

Tarea 2

Profesor: Felipe Tobar

Auxiliares: N. Aramayo, L. Araya, M. Campos, A. Cuevas y C. Valenzuela

Consultas: Mauricio Campos (mcampos AT dim.uchile.cl)

Fecha entrega: 13/4/2018

Formato entrega: Entregue un informe en formato PDF con una extensión de a lo más 2 páginas presentando y analizando sus resultados, detalle la metodología utilizada y adicionalmente debe entregar un jupyter notebook con los códigos que creó para resolver la tarea.

P1. Regresión No Lineal (7 puntos)

En el archivo **datosT2.txt** están los datos de la cantidad de pasajeros de una aerolínea medidos de forma mensual. Los datos son de la forma $\{(x_i, y_i)\}_{i=1}^N$ donde x_i representa un mes, y_i la cantidad de pasajeros transportados en el mes correspondiente.

El objetivo de esta tarea es modelar la cantidad de pasajeros (y) respecto al instante de tiempo (x), para este fin, se asumirá el siguiente modelo:

$$y = f_{\theta}(x) + \eta$$

Donde θ corresponde a los parámetros de la función f los cuales debe ajustar y $\eta \sim \mathcal{N}(0, \sigma_{\eta}^2)$ corresponde a ruido gaussiano.

Para la parte que siguen, considere 75% de los datos (los primeros 9 años) para entrenamiento de su modelo y el 25% restante para validar sus resultados.

Para esto deberá:

- (i) Cargue los datos en **datosT2.txt** y grafíquelos de forma que se pueda distinguir entre el conjunto de entrenamiento y validación.
- (ii) (1 punto) Modele f_{θ} como un polinomio, es decir:

$$f_{\theta}(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \dots$$

Considere un prior gaussiana sobre los parámetros y encuentre los parámetros usando el método maximum a posteriori.

En base a los resultados obtenidos indique y discuta cual es el grado del polinomio que es más probable que haya generado los datos.

HINT: Considere polinomios entre grado 1 y 4.

- (iii) (3 puntos) Como podrá observar el modelo polinomial no es capaz de capturar las componentes periódicas observadas, para resolver esto, denote el polinomio encontrado en la parte anterior como f^{pol} y modele la señal como f^{pol} más una componente sinusoidal modulada por una exponencial, es decir:

$$y = f_{\theta}(x) = f^{pol} + \theta_1 \sin(\theta_2 x + \theta_3) e^{\theta_4 x} + \eta, \quad \eta \sim \mathcal{N}(0, \sigma_{\eta}^2)$$

Solo ajuste los parámetros $(\theta_{1:4})$ del modelo mediante máxima verosimilitud, note que este modelo no es lineal en los parámetros por lo cual la solución no puede ser escrita de forma exacta. Para ello construya la función de verosimilitud y optimícela usando el método BFGS. Recuerde que no es necesario calcular analíticamente el gradiente de la función a optimizar dependiendo de la implementación de BFGS utilice.

Discuta la capacidad del modelo para predecir el conjunto de evaluación e interprete los parámetros del modelo.

HINT: Utilice 0.01 como condición inicial para θ_4 .

- (iv) (3 puntos) La 'forma de onda' de los datos no es puramente sinusoidal, por esta razón, denote la función encontrada en la parte anterior como $f^{pol-sin}$ y agregue una segunda componente senoidal modulada por una exponencial, es decir:

$$y = f_{\theta}(x) = f^{pol-sin} + \theta_1 \sin(\theta_2 x + \theta_3) e^{\theta_4 x} + \eta, \quad \eta \sim \mathcal{N}(0, \sigma_{\eta}^2)$$

Recuerde los parámetros de $f^{pol-sin}$ son fijos y encuentre los nuevos parámetros $\theta_{1:4}$ usando máxima verosimilitud y BFGS. Evalúe su solución en la prediciendo el 25% restante de los datos.

HINT: Utilice 0.01 como condición inicial para θ_4 .

Presente sus resultados y discuta el método en que los obtuvo, en particular, explique el rol de cada una de las componentes del modelo final. ¿Habría sido posible entrenar el modelo final de una vez? Evalúe los modelos ajustados en función de su verosimilitud y del error de predicción en el conjunto de evaluación.

P2. Proyecto curso (3 puntos)

- (i) Para el proyecto del curso, formen grupos de 3 personas y definan cual va a ser el tema de proyecto que van a desarrollar. Luego explique y muestre los datos que piensan utilizar.

Los alumnos de postgrado deben definir cual será su tema de proyecto individualmente, explicar y mostrar los datos que piensan utilizar.