

Métodos Numéricos

Departamento de Computación
Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires

9 de junio de 2017



Hasta ahora...

- ▶ Generación de rankings deportivos
- ▶ Reconstrucción de imágenes eliminando ruido
- ▶ Reconocimiento facial
- ▶ Clasificación de noticias
- ▶ PageRank
- ▶ Generación de imágenes de tomografía computada
- ▶ Hoy: generación de modelos climáticos

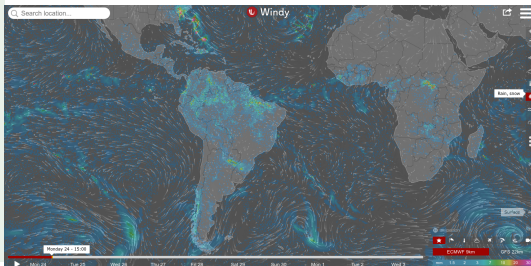
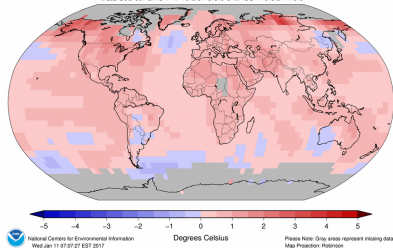
Motivación

Predicciones climáticas



Land & Ocean Temperature Departure from Average Jan–Dec 2016
(with respect to a 1981–2010 base period)

Data Source: GHCN–M version 3.3.0 & ERSST version 4.0.0



Pronósticos de clima

- ▶ Generación de pronósticos a partir de datos de la realidad: temperatura, presión, dirección y velocidad del viento, humedad, nivel de nubosidad, precipitaciones, etc.
- ▶ Los modelos más complejos utilizan ecuaciones diferenciales que expresan diversas variables
- ▶ En este trabajo haremos análisis y predicciones de las temperaturas

Pronósticos de clima

- ▶ Generación de pronósticos a partir de datos de la realidad: temperatura, presión, dirección y velocidad del viento, humedad, nivel de nubosidad, precipitaciones, etc.
- ▶ Los modelos más complejos utilizan ecuaciones diferenciales que expresan diversas variables
- ▶ En este trabajo haremos análisis y predicciones de las temperaturas

Pronósticos de clima

- ▶ Generación de pronósticos a partir de datos de la realidad: temperatura, presión, dirección y velocidad del viento, humedad, nivel de nubosidad, precipitaciones, etc.
- ▶ Los modelos más complejos utilizan ecuaciones diferenciales que expresan diversas variables
- ▶ En este trabajo haremos análisis y predicciones de las temperaturas

Pronósticos de clima

- ▶ Generación de pronósticos a partir de datos de la realidad: temperatura, presión, dirección y velocidad del viento, humedad, nivel de nubosidad, precipitaciones, etc.
- ▶ Los modelos más complejos utilizan ecuaciones diferenciales que expresan diversas variables
- ▶ En este trabajo haremos análisis y predicciones de las temperaturas

¿Analytics?, ¿para qué?

También Big Data o Data Science

1. Observar los datos
2. Analizarlos y entenderlos
3. Tomar decisiones



En este TP: modelar los patrones o tendencias vistos en la temperatura de ciudades o países con el fin de predecir comportamientos futuros

¿Analytics?, ¿para qué?

También Big Data o Data Science

1. Observar los datos
2. Analizarlos y entenderlos
3. Tomar decisiones

En este TP: modelar los patrones o tendencias vistos en la temperatura de ciudades o países con el fin de predecir comportamientos futuros

Datos

- ▶ Temperatura promedio del planeta por año
- ▶ Temperatura promedio por país por mes
- ▶ Temperatura promedio por ciudad por mes (junto con la latitud y longitud de la ciudad)
- ▶ Los que quieran usar ustedes relacionados a los anteriores

TP3: El problema

- ▶ Utilizar Cuadrados Mínimos Lineales (CML) como técnica e identificar modelos que describan la temperatura
- ▶ Realizar análisis y pronóstico de la temperatura del planeta
 - ▶ En función de la temperatura de los países
 - ▶ Utilizando expresiones matemáticas calculadas usando el año
- ▶ Realizar análisis y pronóstico para una ciudad a partir de otras ciudades o países
 - ▶ Considerar la posición geográfica o la distancia al mar o altitud, etc.
 - ▶ Minimizar la cantidad de features usados
- ▶ Se debe proponer alguna variante original para el análisis
 - ▶ Por ejemplo, podría buscarse explicar cierto fenómeno climático o predecir las temperaturas de ciudades (o regiones) similares en sus características pero distantes geográficamente

TP3: El problema

- ▶ Utilizar Cuadrados Mínimos Lineales (CML) como técnica e identificar modelos que describan la temperatura
- ▶ Realizar análisis y pronóstico de la temperatura del planeta
 - ▶ En función de la temperatura de los países
 - ▶ Utilizando expresiones matemáticas calculadas usando el año
- ▶ Realizar análisis y pronóstico para una ciudad a partir de otras ciudades o países
 - ▶ Considerar la posición geográfica o la distancia al mar o altitud, etc.
 - ▶ Minimizar la cantidad de features usados
- ▶ Se debe proponer alguna variante original para el análisis
 - ▶ Por ejemplo, podría buscarse explicar cierto fenómeno climático o predecir las temperaturas de ciudades (o regiones) similares en sus características pero distantes geográficamente

TP3: El problema

- ▶ Utilizar Cuadrados Mínimos Lineales (CML) como técnica e identificar modelos que describan la temperatura
- ▶ Realizar análisis y pronóstico de la temperatura del planeta
 - ▶ En función de la temperatura de los países
 - ▶ Utilizando expresiones matemáticas calculadas usando el año
- ▶ Realizar análisis y pronóstico para una ciudad a partir de otras ciudades o países
 - ▶ Considerar la posición geográfica o la distancia al mar o altitud, etc.
 - ▶ Minimizar la cantidad de features usados
- ▶ Se debe proponer alguna variante original para el análisis
 - ▶ Por ejemplo, podría buscarse explicar cierto fenómeno climático o predecir las temperaturas de ciudades (o regiones) similares en sus características pero distantes geográficamente

TP3: El problema

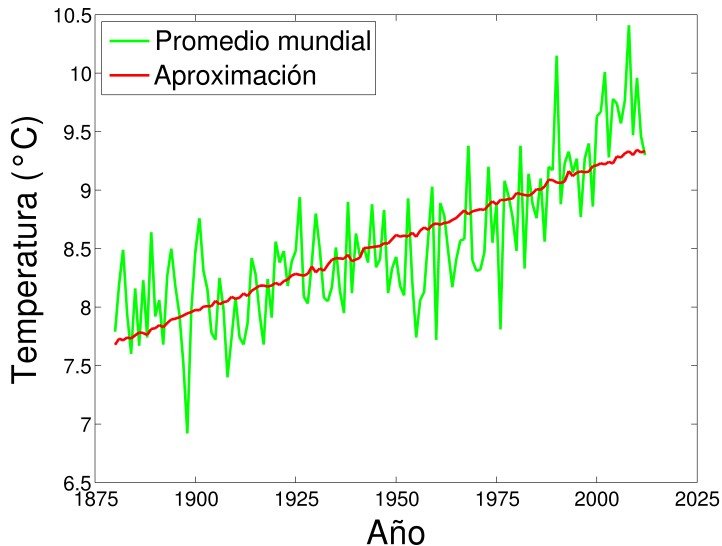
- ▶ Utilizar Cuadrados Mínimos Lineales (CML) como técnica e identificar modelos que describan la temperatura
- ▶ Realizar análisis y pronóstico de la temperatura del planeta
 - ▶ En función de la temperatura de los países
 - ▶ Utilizando expresiones matemáticas calculadas usando el año
- ▶ Realizar análisis y pronóstico para una ciudad a partir de otras ciudades o países
 - ▶ Considerar la posición geográfica o la distancia al mar o altitud, etc.
 - ▶ Minimizar la cantidad de features usados
- ▶ Se debe proponer alguna variante original para el análisis
 - ▶ Por ejemplo, podría buscarse explicar cierto fenómeno climático o predecir las temperaturas de ciudades (o regiones) similares en sus características pero distantes geográficamente

TP3: El problema

- ▶ Analizar los resultados y evaluar los modelos usando cross validation
- ▶ Aplicar las técnicas y metodologías aprendidas en la materia
- ▶ Libertad para plantear los modelos y para utilizar los datos pero rigurosidad en su justificación y análisis

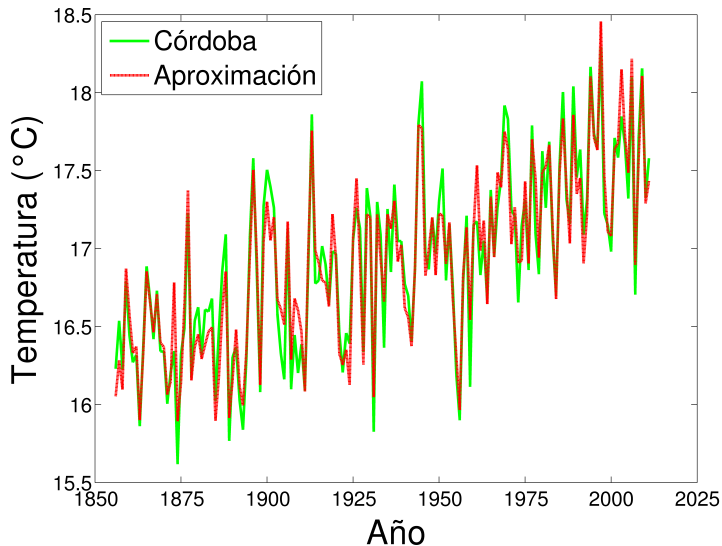
Predicción de la temperatura del planeta

Usando Argentina, Sudáfrica y Rumania + constante + lineal



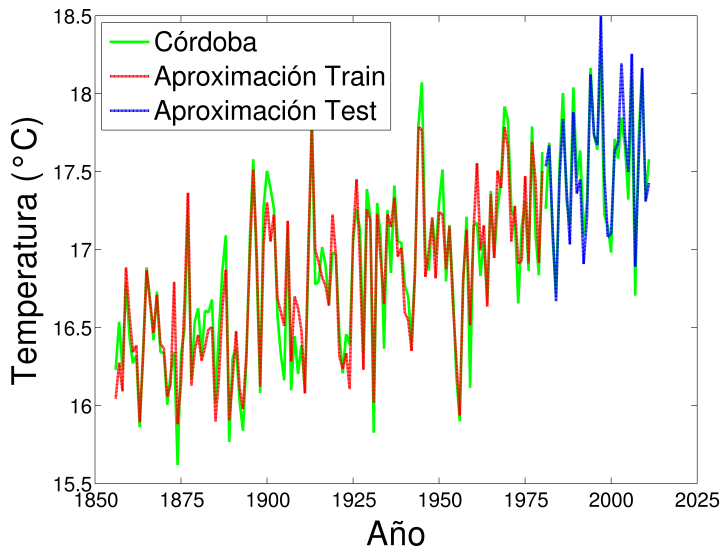
Predicción de ciudad

Usando Bahía Blanca, Catamarca, Corrientes, Bariloche



Predicción de ciudad (train y test)

Usando Bahía Blanca, Catamarca, Corrientes, Bariloche



Usando CML

- ▶ Considerar los datos disponibles y elegirlos con criterio
- ▶ Considerar no sólo datos sino eventualmente funciones aplicadas sobre los mismos (i.e. senos, polinomios, etc.)
- ▶ Considerar datos de distintos períodos y evaluar sobre períodos futuros

Métricas de evaluación (1/2)

- ▶ N observaciones $(x_{(i)}, y_{(i)})$, con $x_{(i)} \in \mathbb{R}^k$ el vector de *features* e $y_{(i)} \in \mathbb{R}$ nuestra variable dependiente.
- ▶ Suponemos $y_{(i)} = f(x_{(i)}) + \epsilon_i$, $i = 1, \dots, N$, donde ϵ_i es el error de la medición i -ésima.
- ▶ Dado un modelo \hat{f} de f y $(x_{(i)}, y_{(i)})$, definimos $\hat{y}_{(i)} = \hat{f}(x_{(i)})$ y $e_{(i)} = y_{(i)} - \hat{y}_{(i)}$. Con estas definiciones, podemos calcular el MSE del modelo \hat{f} como

$$MSE(\hat{f}) = \frac{1}{N} \sum_{i=1}^N e_{(i)}^2.$$

- ▶ Para evaluar cómo se comporta como modelo predictivo, podemos usar *Cross-Validation* combinado con MSE:

Métricas de evaluación (2/2)

- Consideramos que cada observación está asociada a un determinado período de tiempo t , con $t = 1, \dots, T$, $(x_{(i)}^t, y_{(i)}^t)$, y asumimos que al menos K períodos de tiempo son necesarios para poder conformar el conjunto de *training*. Para evaluar los resultados de la predicción en el período $\tau \in [K, T]$ se puede:
 1. Tomar los conjuntos de observaciones correspondientes a períodos $1, \dots, \tau - 1$ como training
 2. Calcular las métricas correspondientes tomando como test el período τ
 3. Al finalizar, reportar alguna medida sobre los resultados parciales obtenidos

Posibles análisis

- ▶ Evaluar correlación entre variables
- ▶ Considerar las posiciones geográficas (posibilidad de utilizar la latitud y longitud de las ciudades u otro indicador propuesto por el grupo)
- ▶ Evaluar distintos períodos de tiempo (para train y test)
- ▶ Considerar la disponibilidad de datos (no todos los países/ciudades tienen datos en los mismos períodos)
- ▶ Posibles niveles de granularidad

Desarrollo

- ▶ El TP se puede hacer en MATLAB, Python y/o C++. Se pueden usar bibliotecas con los métodos implementados.
- ▶ Junto al enunciado se presentan scripts en bash que extraen algunos datos de los archivos con la información de las temperaturas. Estos scripts pueden resultar útiles como inspiración para plantear sus propias herramientas. Éstas pueden ser compartidas entre ustedes, siempre y cuando se mantenga reservada la información de los experimentos. Deben hacerlo a través de la lista de alumnos.

Presentación

- ▶ Además del informe usual (esta vez bajo otro formato), el trabajo será expuesto en una presentación oral frente a alumnos y docentes.
- ▶ Para ello, harán una presentación mostrando lo que hicieron. Vamos a dar soporte y ayuda para prepararla.
- ▶ La exposición contará con 30 minutos totales: 15 minutos para presentar, 15 minutos de preguntas y respuestas. La nota de aprobación es individual.
- ▶ Para poder presentar en primera fecha, se deberá tener la aprobación previa por parte de los docentes correctores.

Trabajo Práctico

Fecha de entrega

- ▶ *Formato Electrónico*: Lunes 26/6, hasta las 23:59 hs.
- ▶ *Confirmación presentación oral*: Miércoles 28/6, por correo electrónico.
- ▶ *Presentación oral*: Lunes 3/7, en horario a determinar luego de la confirmación. Será en horario de clase de la materia.

Importante

El horario es estricto. Los correos recibidos después de la hora indicada serán considerados re-entrega.