CS 419

Lecture 16: Gaussian Processes

17 March 2023

Lecturer: Abir De *Scribe: Aditya, Jay, Soham & Sooraj*

Till now we had studied about kernels and their properties. In the last lecture, we learned about Gaussian Processes(GP).

1 Recap of the last lecture

Let's begin with a brief introduction to Gaussian processes.

Gaussian Processes are a class of probabilistic models that can be used for supervised machine-learning tasks such as regression and classification. GPs are a non-parametric approach, meaning they do not assume any particular functional form for the relationship between the inputs and outputs.

In a GP, a function is modeled as a probability distribution over functions. This distribution is defined by a mean function and a covariance function. The mean function specifies the expected value of the function at each input point, while the covariance function specifies how much the function values at different input points are correlated.

During training, the GP is fitted to the training data by adjusting the hyper-parameters to maximize the likelihood of the observed data. Once the GP is trained, it can be used to make predictions for new inputs by computing the posterior distribution over functions given the observed data.

2 Gaussian Processes and Multivariate Gaussian Distribution

A multivariate Gaussian distribution is a probability distribution that describes the joint probability distribution of a set of random variables that are normally distributed. It is an extrapolation of the univariate Gaussian distribution to higher dimensions, where the mean vector and covariance matrix fully specify the distribution.

The multivariate Gaussian distribution is widely used in statistics, machine learning, and many other fields, due to its flexibility and tractability. It is used in applications such as clustering, classification, regression and data analysis among others.

2.1 2-dimensional Gaussian Distribution

$$y \sim \mathcal{N}(\mu, \Sigma) \tag{1}$$

where y and μ are two-dimensional vectors and Σ is 2x2 matrix given by:

$$\Sigma = \begin{bmatrix} \Sigma_{AA} & \Sigma_{AB} \\ \Sigma_{BA} & \Sigma_{BB} \end{bmatrix} \tag{2}$$

where Σ_{aa} represents the variance of a and Σ_{ab} is a co-variance of a and b.

$$f(x) = \frac{1}{\sqrt{2\pi|\Sigma|^2}} e^{-\frac{(x-\mu)^T \Sigma^{-1} (x-\mu)}{2}} \tag{3}$$

2.2 Conditional Gaussian Distribution

We have two random variables X and Y, and we know that Y is dependent on X, then the conditional distribution of Y given X can be represented as a Gaussian distribution.

$$Y_A | Y_B \sim \mathcal{N}(\mu'_A, \Sigma_{A|B}) \tag{4}$$

Now, let's see the effect of conditioning on the mean of a distribution:
If we observe one variable,
let's say Y_B = Y_0, then Y_B is a distribution with mean Y_0 and variance 0.
Start with the intuition:

$$\mu'_A = \mu_A + \Sigma_{AB}\Sigma_{BB}^{-1}(Y_0 - \mu_B) \tag{5}$$

Verification of our intuition:

- If $Y_B = \mu_B$, then the mean of Y_A won't change which can be seen in the equation. $Y_0 = \mu_B$ then $\mu'_A = \mu_A + 0$.
- Substituting A with B yields, $\mu'_B = \mu_B + \Sigma_{BB}\Sigma_{BB}^{-1}(Y_0 - \mu_B) = Y_0$
- If co-variance of Y_A and Y_B is zero, then after observation mean of Y_A doesn't change. $\Sigma_{AB} = 0$, so $\mu'_A = \mu_A$
- If Σ_{BB} is very large, it means that there is a large variation in Y_B. Hence, we shouldn't rely on the observed value of Y_B and thus, the mean remains unaffected.
- Now $\Sigma_{BB} \to 0$ which means that Y_B is a constant then $(Y_B - \mu_B) \to 0$

So, we have to calculate the limit for μ'_A,

$$\lim_{\Sigma_{BB}\to 0}(\mu'_A) = \mu_A + \lim_{\Sigma_{BB}\to 0}[\Sigma_{AB}\Sigma_{BB}^{-1}(Y_0 - \mu_B)] \tag{6}$$

Calculating the limit:

$$\mathbf{L} = \lim_{\Sigma_{BB} \to 0} [\Sigma_{AB}\Sigma_{BB}^{-1}(Y_0 - \mu_B)] \tag{7}$$

where, Σ_{AB} = $\mathbf{E}[(Y_A - \mu_A)(Y_B - \mu_B)]$, Σ_{BB} = $\mathbf{E}[(Y_B - \mu_B)^2]$.

Since variance is tending towards zero, $(Y_B - \mu_B)$ can be assumed as constant.

$$\mathbf{L} = \lim_{\Sigma_{BB} \to 0} \frac{(Y_B - \mu_B)^2 \mathbf{E}[Y_A - \mu_A]}{(Y_B - \mu_B)^2} \tag{8}$$

$$= \lim_{\Sigma_{BB} \to 0} \mathbf{E}[Y_A - \mu_A] = 0 \tag{9}$$

so,

$$\lim_{\Sigma_{BB} \to 0} (\mu'_A) = \mu_A + 0 = \mu_A \tag{10}$$

Now that we have observed the effect of conditioning on the mean of a distribution, we proceed with the variance of $Y_A|Y_B$. Intuitively, one can guess the variance to be the following:

$$\Sigma_{A|B} = \Sigma_{AA} - \Sigma_{AB}\Sigma_{BB}^{-1}\Sigma_{BA} \tag{11}$$

Verification of our intuition:

- If $A = B$, then the value of variance will be:

$$\Sigma_{B|B} = \Sigma_{BB} - \Sigma_{BB}\Sigma_{BB}^{-1}\Sigma_{BB} \tag{12}$$

$$= \Sigma_{BB} - \Sigma_{BB} I = 0 \tag{13}$$

  Hence, we obtain $\Sigma_{B|B} = 0$.

- When $\Sigma_{BB} \to \infty$:

$$\Sigma_{A|B} = \lim_{\Sigma_{BB} \to \infty} (\Sigma_{AA} - \Sigma_{AB}\Sigma_{BB}^{-1}\Sigma_{BA}) \tag{14}$$

Now, we need to calculate the limit for $\Sigma_{A|B}$,
Calculating the limit:

$$\mathbf{L} = \lim_{\Sigma_{BB} \to \infty} (\Sigma_{AA} - \Sigma_{AB}\Sigma_{BB}^{-1}\Sigma_{BA}) \tag{15}$$

where, Σ_{AB} = $\mathbf{E}[(Y_A - \mu_A)(Y_B - \mu_B)]$= Σ_{BA}, Σ_{BB} = $\mathbf{E}[(Y_B - \mu_B)^2]$.

$$\mathbf{L} = \lim_{(Y_B - \mu_B)^2 \to \infty} \left(\mathbf{E}[(Y_A - \mu_A)^2] - \frac{(\mathbf{E}[(Y_A - \mu_A)(Y_B - \mu_B)])^2}{\mathbf{E}[(Y_B - \mu_B)^2]} \right) \tag{16}$$

As $\mathbf{E}[(Y_B - \mu_B)^2]$ becomes arbitrarily large, we can conclude that:

$$\mathbf{L} = \lim_{(Y_B - \mu_B)^2 \to \infty} \left(\mathbf{E}[(Y_A - \mu_A)^2] - \frac{(\mathbf{E}[(Y_A - \mu_A)(Y_B - \mu_B)])^2}{\mathbf{E}[(Y_B - \mu_B)^2]} \right) = \mathbf{E}[(Y_A - \mu_A)^2] = \Sigma_{AA} \quad (17)$$

One more argument that can be made in the favour of the above result is that since Σ_{BB} is arbitrarily large Y_B is effectively a noise term and the distribution of $Y_A|Y_B$ is identical to that of Y_A. Thus, we have modeled the bi-variate Gaussian distribution in the following manner:

$$Y_A|Y_B \sim \mathcal{N}\left(\mu_A + \Sigma_{AB}\Sigma_{BB}^{-1}(Y_0 - \mu_B), \Sigma_{AA} - \Sigma_{AB}\Sigma_{BB}^{-1}\Sigma_{BA} \right) \quad (18)$$

Following is a plot of the joint, marginal and conditional distribution for a bivariate gaussian function which has been discussed above.

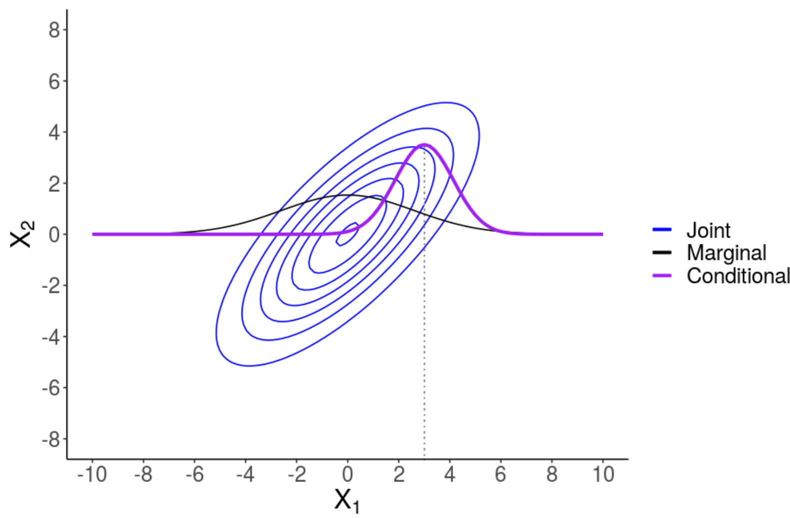

Figure 1: Joint, Marginal and Conditional Distribution for bivariate Gaussian distribution
source: https://fabiandablander.com/statistics/Two-Properties.html

We have used an extremely intuitive approach to modeling the bivariate Gaussian distribution, a more complete and rigorous derivation of the same can be found at `https://fabiandablander.com/statistics/Two-Properties.html`.