

Clasificación supervisada

Utilizando diferentes algoritmos las técnicas de clasificación **supervisada** comparan la similitud que tiene cada píxel de una escena con respecto a los parámetros de subconjuntos de píxeles que se consideran representativos de los elementos a identificar. A diferencia del resultado que se obtiene al aplicar los algoritmos de agrupamiento (clasificación no supervisada), en esta ocasión el resultado proporciona “categorías”, ya que al clasificar se establece la identidad de los elementos que están representando los píxeles o las celdas que son discriminados en cada clase. En otras palabras, la identificación de los elementos se realiza sincrónicamente con la segregación de las celdas o píxeles.

Sobre la base de lo descrito en el párrafo anterior se reconoce que en los procedimientos de clasificación supervisada existen dos etapas ineludibles. Una inicial vinculada a la segregación de los píxeles para conformar clases o subconjuntos que identifican a los elementos que se desea individualizar, y la segunda que está dirigida a determinar la fiabilidad con que las celdas discriminadas representan a los elementos que son de nuestro interés.

Algoritmos de discriminación:

Los algoritmos aplicables en las clasificaciones supervisadas difieren de aquellos que son utilizados en la clasificación no supervisada. Mientras que en esta última la segregación se basa en algoritmos de agrupamiento, la clasificación supervisada se basa en algoritmos de discriminación. Los cuatro métodos más utilizados que suelen estar disponibles en programas de computación son los siguientes:

Paralelepípedo
Mínima distancia (distancia euclidiana)
Distancia de Mahalanobis
Maximum Likelihood (máxima verosimilitud o máxima probabilidad)

Al utilizar los tres primeros algoritmos se toma la decisión de discriminar las celdas en función de la valoración de la similitud, la cual se establece al comparar los parámetros que definen a cada celda con la variabilidad que está presente en una muestra representativa de los elementos que se van a clasificar, dato este último que debe ser generado por el operador a partir de celdas que representan los elementos de interés. A estos sitios de referencia se los suele llamar “regiones de interés”. A diferencia de las tres primeras alternativas, con *Maximum Likelihood* la discriminación entre las celdas se basa en cálculos de probabilidad.

Etapas en la clasificación supervisada

Tres etapas deben ser ejecutadas en el contexto de una clasificación supervisada. Éstas pueden ser identificadas como:

- 1) Etapa de muestreo
- 2) Etapa de Identificación
- 3) Etapa de valoración de la autenticidad (fiabilidad de los resultados)

La primera etapa es el proceso que se transita con el propósito de generar muestra con los píxeles representativos de los elementos que se van a clasificar e identificar. La etapa de identificación consiste en la aplicación y realización de los cálculos en función del algoritmo seleccionado. Mientras que por su parte, la etapa de valoración de la autenticidad tiene el propósito de establecer la exactitud y fiabilidad de la clasificación que resulta.

Práctica en el entorno del programa SAGA para ejecutar clasificaciones supervisadas.

La práctica que realizarás tiene el propósito de ejemplificar cómo es factible discriminar celdas o píxeles en las escenas de imágenes digitales mediante métodos de clasificación supervisada. Para lograrlo se utilizan dos fuentes de información. Una proveniente de imágenes Sentinel 2 que fueron registradas entre julio y diciembre de la campaña agrícola 2018, mientras que la segunda es un relevamiento que se realizó en noviembre de dicho año, en el cual se recopiló la información de la cobertura vegetal en diferentes sitios, relevamiento que constituye la verdad terrestre.

Generación de la firma fenológica

Como la variable a partir de la cual se contrastará la identidad es un índice que deriva de una firma fenológica, el primer proceso que debes realizar es el cálculo del NDVI con las bandas correspondientes a las longitudes del rojo e infrarrojo cercano. Este cálculo es preciso que lo realices para cada fecha disponible, 8 en total. Posteriormente tendrás que combinar las bandas NDVI consecutivamente, de modo que en la configuración de la serie temporal queden representadas las firmas fenológicas de la vegetación. Los materiales necesarios para calcular el NDVI y configurar la firma fenológica, están dispuestos en una carpeta en el aula virtual denominada Métodos de Clasificación (séptimo encuentro).

Una vez finalizado este proceso tienes que analizar las firmas fenológicas que se construyen a partir de parcelas donde es conocido que se sembró cebada o trigo (figura 1). El objetivo en esta oportunidad es discernir cómo definir un índice que facilite la discriminación entre los dos cultivos. En función de lo sucedido durante el período analizado (ver figura 1), la siguiente relación entre bandas que representan valores de NDVI será la utilizada para discriminar el trigo de la cebada.

$$\text{IDTyC} = (b5 - b8) / (b5 + b8) \text{ donde}$$

IDTyC: índice para discriminar trigo de cebada

b5: NDVI del 3 de octubre

b8: NDVI del 2 de diciembre

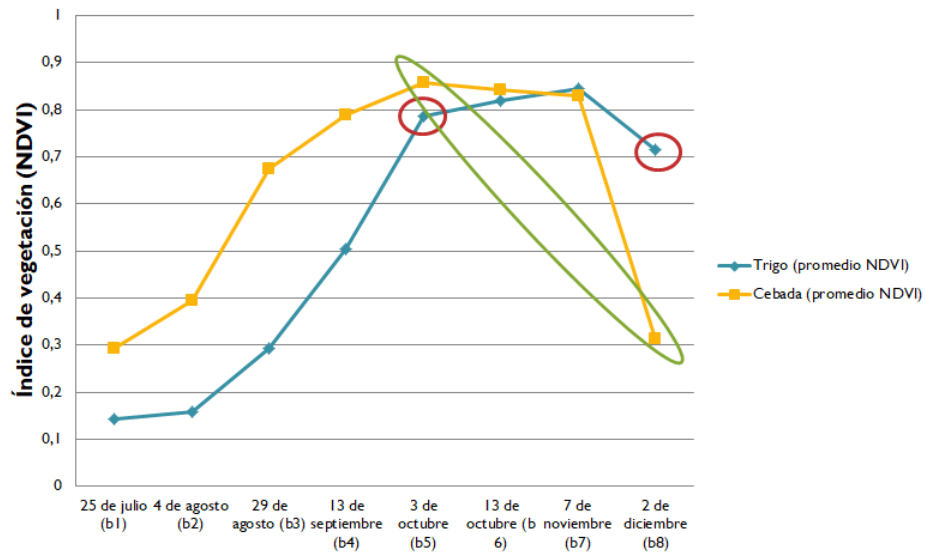


Figura 1. Firmas fenológicas en cultivos de cebada y trigo distribuidos sobre el corredor Mar del Plata - Balcarce (NDVI calculado a partir de la reflectancia de imágenes Sentinel 2).

En esta oportunidad los valores del índice que representan coberturas vegetales varían entre -1 y 1. En función de la representación de los promedios se puede prever que si el valor es positivo, cuanto más se aproxima al límite máximo (1) es más probable que la cobertura vegetal se corresponda a la de un cultivar de cebada. En cambio, los valores positivos o negativos que circundan al cero es probable que identifiquen píxeles que representan parcelas donde se sembró trigo. **Este proceso lo podés realizar utilizando la herramienta calculadora ráster del programa Qgis.** En esta oportunidad la Firma Fenológica te es provista, así como también el IDTyC.

Etapa de muestreo

Como hemos mencionado esta etapa tiene como meta compilar píxeles en la escena que conformarán las muestras que representan la variabilidad que caracteriza a los elementos que se desea discriminar e identificar. A continuación te proporcionamos los comandos que en el entorno del programa SAGA permiten la generación de las regiones de interés. Lo primero que debes realizar es lograr visualizar la imagen de la firma fenológica y también los puntos que muestran la distribución de los sitios que constituyen la verdad terrestre.

Visualización de la Firma Fenológica y los puntos de control:

Para comenzar debes recuperar la información de dos archivos, uno raster y otro vectorial. El archivo raster es el correspondiente a la Firma Fenológica, el vectorial es el de los puntos de control a partir de los cuales tomaremos las muestras que constituirán la verdad terrestre que se utilizará como referencia. Para ello dirige el mouse a **FILE**, luego a **GRID** y cliquea sobre **LOAD**. En la ventana que se despliega donde dice *All recognised Filed* selecciona **All Files**. Luego navega hasta la carpeta donde has guardado el archivo que representa la Firma Fenológica y selecciona un archivo xxxx.tif, cliquea en **Open**. En la ventana de la izquierda puede suceder que se incorporen todas las bandas que conforman la firma fenológica, en esta oportunidad son 8, o que sólo se incorpore el archivo donde se

compilaron las ocho bandas. En éste último caso debes pararte sobre el nombre de dicho archivo e ir a **Geoprocessing, Grid, Grid Collection** o, **Geoprocessing, Grid Collection, Tools** y finalmente clicar sobre **Extract Grids from Grid Collection**. Si se cierra el programa vuelve a intentarlo nuevamente hasta que permanezca activo. Luego de visualizar la imagen sobre la ventana de la derecha, debes desplegar el archivo shape file de puntos que representa los sitios geográficos donde se observaron las coberturas vegetales para definir la verdad terrestre. Para lograrlo dirige el mouse a **FILE**, luego a **SHAPES** y cliquea sobre **LOAD**. En la ventana que se despliega donde dice *All recognised Filed* selecciona **ESRI shape files**. Luego navega hasta la carpeta donde has guardado el archivo que representa los sitios de reconocimiento de las coberturas vegetales y selecciona el archivo xxxx.shp. Ahora configuraremos la ventana del programa para facilitar la ejecución de acciones que tendrás que realizar. Dirige el cursor del MOUSE a la opción **WINDOW** situada en el primer renglón de la ventana principal. En la ventana que se despliega asegúrate que estén tildadas **Show Manager Window** y **Show Object Properties Window**. Luego cliquea dos veces sobre el archivo de puntos "shape" que está en el panel de la izquierda y al abrirse la ventana elige que se despliegue sobre el mapa que ya has desplegado y que corresponde a la imagen de la Firma Fenológica. En la ventana donde se encuentran las propiedades (**properties**), *es la del centro de la pantalla*, desciende hasta hallar **Labels** y en **Attributes** elige la columna "cobertura" y cliquea sobre "**Apply**". Luego cliquea sobre el nombre del shape file de puntos de control con el botón derecho del mouse y selecciona **Classify**. En la ventana que se despliega en **Attributes** selecciona el campo "cobertura", en **Colors** elige (11 colors) y en **Classifications** selecciona **unique values**. Aplica y haz un zoom en el mapa. Deberías ver los puntos de control clasificados y etiquetados sobre el fondo de la imagen de la Firma Fenológica.

Generación de las regiones de interés que constituyen la verdad terrestre:

Para generar las regiones de interés debo visualizar las herramientas que están disponibles en la librería. Para lograrlo en el **Panel "Manager"** cliquea sobre **Tools** y se despliega la lista de la **Tool libraries**. Dirige el puntero del mouse a **Shapes**, despliega las opciones, selecciona **Tools** y dentro de estas **Create New Shapes Layer**. En la ventana que se despliega en **Name** asigna un nombre al archivo que crearás para representar las regiones de interés, es decir *cebada* y *trigo*. En **Shape Type** o **Geometry Type** y seleccione **Polygon**. El resto de las opciones déjalas por defecto y cliquea **Apply** y **Execute**. Dirige el mouse al panel "Manager" y cliquea en **Data**. En el panel se visualiza una capa con el nombre del *shape layer* que acabas de crear. Cliquea dos veces sobre el nombre del polígono y haz que se despliegue sobre la ventana donde está desplegada la firma fenológica y los puntos control. Ahora comenzaremos a digitalizar los polígonos. Para ello cliquea con el **botón derecho del mouse** sobre el nombre del shape layer, dirige el mouse a **Edit** y selecciona **Add Shape**. Cliquea sobre la barra de herramientas la flecha negra **Interc**, sobre la imagen en un sector que represente un cultivo, cliquea botón derecho del mouse y tilda **Add Part** y verás que en la ventana donde está desplegada la imagen aparece en la esquina superior izquierda un "círculo blanco con un punto negro" en el centro. Esto significa que la capa shape está editada y se puede comenzar a dibujar los polígonos. Dirígete a la imagen y comienza a trazar los polígonos en los sitios donde se ha verificado el cultivo que conforma la cubierta vegetal. Con cada clic del botón izquierdo del mouse se adiciona un nodo al polígono, y con un clic en el botón derecho del mouse se da por concluido el trazado de una de las partes de un polígono. Como al clicar con el botón derecho se te despliega una

ventana, elige **Add part** y ve a otro punto que identifique al mismo cultivo para el cual acabas de dibujar un área de interés. Repite la acción hasta que hayas muestreado todos los sitios que representan a alguno de los elementos que se pretenden clasificar. Cuando todos los polígonos que representan un cultivo estén dibujados, cliquea el botón derecho, quita el tilde a **Editar** y aplica los cambios seleccionando **Yes**. Nuevamente dirige el puntero del mouse al nombre del **Shape Layer** de las regiones de interés y cliquea con el botón derecho del mouse. Dirige el puntero a **Attributes** y selecciona **Show**. Ahora en la tabla donde dice **ID** repite el número de la columna de la izquierda, en este caso (1), y donde dice **Name** el nombre del cultivo que corresponda y cierra la tabla. Vuelve a clicar con botón derecho del mouse dirígelo a **Edit** y cliquea sobre **Add Shape**. Selecciona el ícono con la flecha negra y genera las áreas del polígono que va a representar las regiones de interés para el otro cultivo. Cuando finalices de dibujar la última de las áreas quita el tilde a **Edit Selection** y acepta aplicar los cambios. Una vez más con el botón derecho del mouse cliquea sobre el nombre del archivo de las regiones de interés, ve a **Attributes** y cliquea sobre **Show**. En la tabla que se despliega registra los datos según el criterio explicado anteriormente. En esta práctica deben quedar dos **ID** (1 y 2) y dos **Name** (trigo y cebada). Para visualizar los polígonos en diferentes colores según el tipo de cultivo, con el botón derecho del mouse cliquea sobre el *nombre del archivo de las áreas de interés*, dirige el mouse a **Classify** y elige el **“Attribute”** mediante el cual vas a clasificar, los **“Colors”** a utilizar (11 colores) y el criterio clasificador **Unique Value**. Si desean guardar las regiones de interés cliquea sobre **Save** en la ventana central de la pantalla que corresponde a las **“Properties”** del archivo y asigna una carpeta y un nombre.

Etapa de identificación:

Finalizada la etapa de muestreo es el turno de aplicar el algoritmo que establece el criterio de discriminación en una clasificación supervisada. Para realizar esta acción es preciso desplegar la imagen en la cual están registrados los datos que caracterizan a los elementos a discriminar. Del mismo modo que lo hicimos en la práctica anterior, utilizaremos la imagen donde el dato que se registra es el índice para discriminar trigo de cebada:

$$IDTyC = (b5 - b8) / (b5 + b8)$$

donde:

IDTyC: índice para discriminar trigo de cebada

b5: NDVI del 3 de octubre

b8: NDVI del 2 de diciembre

En esta oportunidad recuperaremos el dato que hemos generado en la práctica anterior, entonces despliega la imagen donde están registrados los IDTyC que representan los píxeles de lotes que han sido destinados a la producción de cultivos de invierno (trigo-cebada), recordá que es un archivo cuya extensión es xxxxxx.sgrd o xxxxx.tif. También debe estar desplegado en el panel de capas el archivo *Shape* que representa las *regiones de interés* (**Training Areas**). Para ejecutar la clasificación dirige el puntero del mouse hacia **Geoprocessing**, luego hacia **Imagery**, **Classification** y cliquea sobre **Supervised classification for Grids o Supervised image classification**. En la ventana que se despliega selecciona el **Grid system** que corresponda, en **Features** las bandas a partir de las cuales se realizará la discriminación, en **Classification** deja <create>, en **Taining Areas** selecciona el **Layer shape** con las áreas de interés, y en **Class Identifier** el

nombre de la columna que identifica a los elementos a individualizar, en este caso cobertura de trigo y cobertura de cebada. Finalmente elige el método en **Method** y cliques en **Okay**, espera el resultado y despliega la imagen que se crea.

Etapa de valoración de la autenticidad

Para valorar la calidad de la clasificación obtenida utilizaremos la opción matriz de confusión. En esta oportunidad dirige el puntero del mouse hacia **Geoprocessing**, luego a **Imagery**, luego **Classification** y finalmente clickea sobre **Confusion Matrix (Polygons / Grid)**. Selecciona el **Grid system** correspondiente a la imagen que resultó de aplicar el método de clasificación supervisada para el cual deseas valorar la fiabilidad de los resultados, asigna el nombre del **Shape Layer** que posee los polígonos a partir de los cuales se establecerá la evaluación y en **Class** el nombre del campo donde se identifican las coberturas trigo y cebada. Mantén <create> en las tres **Table** a generar. Luego despliega las tablas que se crearon. En la figura que se presenta a continuación se muestran los resultados a partir de los cuales podrás evaluar la individualización de los píxeles que se obtuvo mediante el procedimiento que realizamos. Para que puedas interpretar los resultados de la tabla te hemos proporcionado un archivo denominado “Matriz de confusión SAGA”, en el cual se describe cómo se leen los datos que te proporcionan las tres (3) tablas.

01. Confusion Matrix [Regiones de interés para trigo y cebada - Classification [Mahalanobis Distanc...]

	CLASS	Trigo	Cebada	SumUser	AccUser
1	Trigo	7738	165	7903	97.912185
2	Cebada	370	4978	5348	93.081526
3	Unclassified	314	7	0	0
4	SumProd	8108	5143		
5	AccProd	95.436606	96.791756		

02. Class Values [Regiones de interés para trigo y cebada - Classification [Mahalanobis Distance]]

	Class	SumRef	AccProd	SumClassified	AccUser
1	Trigo	8108	0.954366	7903	0.979122
2	Cebada	5143	0.967918	5348	0.930815

03. Summary [Regiones de interés para trigo y ...]

	NAME	VALUE
1	Kappa	0.91561
2	Overall Accuracy	0.959626