Gaussian process models

Chapter 21

- Properties of Multivariate Gaussian Distributions
- Gaussian Process Models

Once Gaussian, Always Gaussian

We first review the definition and properties of Gaussian distribution:
A Gaussian random variable $X \sim \mathcal{N}(\mu, \Sigma)$, where μ is the mean, and Σ is the covariance matrix.

$$P(x; \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{1/2}} e^{-\frac{1}{2}((x-\mu)^{\top} \Sigma^{-1} (x-\mu)}$$

Where $|\Sigma|$ is the determinant of Σ

The Gaussian distribution occurs very often in real world data. As the Central Limit Theorem (CLT) states that the arithmetic mean of m>0 samples is approximately normal distributed - independent of the original sample distribution (provided it has finite mean and variance).

Once Gaussian, Always Gaussian

Let Gaussian random variable $y = \begin{bmatrix} y_A \\ y_B \end{bmatrix}$, mean $\mu = \begin{bmatrix} \mu_A \\ \mu_B \end{bmatrix}$ and covariance matrix $\Sigma = \begin{bmatrix} \Sigma_{AA}, \Sigma_{AB} \\ \Sigma_{BA}, \Sigma_{BB} \end{bmatrix}$

We have the following properties:

- Normalization: $\int_y p(y; \mu, \Sigma) dy = 1$
- Marginalization:

$$y_A \sim \mathcal{N}(\mu_A, \Sigma_{AA})$$

$$y_B \sim \mathcal{N}(\mu_B, \Sigma_{BB})$$

- Summation: If $y \sim \mathcal{N}(\mu, \Sigma)$ and $y' \sim \mathcal{N}(\mu', \Sigma')$, then

$$y + y' \sim \mathcal{N}(\mu + \mu', \Sigma + \Sigma')$$

- Conditioning:

$$y_A | y_B \sim \mathcal{N}(\mu_A + \Sigma_{AB}\Sigma_{BB}^{-1}(y_B - \mu_B), \Sigma_{AA} - \Sigma_{AB}\Sigma_{BB}^{-1}\Sigma_{BA}).$$

Gaussian Process - Definition

- **Definition**: Gaussian Process (GP) is a collection of random variables (RV) such that the joint distribution of every finite subset of RVs is multivariate Gaussian:

  $$f \sim GP(\mu, K)$$

  where $\mu(x)$ and $K(x, x')$ are the mean and covariance function.
- Now, in order to model a distribution, we can use a Bayesian approach with a GP prior: $f \sim GP(\mu, K)$

Broad Applications

- Regression:
  1. Surrogate surfaces for optimization or simulation
  2. Function estimation
- Classification:
  1. Recognition: e.g. handwritten digits on cheques
  2. Filtering: fraud, interesting science, disease screening
- Ordinal regression
  1. User ratings (e.g. movies or restaurants)
  2. Disease screening