

Tarea 3 - Procesos Gaussianos

Camilo Carvajal Reyes

21 de noviembre de 2022

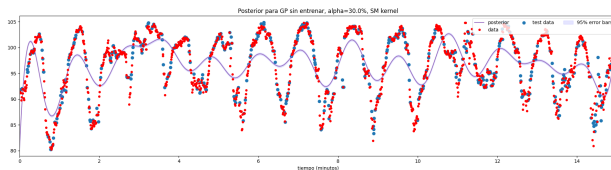
En la presenta tarea se presentarán aplicaciones de Procesos Gaussianos usando diferentes Kernels y variando los respectivos parámetros. Se usará la primera serie del conjunto de **pulsos sanguíneos del MIT**, que parece tener arritmia sinusal respiratoria. Se remueven 30 % de los datos, los cuales se intentan predecir con nuestro proceso ajustado al otro 70 % de los datos. Además se usan las siguientes métricas de evaluación: log-versimilitud negativa (para el conjunto de entrenamiento, i.e., aquella usada para entrenar los hiper-parámetros, y para el conjunto de puntos no-observados) y el error cuadrático medio entre la media y los valores reales (de puntos no observados). Para las métricas en cuestión se usa el posterior computado sobre aquellos puntos de interés.

1. *Kernel Spectral Mixture*

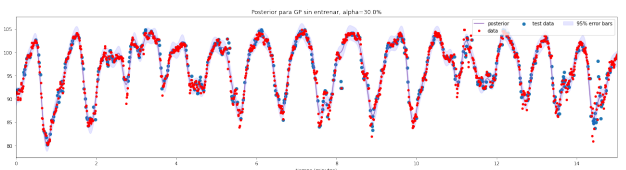
Consideramos primeramente el kernel de Mixtura Espectral dado por:

$$K(x) = \sigma^2 \exp(-2\pi^2 \gamma^2 x^2) \cos(2\pi \mu x).$$

Además poseemos un parámetro μ_n que puede interpretarse como un nivel de confianza en la medición. Para esto, usamos el toolkit **gp_lite**, en la cual este kernel viene implementado. Usar los parámetros por defecto ($\sigma = 10, \gamma = 0.5, \mu = 0.1, \sigma_n = 0.1$) nos da un proceso como el mostrado en la figura 1 (a). Acá nos movemos en el dominio de los minutos. Cuando la serie la definimos en el dominio de los minutos. Vale señalar que los resultados por defecto varían drásticamente al cambiar el dominio inicial (no así luego de entrenar). En la figura 1 (b) tenemos una actualización manual de parámetros ($\sigma = 10, \gamma = 10, \mu = 0.75, \sigma_n = 0.75$) que muestra un ajuste razonable a nuestra serie original.



(a) Parámetros por defecto



(b) Parámetros manuales

Figura 1: Posterior para GP sin entrenar

En lo que sigue se realiza un análisis cualitativo de los parámetros del kernel SM. Primero notemos que la variación de μ se refleja en una frecuencia de oscilación, lo cual se condice con su rol dentro del término coseno. En el gráfico 2 (a) podemos ver el efecto de un valor exageradamente grande, para luego ver un efecto más suave en la parte 2 (b). Por otro lado, las variaciones de el parámetro *sigma* se manifiestan en menores oscilaciones para parámetros pequeños (no consideramos pertinente incluirlos en el informe). Esto es similar para γ , en cuyo caso un valor pequeño se manifiesta en un bajo ajuste a los datos observados (figura 2 (c)). Esto contrasta con el sobre-ajuste que se observa en la figura 2 (d). Esto puesto que el término que multiplica x es demasiado alto y por ende domina el efecto del kernel (con esto no cumple realmente su rol suavizador).

Como parámetro aparte, σ_n , como es de esperar, es proporcional a la varianza del proceso Gaussiano, con lo cual se admiten posibles más puntos alrededor de la media del posterior. Esto se observa en la sub-figura 1 (b) anterior.

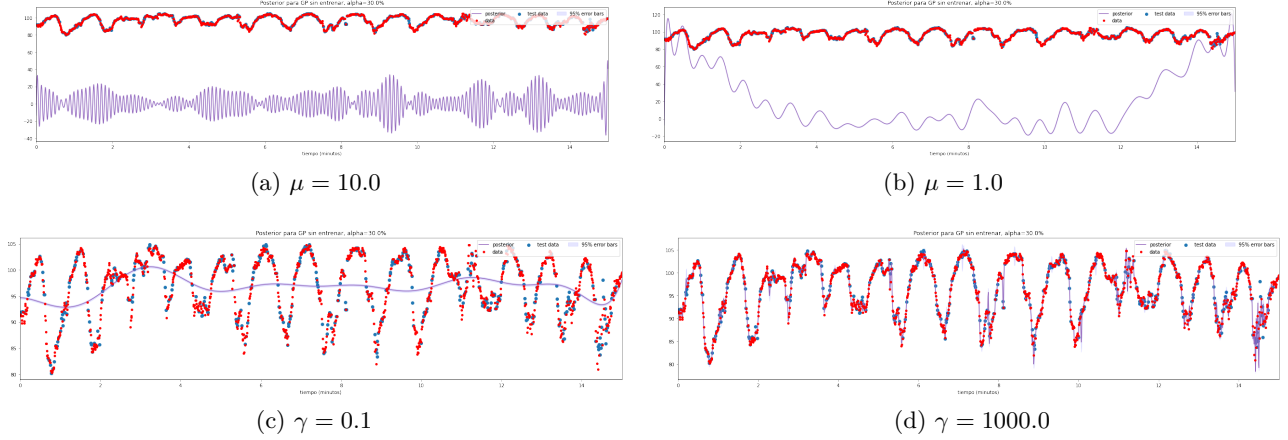


Figura 2: Parámetros variados respecto a kernel por defecto

Al minimizar la log-verosimilitud negativa, logramos obtener los parámetros óptimos. Esto se manifiesta en un Proceso Gaussiano ajustado como se visualiza en la figura 3. En este caso la combinación óptima encontrada es: $\sigma = 61.644$, $\gamma = 29.694$, $\mu = 0.012$ y $\sigma_n = 0.919$.

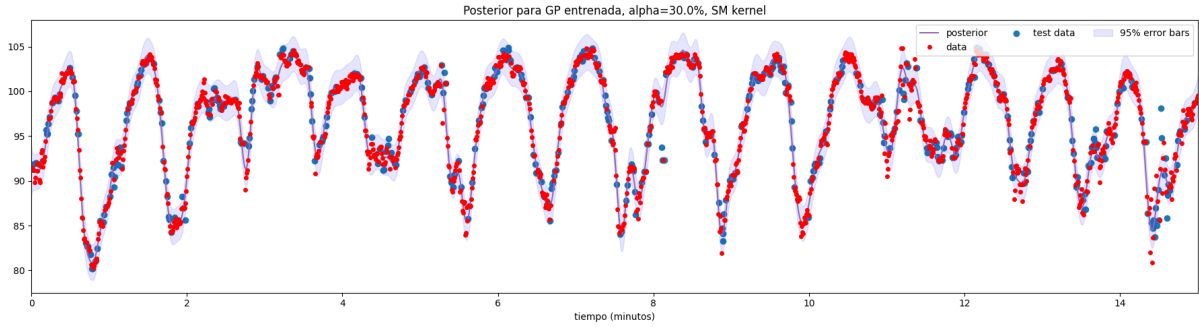


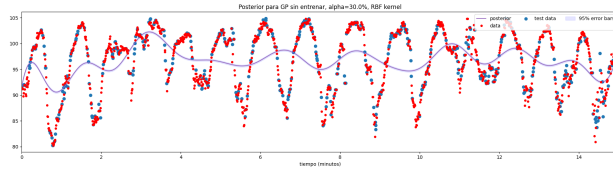
Figura 3: Posterior para GP entrenada

2. Kernel *Radial Basis Function*

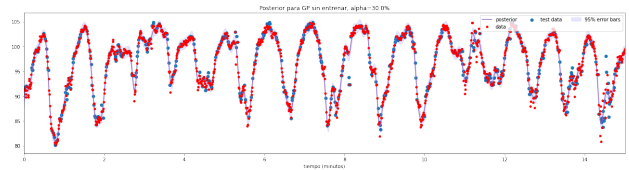
En esta sección abordamos el kernel Función de base radial o RBF por sus siglas en inglés. También llamado exponencial cuadrática ya que está dado por:

$$K(x) = \sigma^2 \exp\left(-\frac{x^2}{2\gamma^2}\right).$$

Es el kernel “por defecto” en algunas implementaciones, lo cual se debe a sus buenas propiedades. Además, comparte similitudes con el kernel de mixtura espectral, sin embargo su expresión es simplificada. Como primera hipótesis, como para nuestra serie de arritmia funcionaron mejor valores de μ pequeños, se conjetura que una forma de kernel como la RBF bastaría para modelar el problema, ya que en Spectral Mixture prácticamente se ignoró el término de coseno.



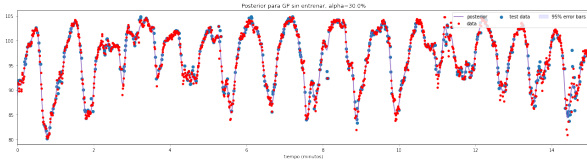
(a) Parámetros por defecto



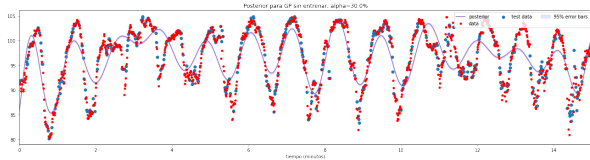
(b) Parámetros manuales

Figura 4: Posterior para GP con kernel RBF sin entrenar

Primero notemos que su forma por defecto, que en este caso consideramos como $\sigma = 1.0$, $\gamma = 1.0$ y $\sigma_n = 0.1$, es más suave, como se vislumbra en 4 (a). De manera análoga, podemos “al ojo” encontrar buenos parámetros, que se ajustan más o menos bien a nuestra serie ($\sigma = 5.0$, $\gamma = 0.1$ y $\sigma_n = 0.5$). Esto ya nos va dando una idea del rol de cada parámetro. Por un lado un γ más bien pequeño será como tomar más en cuenta los puntos de la medición. Esto se ve claramente en la figura 5 (a), donde un valor pequeño de γ genera un sobreajuste sobre las observaciones. Por otro lado, el término de escalamiento nos hace acercarnos a la serie original al crecer como se ve en 5 (b), pese a dejar $\gamma = 1$ fijo.



(a) $\gamma = 0.1$



(b) $\sigma = 50.0$

Figura 5: Parámetros variados en Kernel RBF

Finalmente, la minimización de la log-verosimilitud negativa nos entrega los parámetros: $\sigma = 57.445$, $\gamma = 0.1569$ y $\sigma_n = 1.033$. La figura 6 nos sugiere que los resultados son muy similares a lo logrado con el kernel mixtura espectral. Acá ganamos en complejidad, pues tenemos menos términos y por ende una convergencia más rápida a los valores apropiados (la mitad de iteraciones de L-BFGS-B aproximadamente). En cuanto a las métricas usadas, el kernel de mixtura espectral posee una mejor log-verosimilitud negativa. Sin embargo, el kernel RBF presenta un menor error cuadrático medio, alcanzado (curiosamente) por la versión con parámetros encontrados de modo manual (todas reportadas respecto al conjunto de test).

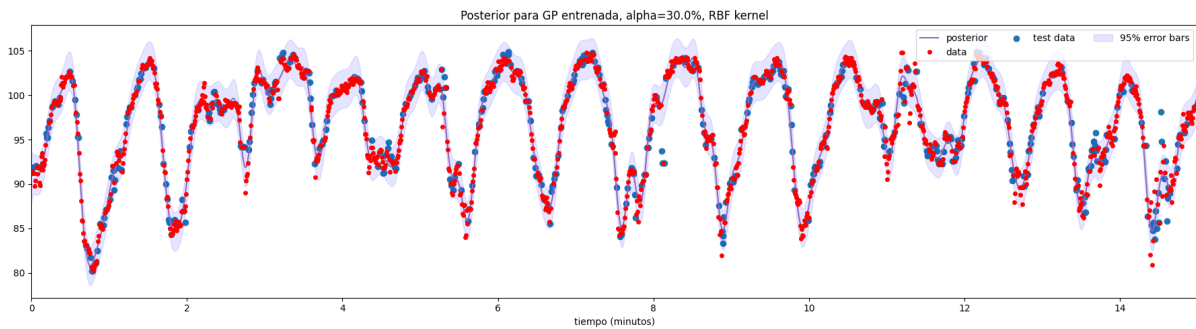


Figura 6: Posterior para GP con kernel RBF entrenada

	SM def	SM manual	SM trained	RBF def	RBF manual	RBF trained
NLL	35102	1236	821	1873414	1255	874
MSE	19.39	2.05	1.25	26.64	1.23	1.52