

## Segundo Examen Parcial

DOCENTE	CARRERA	CURSO
MSc. Vicente Machaca Arceda	Escuela Profesional de Ciencia de la Computación	Estructura de Datos Avanzada

### 1. Competencias del curso

- Conocer e investigar los métodos de acceso multidimensional, métrico y aproximado.
- Analiza, diseña y propone soluciones utilizando estructuras de datos avanzadas.
- Comprende la importancia e impacto de los algoritmos estudiados y las nuevas propuestas.
- Aplica principios matemáticos para la solución de problemas.

### 2. Competencias del examen

- Comprende e implementa la estructura multidimensional *KD-Tree*.

### 3. Equipos y materiales

- Python
- Cuenta en Github
- IDE de desarrollo

### 4. Entregables

- Un informe en PDF donde se detalle el trabajo. Respecto a este informe:
  - Debe estar hecho en Latex.
  - Solo un integrante de cada grupo sube el informe a Classroom.
  - El informe debe ser nombrado "GRUPO-X", donde "X" es el nombre del grupo (1A, 2A, 1B, etc.).
  - Debe incluir un enlace al repositorio Github donde esta el código.
  - Debe tener el código fuente así como capturas de pantalla de la ejecución y resultados del mismo.

## 5. Descripción del trabajo

La estructura KD-Tree es una estructura multidimensional de  $k$  dimensiones. Esta permite implementar búsquedas por similitud como *K Nearest Neighbor* o *Closest point*. Adicionalmente, se puede usar esta estructura como un clasificador. Usted debe implementar este clasificador en el tema de su preferencia. A continuación detallamos el algoritmo:

---

**Algorithm 1:** KNN Classifier

---

**Input:**  $X$ : training data;  $y$ : object to be classified.

**Output:** Classification for  $y$ .

Extract features of each sample;

Build KD-Tree;

Select KNN of  $y$  in  $X$ ;

Class( $y$ )  $\leftarrow$  max of classes (  $k$  closest objects );

---

Para mas detalles de este algoritmo puede revisar el trabajo de Hou [1]. En cuanto al clasificador, usted es libre de escoger el tema, algunos ejemplos pueden ser:

- Clasificación de señales de transito.
- Clasificación de emociones faciales.
- Clasificación de tumores malignos y benignos.
- Clasificación de correo *spam*.
- Etc.

Usted tambien es libre de escoger el descriptor. Este descriptor es un método que toma como entrada una muestra de la base de datos y retorna un vector de características, luego este vector representa un punto en el KD-Tree. Usted tambien, puede evaluar varios descriptores y hacer comparativas. Entre algunos descriptores tenemos:

- Para imágenes:
  - Key points: SIFT.
  - Bordes: Canny.
  - Texturas: *Local Binary Patterns*.
  - *Histograms*.
- Para texto:
  - Frecuencia de palabras clave.
  - Bolsa de palabras.
- Para otras señales en general:
  - Fourier.
  - Wavelets.
- Seguido a un descriptor, puede aplicar reducción de dimensiones como PCA o eliminar atributos: *recursive feature elimination*.

## 6. Rúbricas

Rúbrica	Cumple	Cumple con obs.	No cumple
Implementa el algoritmo de clasificación en el KD-Tree.	6	3	0
Evalúa los descriptores y mide los resultados (accuracy).	6	3	0
El informe contiene todo el código fuente, pantallazos de la aplicación y buena redacción.	4	2	0
Durante la presentación el alumno presenta dominio del tema.	4	2	0

## Referencias

- [1] W. Hou, D. Li, C. Xu, H. Zhang, and T. Li, “An advanced k nearest neighbor classification algorithm based on kd-tree,” in *2018 IEEE International Conference of Safety Produce Informatization (IICSPI)*, pp. 902–905, IEEE, 2018.