

Tarea 2

Profesor: Felipe Tobar

Auxiliares: Mauricio Araneda, Alejandro Cuevas, Mauricio Romero

Consultas: Alejandro Cuevas

Fecha entrega: 26/4/2019

Formato entrega: Entregue un informe en formato PDF con una extensión de a lo más **3** 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

Se tienen 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. Los datos los puede encontrar en:

<https://github.com/GAMES-UCHile/Curso-Aprendizaje-de-Maquinas/blob/master/datos/datosT2.txt>

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 , y $\eta \sim \mathcal{N}(0, \sigma_{\eta}^2)$ corresponde a ruido gaussiano. Luego, los parámetros a ajustar se denotan:

$$\theta' = [\theta, \sigma_{\eta}^2]$$

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 **datosT2.txt** y gráfíquelos de forma que se pueda distinguir entre el conjunto de entrenamiento y test.
- (ii) Modele f_{θ} como un polinomio, es decir:

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

Considere un prior Gaussiano sobre los parámetros θ' y encuentre dichos parámetros usando el método *máximo 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) 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 $(\theta_{1:4})$ y el ruido σ_{η}^2 del modelo mediante máxima verosimilitud, es decir, mantenga sin modificar la componente polinomial f^{pol} del inciso anterior.

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) Como podrá notar, 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:

$$y = f_{\theta}(x) = f^{pol-sin} + \theta_5 \sin(\theta_6 x + \theta_7) e^{\theta_8 x} + \eta, \quad \eta \sim \mathcal{N}(0, \sigma_{\eta}^2)$$

Recuerde los parámetros de $f^{pol-sin}$ son fijos y encuentre los nuevos parámetros $\theta_{5:8}$ y σ_{η}^2 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_8 .

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, es decir el modelo que considera la parte polinomial y ambas componentes sinusoidales. Además, comente sobre los distintos estimados de la varianza del ruido σ_{η}^2 en cada etapa.

¿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 test.

P2. Proyecto curso

- (i) Forme grupos de a lo más **3** personas (Si es de postgrado, máximo 1). Como grupo elijan una propuesta de proyecto y descríballo.
- (ii) Sobre el proyecto que piensa realizar, muestre y describa los datos que piensa utilizar.