

Trabajo práctico especial

Teoría de la información



Hermida Matias

mhermida@alumnos.exa.unicen.edu.ar

Soutrelle Perez Axel Nahuel

asoutrelleperez@alumnos.exa.unicen.edu.ar



Índice

| | |
|--------------------------|----------|
| Introducción..... | 1 |
| Parte 1..... | 1 |
| Inciso 1..... | 1 |
| Inciso 2..... | 3 |
| Parte 2..... | 3 |
| Inciso 1..... | 3 |
| Inciso 2.1..... | 4 |
| Inciso 2.2..... | 6 |
| Parte 3..... | 6 |
| Inciso 1..... | 6 |
| Inciso 2..... | 7 |
| Inciso 3..... | 7 |
| Parte 4..... | 8 |
| Inciso 1..... | 8 |
| Inciso 2..... | 8 |
| Conclusión..... | 9 |

Introducción

En el presente trabajo práctico se aplicarán conceptos de Teoría de la Información para analizar, comprimir y transmitir datos climáticos reales de tres ciudades: Quito, Melbourne y Oslo.

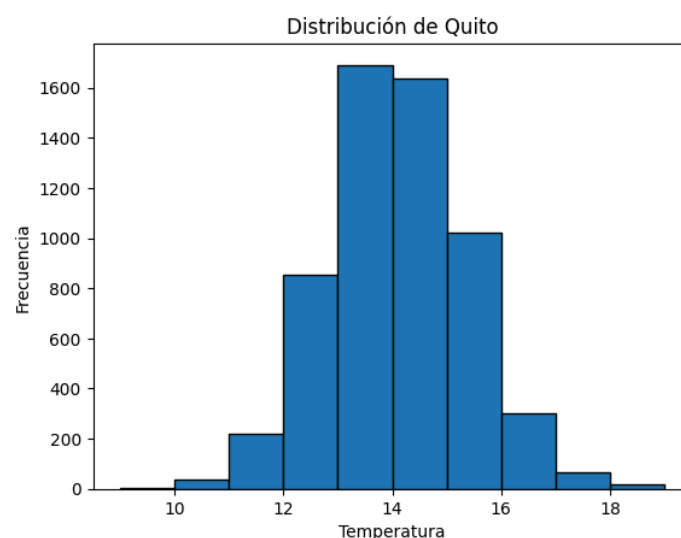
El objetivo es modelar las señales como fuentes de información, evaluando su comportamiento estadístico, predictibilidad, entropía y eficiencia de codificación, considerando además un canal de transmisión ruidoso. Se tratan problemas como: análisis estadístico (media, desviación estándar, correlación cruzada). Modelado de fuente con memoria para temperaturas discretas (frío, templado, cálido), simulando y estimando probabilidades estacionarias y tiempos de 1° recurrencia. Cálculo de entropía y compresión de datos usando Huffman, aplicación del Teorema de Shannon. Análisis de canal de comunicación, calculando ruido e información mutua, para juzgar la calidad de transmisión.

Parte 1

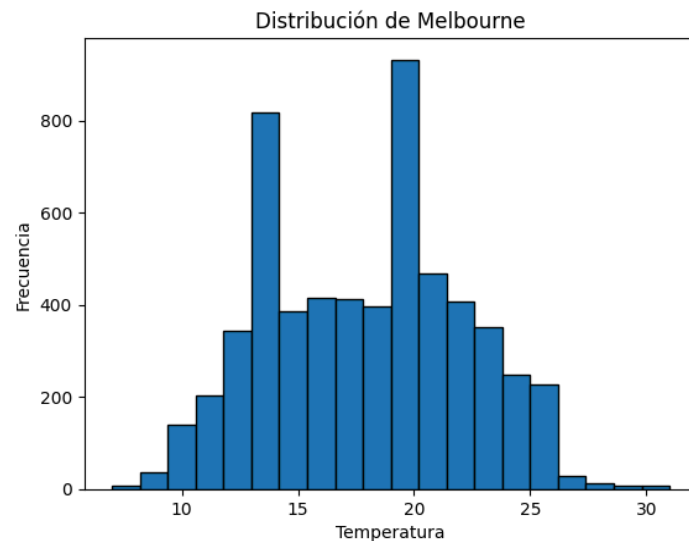
El análisis se inició con la visualización de los valores de temperaturas por cada ciudad. Se detectaron outliers en Melbourne y Oslo, los cuales fueron reemplazados mediante imputación por la media para preservar la coherencia de los datos.

Inciso 1

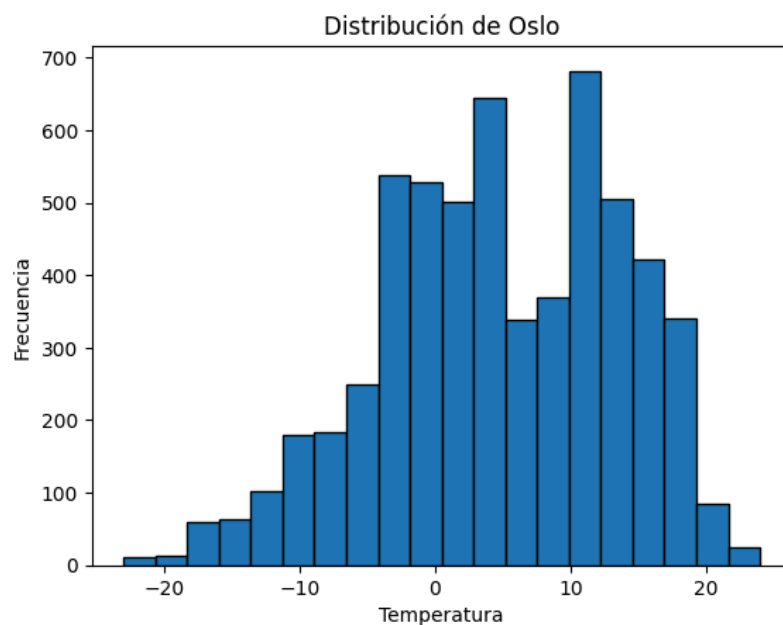
La temperatura promedio en Quito es de aproximadamente 13.6°C , con una desviación estándar de 1.30°C . Esto indica que las temperaturas tienden a concentrarse cerca del valor medio, con una variabilidad baja. La distribución, como se observa en el histograma, tiene una forma casi simétrica, lo que sugiere un comportamiento estable del clima de la ciudad a lo largo del tiempo.



La temperatura promedio en Melbourne es de 17.79°C , con una desviación estándar de 4.27°C , por lo que presenta una mayor variabilidad a comparación de Quito. El histograma muestra dos picos, en torno a los 14°C y 20°C .

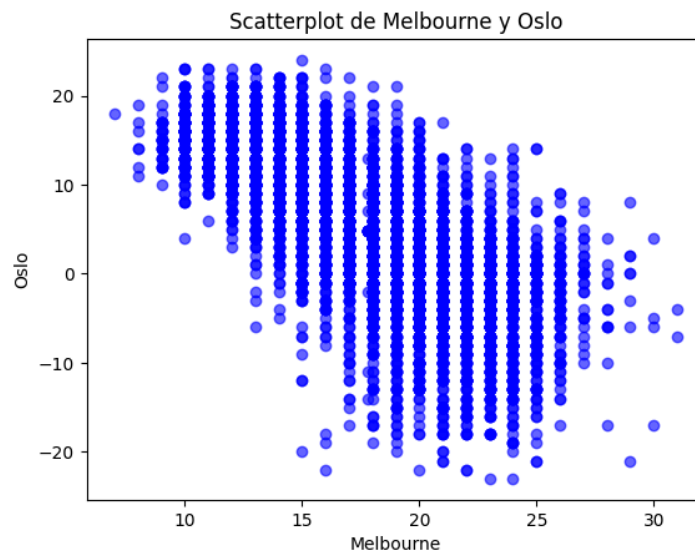


La temperatura promedio en Oslo es de aproximadamente 4.77°C , con una desviación estándar de 8.78°C , lo que indica la mayor variabilidad térmica de las tres ciudades. El histograma muestra una distribución con dos picos: cercanos a los 2°C y otro a los 11°C . Esta forma puede representar dos estaciones bien diferenciadas, ya que las temperaturas van desde valores negativos extremos hasta más de 20°C .



Inciso 2

De las tres correlaciones analizadas, la única que muestra una relación significativa se da entre Melbourne y Oslo, con un coeficiente de correlación de -0.7284 . Este valor indica una fuerte correlación negativa, lo que significa que cuando las temperaturas en Melbourne aumentan, tienden a disminuir en Oslo, y viceversa. Esta relación inversa puede explicarse ya que Melbourne se encuentra en el hemisferio sur, mientras que Oslo está en el hemisferio norte. Por lo tanto, cuando en Melbourne es verano, en Oslo es invierno.




Parte 2

Para analizar cómo se comportan las temperaturas en las ciudades, discretizamos las series dividiendo los valores en tres estados: frío (F) para temperaturas menores a 11°C , templado (T) entre 11°C y 19°C y cálido (C) mayores o iguales a 19°C .


Inciso 1

Calculamos la matriz de transición de primer orden, que representa la probabilidad de pasar de un estado a otro en dos días consecutivos. Esta matriz se construyó contando las transiciones entre pares de estados consecutivos y normalizando por el total de veces que se estuvo en cada estado.


En Quito, si hoy es un día templado, hay una probabilidad muy alta de que el día siguiente también sea templado, lo que indica que tiene largas rachas de días con temperaturas templadas. Desde templado, es poco frecuente que se pase a frío o cálido, mostrando que las variaciones hacia extremos son raras. Si hoy es un día frío o cálido, hay una probabilidad baja de que el próximo también lo sea, lo que muestra que generalmente se vuelve a una temperatura templada.

|  | F | T | C |
|---|--------|--------|-----|
| F | 0.0541 | 0.006 | 0.0 |
| T | 0.9459 | 0.9933 | 1.0 |
| C | 0.0 | 0.0007 | 0.0 |

En Melbourne, luego de un día templado, hay una alta probabilidad de que el día también sea templado. Lo mismo ocurre con los días cálidos, mientras que para los días fríos hay una probabilidad parecida entre que el proximo dia sea frío o templado

|  | F | T | C |
|---|--------|--------|--------|
| F | 0.4863 | 0.0316 | 0.0 |
| T | 0.5137 | 0.8359 | 0.1467 |
| C | 0.0 | 0.1325 | 0.8533 |

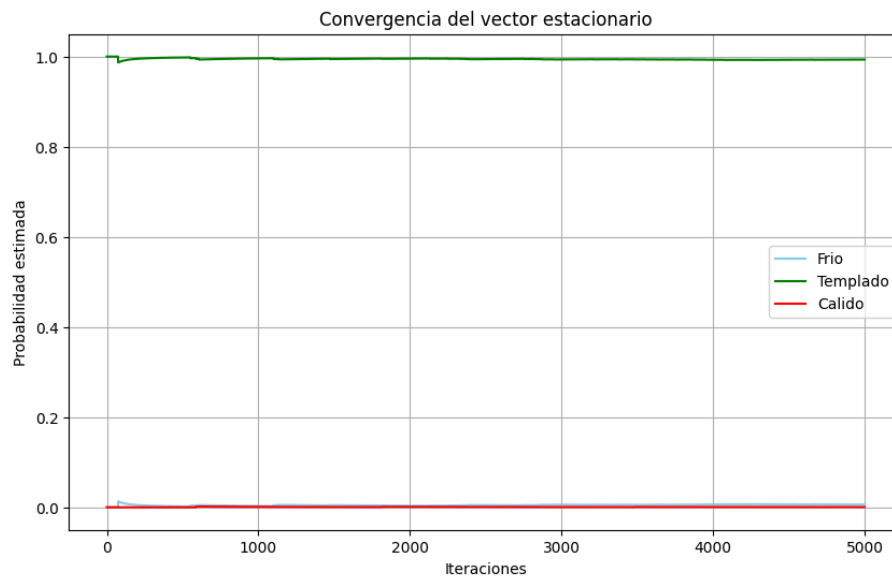
En Oslo, hay una probabilidad muy alta de que haya días de frío consecutivos, esto indica que suele haber inviernos largos, mientras que si se da un día templado, hay una gran probabilidad de que el siguiente también.

|  | F | T | C |
|---|--------|--------|--------|
| F | 0.9593 | 0.0994 | 0.0 |
| T | 0.0407 | 0.8652 | 0.2959 |
| C | 0.0 | 0.0354 | 0.7041 |

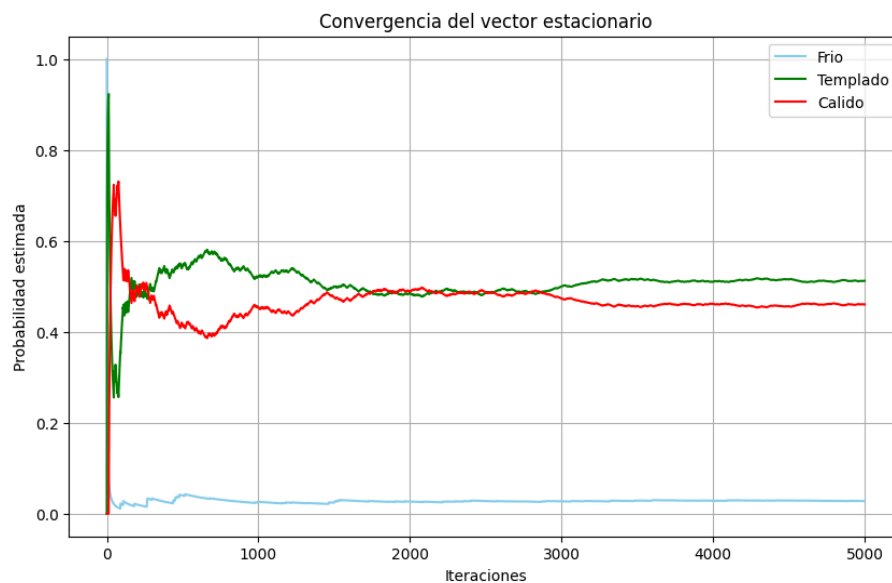
Inciso 2.1

Utilizamos muestreo Monte Carlo para obtener la probabilidad estacionaria para los estados de cada ciudad.

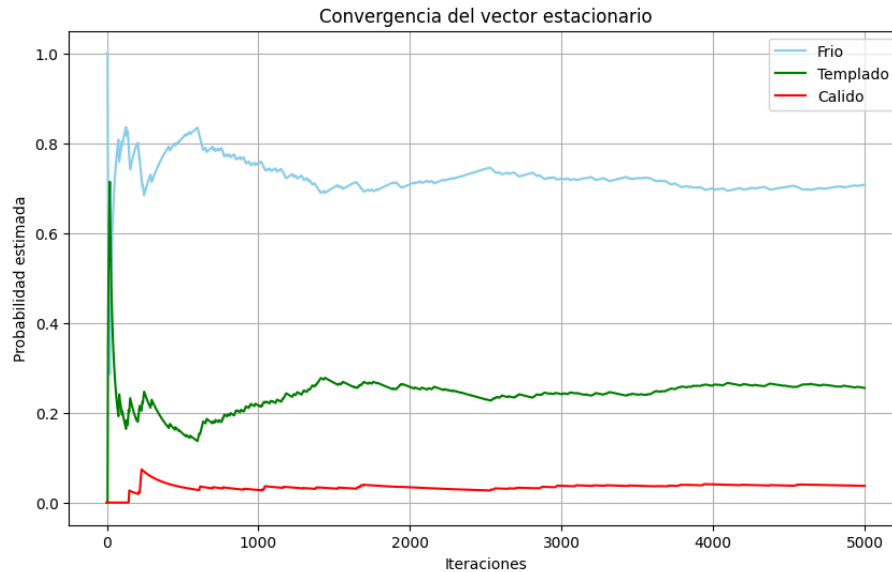
En Quito, en el largo plazo, hay una gran probabilidad de que un día sea templado, los días fríos y cálidos son muy raros. Esto indica que la ciudad tiene un clima constante, con temperaturas templadas casi todo el tiempo, lo que coincide con su ubicación cerca del ecuador.



En Melbourne hay una distribución bastante equilibrada entre días templados y cálidos, reflejando la variación del clima, los días fríos son poco frecuentes, pero no nulos.



En Oslo hay una alta probabilidad de días fríos en el largo plazo, le siguen los días templados, mientras que los días calurosos son muy raros. Esto demuestra que el clima de la ciudad está dominado por el frío, coincide con su ubicación en el norte de Europa.



Inciso 2.2

En Quito el estado templado se vuelve a ver prácticamente al día siguiente. Para que se repita un día de frío van a pasar aproximadamente 5 meses desde el último día que hizo frío, y para que se detecte un día cálido desde el último detectado van a pasar casi 4 años.

En Melbourne templado y cálido tienen recurrencias muy similares, cada dos días aproximadamente. El frío aparece, pero es menos común, se repite cada 1 mes.

En Oslo el estado frío se repite casi todos los días, lo cual es coherente con su alta frecuencia estacionaria. Templado aparece regularmente cada 4 días, y cálido es muy similar al frío en Melbourne, se detecta cada 1 mes.

Parte 3

Inciso 1

Quito tiene muy baja entropía, tanto con como sin memoria. Esto indica que la temperatura en Quito es altamente predecible, probablemente porque se mantiene estable durante todo el año (clima ecuatorial).

Melbourne tiene muy alta entropía sin memoria, pero baja con memoria. Indica que si no consideramos el estado anterior, las temperaturas parecen muy variables, cuando usamos un modelo con memoria, la incertidumbre baja muchísimo, lo que significa que el estado anterior da bastante información sobre el siguiente.

Oslo tiene alta entropía sin memoria (aunque un poco menor que Melbourne) y una entropía con memoria intermedia. Esto sugiere que el clima de Oslo es menos predecible que el de Quito, pero más predecible que Melbourne.

Inciso 2

Luego de aplicar algoritmos de codificación de Huffman tanto para orden 1 (sin memoria) como para orden 2 (con memoria), podemos observar que en todos los casos se cumple la desigualdad de Shannon, lo que implica que los códigos generados son válidos y eficientes, es decir, no desperdician espacio, pero tampoco son imposibles de decodificar.

| Códigos Huffman Quito orden 2: | Códigos Huffman Melbourne orden 2: | Códigos Huffman Oslo orden 2: |
|--------------------------------|------------------------------------|-------------------------------|
| TF: 00 | TT: 0 | FC: 0000000 |
| TC: 0100 | TC: 100 | CF: 0000001 |
| CC: 0101000 | TF: 10100 | TC: 000001 |
| FC: 01010010 | FC: 10101000 | CT: 00001 |
| CF: 01010011 | CF: 10101001 | TF: 0001 |
| FF: 010101 | FF: 1010101 | CC: 0010 |
| CT: 01011 | FT: 101011 | FT: 0011 |
| FT: 011 | CT: 1011 | TT: 01 |
| TT: 1 | CC: 11 | FF: 1 |

Nota: La codificación de Huffman depende de las probabilidades dadas por el muestreo Monte Carlo.

Inciso 3

Quito alcanza la mayor tasa de compresión. Esto se debe a que sus datos presentan muy baja entropía, es decir, son altamente repetitivos y predecibles. En consecuencia, el algoritmo de codificación de Huffman puede representar la información con códigos muy cortos, lo que reduce significativamente la cantidad de bits necesarios.

Melbourne tiene la menor tasa de compresión, lo que refleja mayor variabilidad climática. Logra una compresión considerable al usar memoria, pero sigue siendo la ciudad con menor tasa entre las tres.


Oslo se encuentra en un punto intermedio, cierta regularidad en los datos permite una buena compresión, pero no tan alta como Quito. Mejora mucho su compresión al incorporar memoria.

Las tasas de compresión están directamente relacionadas con la entropía: a menor entropía, mayor compresión. Quito, con su clima casi constante, es la fuente que permite una mayor compresión, mientras que Melbourne es la que permite una menor compresión debido a su mayor variabilidad. Incorporar memoria en el modelado de las fuentes siempre mejora la compresión, ya que así podemos tener en cuenta las dependencias entre estados.

Parte 4

Inciso 1

En el canal, aumenta la probabilidad de permanecer frío y también la probabilidad de que un día templado sea recibido como frío, mostrando que el canal introduce ruido que dispersa las temperaturas medidas. El canal introduce ruido principalmente en el estado templado, mientras que los extremos (frío y cálido) tienden a conservarse.

|  | F | T | C |
|---|--------|--------|--------|
| F | 0.651 | 0.138 | 0.0 |
| T | 0.4836 | 0.6722 | 0.2428 |
| C | 0.0 | 0.1834 | 0.7563 |

Inciso 2

El ruido del canal es relativamente alto. Esto indica que el canal introduce una cantidad considerable de incertidumbre o error. La información mutua es un valor bastante bajo, indicando que la salida del canal (señal recibida) contiene poca información útil sobre la señal original transmitida. En otras palabras, gran parte de la señal se pierde o se distorsiona. El canal es poco confiable y puede ser considerado un canal ruidoso y defectuoso para transmitir la señal climática, introduce confusión especialmente en los días templados, lo que disminuye la capacidad para predecir el estado real a partir de la señal recibida. Este canal no es útil para la transmisión fiable de la información climática de Melbourne.

Conclusión

A lo largo de este trabajo se aplicaron conceptos fundamentales de la Teoría de la Información para modelar y analizar datos climáticos de tres ciudades con características muy distintas: Quito, Melbourne y Oslo. Se demostró cómo la variación climática impacta directamente en la entropía de las señales y en la tasa de compresión de los algoritmos aplicados.

Quito, con un clima estable, presentó baja entropía y permitió una alta tasa de compresión, mientras que Melbourne, con mayor variabilidad térmica, mostró una entropía más alta y una menor tasa de compresión. Oslo se posicionó en un punto intermedio. Además, se comprobó cómo el uso de modelos con memoria mejora tanto la predicción como la compresión de los datos.

Por otro lado, el análisis del canal de transmisión evidenció la importancia de la calidad del medio en la preservación de la información. Un canal ruidoso puede distorsionar los datos, reduciendo la información mutua y dificultando la interpretación correcta de la señal recibida.

Este trabajo permitió no solo aplicar herramientas matemáticas y computacionales, sino también comprender cómo los conceptos teóricos de la información, como la compresión, la codificación y el análisis de canales, se pueden implementar reforzando así la utilidad práctica de la teoría estudiada.