde n es el número total de muestras.



Ejemplo

Vamos a tomar sólo tres coberturas a modo de ejemplo

	A	S	V	
A	50	10	20	80
S	5	100	15	120
V	10	10	80	100
	65	120	115	300



Matriz de confusión

Presición total

$$\frac{\sum_i n_{ii}}{n}$$

Presición usuario

$$\frac{n_{ii}}{n_{i+}}$$

Presición productor

$$\frac{n_{jj}}{n_{+j}}$$



Matriz de confusión

Fracción de la muestra

$$p_{ij} = \frac{n_{ij}}{n}$$

Probabilidad de j en los datos de campo

$$p_{+j} = \sum_i p_{ij}$$

Probabilidad de i en la clasificación

$$p_{i+} = \sum_j p_{ij}$$



Ejemplo

	A	S	V	
A	0,17	0,03	0,07	0,625
S	0,02	0,33	0,05	0,833
V	0,03	0,03	0,27	0,800
	0,769	0,833	0,696	0,767



Matriz de confusión

Matriz de confusión

Cualquier análisis sobre el error de una clasificación parte de la matriz de confusión.



Definición

El índice kappa nos permite estimar si dos matrices de confusión son distintas una de la otra o no.

Mide cuanto se acerca mi clasificación a una clasificación al azar.



Definición

$$\hat{K} = \frac{p_0 - p_c}{1 - p_c}$$

donde

$$p_0 = \sum_i p_{ii}$$

y

$$p_c = \sum_i p_{i+} p_{+i}$$



Ejemplo

En este caso, $p_0 = 0,77$ y $p_c = 0,35$ entonces el índice kappa nos queda

$$\hat{K} = \frac{0,77 - 0,35}{1 - 0,35} = 0,64$$

Ahora hay que interpretar esto.



Índice kappa

Interpretación

Interpretaciones del índice kappa hay muchas. Lo mas básico es que cuanto más cerca de cero este el valor, más se parece la clasificacion una clasificación aleatoria.

Cortes

0.0 - 0.4	0.4 - 0.8	0.8 - 1.0
Pobre	Medio	Bueno

Rangos de acuerdo del índice kappa⁴

⁴ J Richard Landis y Gary G Koch. "The measurement of observer agreement for categorical data". En: *biometrics* (1977), págs. 159-174.



4 preguntas

1. ¿Qué categorías tengo?
2. ¿Qué unidad de muestreo usar?
3. ¿Cuántas muestras tomar?
4. ¿Cómo elegir las muestras?



¿Que categorías tengo?

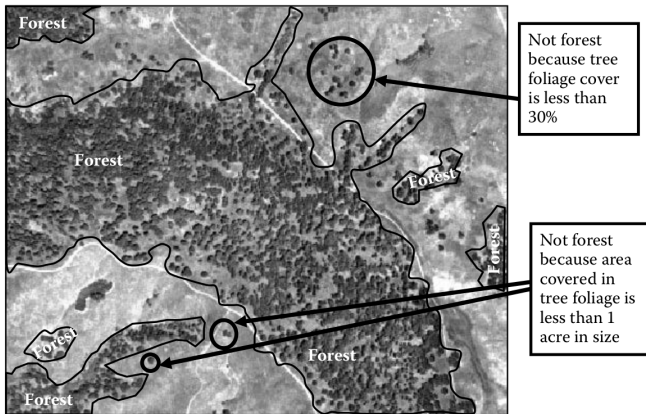
Las clases tienen que ser

- ▶ Mutualmente exclusivas
- ▶ Totalmente exhaustivas

Además de tener un tamaño mínimo para ser considerado de esa clase.



Muestreo



Clases de muestreo definidas en el terreno.⁵

⁵Russell G Congalton y Kass Green. *Assessing the accuracy of remotely sensed data: principles and practices*. CRC press, 2008.



¿Qué unidad de muestreo usar?

- ▶ Un solo píxel.
- ▶ Un clúster de píxeles
- ▶ Un polígono
- ▶ Un clúster de polígonos



¿Cuántas muestras tomar?

$$N = \frac{B}{4b^2}$$

donde B se obtiene a partir de la distribución χ^2 con un grado de libertad y b es la precisión que uno acepta.



¿Como elegir las muestras?

- ▶ Al azar.
- ▶ Estratificado al azar.
- ▶ Sistemático.
- ▶ Clusters



Logística

Todo lo que vimos va a estar supeditado a mi capacidad de realizar el muestreo.



Esquema de presentación

Introducción

Nociones básicas

Matriz de confusión

Índice kappa

Muestreo

Práctica

Práctica



Actividades prácticas de la sexta clase

1. Abrir las imágenes clasificadas y fusionadas por el método de clasificación supervisada y no supervisada.
2. Cargar los polígonos de validación correspondientes a cada clase.
3. Calcular la matriz de confusión correspondiente a cada clasificación.
4. Obtener la precisión global, del usuario, productor y el índice kappa.



Esquema de presentación

Introducción

Nociones básicas

Matriz de confusión

Índice kappa

Muestreo

Práctica

Práctica



Actividades prácticas de la sexta clase

1. Abrir las imágenes clasificadas y fusionadas por el método de clasificación supervisada y no supervisada.
2. Cargar los polígonos de validación correspondientes a cada clase.
3. Calcular la matriz de confusión correspondiente a cada clasificación.
4. Obtener la precisión global, del usuario, productor y el índice kappa.

