Abstract:
Simulación y Estimación Bayesiana de Procesos Gaussianos

M. Cabezas & A. Kolm

1. Introducción

En este trabajo se estudiarán los procesos Gaussianos o $\mathcal{GP}$, investigando cómo se comportan, cómo se simulan y sus usos en Inferencia Bayesiana y Aprendizaje de Máquinas.
Brevemente, un $\mathcal{GP}$ se define como una distribución de probabilidad en un espacio de funciones. La particularidad de los $\mathcal{GP}$ es que dado una función de media $m(\cdot)$ y una función kernel $K(\cdot,\cdot)$, definidas en $\mathcal{X}$ y $(\mathcal{X} \times \mathcal{X})$ (respectivamente,) con $\mathcal{X}$ un conjunto cualquiera (por ejemplo $\mathcal{X} = R$) es posible generar una función como un ”vector” normal multivariado de dimensión arbitraria. Por ejemplo, en Figura 1 se muestra (1) un vector que distribuye como una normal multivariada de dimensión 10 y (2) una función (vector de dimensión infinita) que distribuye como un $\mathcal{GP}$.

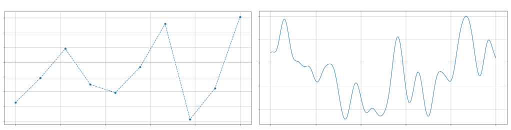

Figure 1. (1) $x \sim \mathcal{N}(\mu, \Sigma)$; (2): $f \sim \mathcal{GP}(m(\cdot), K(\cdot,\cdot))$

2. Kernels

Muchos procesos estocásticos tienen una tendencia clara, como por ejemplo los procesos estacionarios o con tendencia a crecer; en dichos procesos la función media es la que da esta característica. Sin embargo, la aleatoriedad de los GP está encapsulada en los Kernels.
Los kernels pueden verse como una matriz de covarianza de dimensión arbitraria, y es lo que controla la regularidad de las funciónes que viven en un GP dado. Por ejemplo, el movimiento brawniano (GP estudiado en el curso) tiene el kernel $K(s,t) = s \wedge t$, mientras que si se define el kernel $K_{SE}(s,t) = \sigma^2 exp\left(\frac{-(X_s-X_t)^2}{2l^2}\right)$ obtenemos funciones como la de la Figura 2.

3. Simulación

Como un computador no puede representar una función continua, se debe discretizar para obtener una aproximación de un $\mathcal{GP}$.Dado que existen infinitas funciones que pasan por n puntos (pues n es finito), lo importante es tener en cuenta que como $\mathcal{GP}$ es una distribucion de probabilidad en el espacio de funciones, hay funciones con mayor probabilidad de pasar por estos puntos que otras. Esto está definido por la funcion media y el kernel.

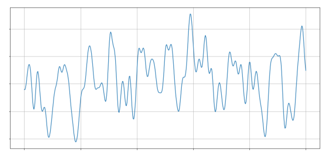

Figure 2. $f \sim \mathcal{GP}(0, K_{SE}(\cdot,\cdot))$

4. Inferencia Bayesiana

Gracias al teorema de Bayes, es posible ajustar $\mathcal{GP}$ a datos obtenidos en la vida real para obtener un modelo mediante regresión, con el fin de predecir o clasificar. Tipicamente, se definen Priors sobre los datos, como el tipo de funcion de media, kernel y los hiperparámetros que en ellos se puedan encontrar, para luego optimizar la función de verosimilitud $p(modelo|Datos)$.

Por último, se mostrarán dos resultados de inferencia bayesiana: (1) regresión sobre datos simulados y (2) regresión sobre datos reales obtenidos de una base de datos externa (cantidad de pasajeros en función del tiempo).

5. Referencias

1. Apunte del curso MA5204-1 Aprendizaje de máquinas: https://github.com/GAMES-UChile/Curso-Aprendizaje-de-Maquinas/blob/master/notas_de_clase.pdf

2. C. E. Rasmussen & C. K. I. Williams, Gaussian Processes for Machine Learning, the MIT Press, 2006, Massachusetts Institute of Technology. www.GaussianProcess.org/gpml

3. Duvenaud, D. (2014). Automatic model construction with Gaussian processes (Doctoral dissertation, University of Cambridge). https://www.cs.toronto.edu/ duvenaud/cookbook/