

Práctica 19 de marzo

Visualización de campos escalares

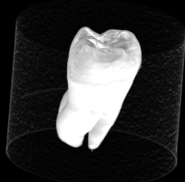
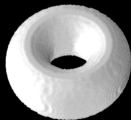
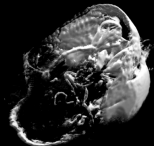
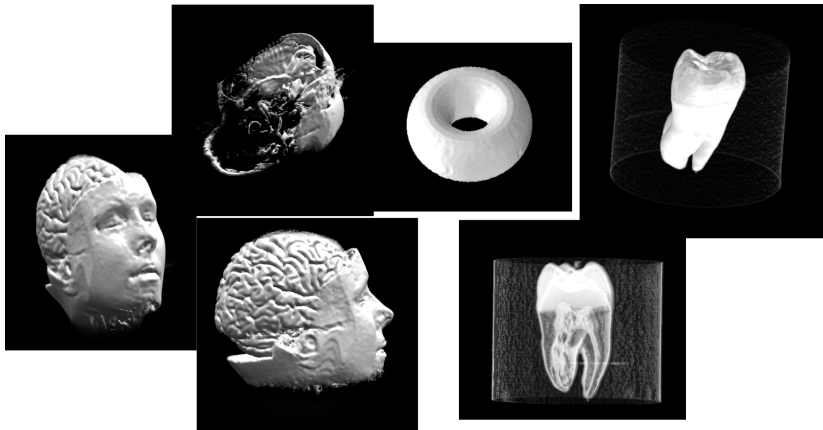
- Instalar Image-J
- Instalar plugin Bio-formats
- instalar plugin Volume Viewer

<https://imagej.nih.gov/ij/download.html>

<https://imagej.net/Bio-Formats>

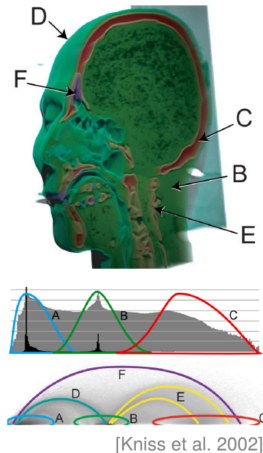
<https://docs.openmicroscopy.org/bio-formats/5.8.2/users/imagej/installing.html>

<https://www.openmicroscopy.org/bio-formats/downloads/>



Función de transferencia 1D:
horizontal: valor escalar
vertical: núm. de voxeles

Función de transferencia 2D:
horizontal: valor escalar
vertical: magnitud del gradiente



Actividad

Para los conjuntos de datos disponibles para esta práctica (aunque también puede buscar otros conjuntos de datos de su interés):

- Genere renders volumétricos de los conjuntos de datos usando ImageJ y los plugins instalados.
- Para explorar la información que despliega en pantalla puede usar las herramientas Función de transferencia 1-D o 2-D (puede modificar también los parámetros del modelo de iluminación)
- ¿Qué efecto tiene sobre los datos que se despliegan en pantalla el uso de estas funciones de transferencia?
- Describa el proceso que tendría que llevar a cabo para visualizar una estructura específica (por ejemplo, los ventrículos cerebrales) en los datos de resonancia magnética de la cabeza. ¿Se podría lograr esto únicamente usando funciones de transferencia? Explique.

Filtrado de datos

Actividad

1) Haga la lectura del fichero “titanic3.csv”. Según este dataset:

- Cuántos cuerpos fueron encontrados? Cuántos de éstos eran hombres mayores de 40 años?
- Cuántas mujeres desaparecidas con edades entre 15 y 35 años?
- Cuántos hombres mayores de 20 años sobrevivieron?
- Cuántas mujeres menores de 25 años sobrevivieron?

2) Genere una copia del dataset y rellene los datos faltantes (NA's) con un valor de 0 en el caso de datos numéricos usados como identificador, la palabra “desconocido” en el caso de datos tipo cadena de caracteres y en el caso de variables numéricas use el promedio de los valores de esa columna (p.e., la edad y la tarifa)

3) De los campos “age” y “fare” agregue columnas al dataset que contenga los valores normalizados. Elija la normalización tipo

$$\frac{x_i - \bar{x}}{\sigma},$$

en el caso de que la variable tenga una distribución normal, y la normalización tipo

$$\frac{x_i - x_{min}}{x_{max} - x_{min}},$$

en otro caso.

Distancias

- Un conjunto de datos puede denotarse $\mathbf{X} = \{\mathbf{x}_i^T\}$ con $i = 1, 2, \dots, N$. N es tamaño del conjunto.
- donde cada $\mathbf{x}_i^T \in \mathbb{R}^n$; $\mathbf{x}_i^T = (x_{i,1}, x_{i,2}, \dots, x_{i,n})$. n es el número de variables de cada elemento del conjunto \mathbf{X} .

Tipos de distancias

Distancia de Manhattan

$$D_1(\mathbf{x}_i, \mathbf{x}_j) = \sum_{k=1}^n |x_{i,k} - x_{j,k}|,$$

para $i, j = 1, \dots, N$

Tipos de distancias

Distancia Euclidea

$$D_2(\mathbf{x}_i, \mathbf{x}_j) = \left[\sum_{k=1}^n (x_{i,k} - x_{j,k})^2 \right]^{1/2},$$

para $i, j = 1, \dots, N$

Tipos de distancias

Distancia de Minkowski

$$D_p(\mathbf{x}_i, \mathbf{x}_j) = \left[\sum_{k=1}^n (x_{i,k} - x_{j,k})^p \right]^{1/p},$$

para $i, j = 1, \dots, N$

Usando el método **distance_matrix** de **scipy.spatial** se puede construir tambien la matriz de distancias.

... en python

user_id	1	2	3	4	5	6	7	8	9	10
user_id										
1	0.0	9.9	15.9	9.1	4.2	6.9	10.5	7.4	5.6	0.7
2	9.9	0.0	12.4	17.2	6.1	6.2	0.8	4.9	11.7	9.6
3	15.9	12.4	0.0	12.4	18.5	9.0	12.0	17.3	12.9	15.2
4	9.1	17.2	12.4	0.0	12.7	11.0	18.0	15.3	5.5	8.8
5	4.2	6.1	18.5	12.7	0.0	9.5	6.5	3.2	8.2	3.9
6	6.9	6.2	9.0	11.0	9.5	0.0	7.0	8.3	5.5	6.2
7	10.5	0.8	12.0	18.0	6.5	7.0	0.0	5.3	12.5	10.2
8	7.4	4.9	17.3	15.3	3.2	8.3	5.3	0.0	9.8	7.1
9	5.6	11.7	12.9	5.5	8.2	5.5	12.5	9.8	0.0	4.9
10	0.7	9.6	15.2	8.8	3.9	6.2	10.2	7.1	4.9	0.0

Usando los datos en el archivo “movies.csv”

- Construya una o varias funciones que permitan calcular la matriz de distancias para los datos numéricos en el `dataFrame`. La función debe permitir construir la matriz de distancia usando las medidas de Manhattan, Euclideana y de Minkowski (para $p \geq 3$).
- Compare sus resultados con los del método **`distance_matrix`** de **`scipy.spatial`**.

Actividad - Clustering jerárquico

- Usando los métodos “dendrogram” y “linkage” construya un diagrama en forma de árbol (dendrograma) para el conjunto de datos en “movies.csv”.
- Repita el proceso ahora usando algún esquema de normalización del rango de los datos.
- ¿Que diferencias puede encontrar en los resultados previos?
- ¿En qué casos resulta importante llevar a cabo un proceso de normalización del rango de datos?
- Consulte los diferentes tipos de distancias que se pueden usar como parámetro en el método “linkage”. ¿En qué características de los datos se podría basar para elegir una determinada distancia?