**Nombre del Algoritmo:** Variable kernel density estimation

**Descripción del Problema Computacional para la cual el algoritmo es una solución.**

**Delimitación del Problema:**

El algoritmo de estimación de densidad con núcleo variable (VKDE) aborda el problema de estimar una función de densidad de probabilidad 𝑓(𝑥) para un conjunto de datos univariados o multivariados, sin asumir una distribución específica. A diferencia de la estimación de densidad tradicional con núcleos, VKDE ajusta la anchura del núcleo ℎ de manera adaptativa en función de la densidad local de los datos.

**Comprensión del Problema:**

El objetivo es obtener una representación suave de la distribución subyacente de los datos observados que capture tanto las áreas densamente pobladas como las áreas más dispersas. Esto se utiliza en problemas como:

* Identificación de patrones no lineales.
* Modelado de distribuciones no paramétricas.
* Análisis exploratorio de datos (detección de modos o "picos").

**Clasificación del Problema:**

Este problema se clasifica como un problema de análisis no paramétrico dentro de la estadística computacional. También se relaciona con la visualización y modelado probabilístico en el aprendizaje automático y la minería de datos.

**Viabilidad Computacional:**

La viabilidad depende de:

* Tamaño del conjunto de datos (𝑛).
* Dimensionalidad de los datos (𝑑).
* Selección del tipo de núcleo y del mecanismo para ajustar las anchuras ℎ.

**Contexto del Problema:**

Se encuentra en aplicaciones como:

* Análisis de datos geoespaciales.
* Reconocimiento de patrones (por ejemplo, detección de anomalías).
* Modelado probabilístico para sistemas dinámicos.

**Análisis de Datos y Estructura del Problema:**

1. Datos de entrada:

* Un conjunto de puntos {𝑥1, 𝑥2,…,𝑥𝑛}.

1. Estructura del problema:

* Cada punto tiene una "vecindad" adaptativa determinada por el parámetro de suavizamiento local ℎ𝑖.
* La densidad se estima como una suma ponderada de núcleos centrados en cada punto.

**Evaluación de Complejidad:**

La complejidad típica del cálculo de densidad para un punto de evaluación es 𝑂(𝑛⋅𝑑), donde 𝑛 es el número de puntos y 𝑑 es la Dimensionalidad de los datos. La complejidad puede reducirse con métodos de optimización, como estructuras de datos jerárquicas (por ejemplo, árboles KD).

**Análisis del Contexto del Problema para la cual el algoritmo es una solución**.

**Contexto general del Problema:**

La estimación de densidad con núcleo variable es crucial en contextos donde:

* Las distribuciones no se ajustan a un modelo paramétrico conocido.
* Se requiere precisión tanto en regiones densamente pobladas como en regiones con pocos datos.

**Identificación de los conjuntos y estructuras de datos:**

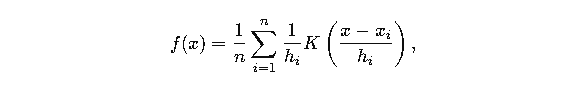
1. Datos de entrada (𝑋):

* Una matriz o vector de datos de muestra 𝑋={𝑥1,𝑥2,…,𝑥𝑛}

1. Parámetros del Algoritmo:

* Núcleo base 𝐾(𝑢), como gaussiano o epanechnikov.
* Ancho de banda adaptativo ℎ𝑖, que puede depender de la densidad local.

**Formalización del problema:**

Dado un conjunto de datos 𝑋= {𝑥1, 𝑥2,…,𝑥𝑛}, estimar 𝑓(𝑥) como:

Donde:

* K(u) es una función de núcleo que satisface ∫ 𝐾 (𝑢) 𝑑𝑢 = 1
* ℎ𝑖 es un ancho de banda adaptativo.

**Análisis de la Dinámica del Problema:**

El núcleo variable ajusta ℎ𝑖 según la densidad local:

* En áreas densas (𝑥𝑖 cerca de otros puntos): ℎ𝑖 es más pequeño para capturar los detalles.
* En áreas dispersas: ℎ𝑖 es mayor para suavizar la estimación.

**Evaluación de la Complejidad del Contexto:**

* La implementación básica es costosa para grandes conjuntos de datos debido a la dependencia cuadrática 𝑂(𝑛2).
* Métodos como aproximaciones en ventanas deslizantes o submuestreo ayudan a mejorar la escalabilidad.

**Identificación de los casos límites y excepciones:**

**Casos límite:**

* Muestras pequeñas: VKDE puede sobreajustar los datos.
* Muestras grandes: Requiere estrategias de optimización para ser eficiente.

**Excepciones:**

* Densidades disjuntas o distribuciones multimodales extremas: el algoritmo puede no estimar bien sin ajustes específicos.

**Interacción con otros sistemas:**

VKDE puede interactuar con:

* Sistemas de visualización para mostrar la densidad estimada.
* Algoritmos de clustering para refinar grupos detectados.
* Algoritmos de detección de anomalías para identificar áreas de baja densidad.

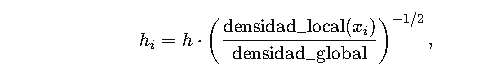
**Especificación Formal:**

**Definición de las entradas y salidas:**

* Entradas:
  + X= {x1, x2 ,…,x n}: Conjunto de datos.
  + K(u): Función de núcleo.
  + {h1,h2 ,…,hn}: Anchos de banda variables.
  + Puntos de evaluación {𝑥𝑒𝑣𝑎𝑙}(opcional).
* Salidas:
  + Densidad estimada 𝑓(𝑥) en cada 𝑥𝑒𝑣𝑎𝑙

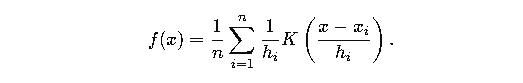
**Definición formal del problema mediante funciones:**

* Calculo de ℎ𝑖:



donde densidad\_local(𝑥𝑖) se aproxima con KDE tradicional.

* Estimación de densidad:



**Relaciones y Restricciones:**

* **Relaciones:**
  + El parámetro ℎ𝑖 debe ajustarse para reflejar la densidad local.
  + K(u) debe ser continuo y simétrico.
* **Restricciones:**
  + La elección de 𝐾(𝑢) y el método de ajuste de ℎ𝑖 impactan la precisión.

**Invariantes y propiedades:**

* El área bajo la curva estimada 𝑓(𝑥) debe integrar a 1.
* La suavidad de 𝑓(𝑥) está garantizada por 𝐾(𝑢).

**Control de Flujo:**

1. Preprocesamiento:
   * Normalización de datos.
   * Selección de núcleo 𝐾(𝑢).
2. Estimación preliminar de densidad para calcular ℎ𝑖
3. Cálculo de la densidad 𝑓(𝑥) usando ℎ𝑖 variable.
4. Visualización y evaluación del resultado.

**Pseudocodigo del Algoritmo:**

Algoritmo EstimacionDensidadVariableKernel

// Entradas:

// datos\_entrenamiento: Matriz[n] con los puntos de datos.

// datos\_consulta: Matriz[k] con los puntos para los que se quiere estimar la densidad.

// kernel: Función kernel (por ejemplo, gaussiana).

// ancho\_base: Escala base del ancho del kernel.

Funcion CalcularDensidad(datos\_entrenamiento, datos\_consulta, kernel, ancho\_base)

Definir densidades como Vector[tamaño(datos\_consulta)]

// Calcular la densidad local en cada punto de consulta

Para i <- 1 Hasta tamaño(datos\_consulta)

consulta <- datos\_consulta[i]

densidad <- 0

// Ajustar el ancho del kernel en función de la densidad local

Para j <- 1 Hasta tamaño(datos\_entrenamiento)

punto <- datos\_entrenamiento[j]

distancia <- Abs(consulta - punto)

ancho\_variable <- ancho\_base / (1 + distancia) // Ancho adaptativo

densidad <- densidad + kernel(distancia / ancho\_variable) / ancho\_variable

FinPara

// Normalizar la densidad

densidades[i] <- densidad / tamaño(datos\_entrenamiento)

FinPara

Retornar densidades

FinFuncion

// Función Gaussiana como ejemplo de kernel

Funcion KernelGaussiano(u)

Retornar (1 / sqrt(2 \* PI)) \* Exp(-0.5 \* u^2)

FinFuncion

// Proceso principal

Escribir "Iniciando el cálculo de densidad con kernel variable..."

Definir datos\_entrenamiento como Vector[n] // Cargar datos

Definir datos\_consulta como Vector[k] // Cargar puntos de consulta

Definir ancho\_base como Real // Definir ancho base del kernel

// Calcular densidades utilizando el kernel gaussiano

densidades <- CalcularDensidad(datos\_entrenamiento, datos\_consulta, KernelGaussiano, ancho\_base)

// Mostrar resultados

Para i <- 1 Hasta tamaño(datos\_consulta)

Escribir "Punto: ", datos\_consulta[i], " - Densidad: ", densidades[i]

FinPara

FinAlgoritmo

**Casos de usos Posible:**

* Análisis de Datos Financieros:
  + Estimar la densidad de precios en un mercado financiero para identificar patrones o áreas con alta concentración de transacciones.
* Procesamiento de Señales:
  + Suavizar datos en series temporales con ruido, como mediciones de sensores.
* Reconocimiento de Patrones:
  + Identificar grupos densos en datos de alta dimensionalidad para clasificación o detección de anomalías.
* Estudio de Distribuciones Demográficas:
  + Analizar la distribución espacial de una población en un área geográfica, ajustando la escala según la densidad local.
* Detección de Anomalías:
  + Detectar eventos raros o inusuales en un conjunto de datos al identificar puntos con baja densidad estimada.
* Imágenes Médicas:
  + Estimar distribuciones de intensidad en imágenes médicas para segmentación o detección de tejidos.
* Optimización de Recursos:
  + Estimar densidades en mapas de tráfico para asignar recursos en áreas congestionadas.