The Derivation Of The Posterior Of Gaussian Process

For a multivariate Gaussian random vector $X \in \mathbb{R}^w$

$$X \sim \mathcal{N}(\mu, \Sigma)$$

$$p_X(x; \mu, \Sigma) = p(X = x) = \frac{1}{det(\Sigma)\sqrt{(2\pi)^n}} e^{-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)}$$

Given a joint probability distribution

$$X \sim \mathcal{N}(\mu, K)$$

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \quad K = \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix}$$

$$X_1, \mu_1 \in \mathbb{R}^{n \times 1}, \quad X_2, \mu_2 \in \mathbb{R}^{m \times 1}$$

$$K_{11} \in \mathbb{R}^{n \times n}, K_{12} \in \mathbb{R}^{n \times m}, K_{21} \in \mathbb{R}^{m \times n}, K_{22} \in \mathbb{R}^{m \times m}$$

With Bayesian inference, we can inference the conditional probability of the test data points X_2 based on these known data points X_1.

$$p(X_2|X_1) = \frac{p(X_2, X_1)}{\int p(X_2, X_1) dX_2} = \frac{p(X_2, X_1)}{p(X_1)}$$

$$X_1|X_2 \sim \mathcal{N}(\mu_1 + K_{12}K_{22}^{-1}(X_2 - \mu_2), K_{11} - K_{12}K_{22}^{-1}K_{21})$$

Here, we provide a tricky proof.

Construct a new random vector $Z \in \mathbb{R}^{n \times 1}, Z = X_1 + AX_2$ and $A = -K_{12}K_{22}^{-1}, A \in \mathbb{R}^{n \times m}$. We can show the Z and X_2 are independent.

$$Cov[Z, X_2] = Cov[X_1 + AX_2, X_2] = Cov[X_1, X_2] + Cov[AX_2, X_2]$$

$$= Cov[X_1, X_2] + ACov[X_2, X_2] = K_{12} + (-K_{12}K_{22}^{-1})K_{22} = 0$$

Lemma 1: Expectation & Variance of Conditional Distribution

$$E[X_1|X_2 = x_2] = \int p(X_1|X_2 = x_2)X_1 dX_1$$

Lemma 2: Variance of Sum of Random Variables

$$Var[X_1+X_2] = E[((X_1+X_2)-(E[X_1]+E[X_2]))((X_1+X_2)-(E[X_1]+E[X_2]))^\top]$$

$$= E[((X_1 - E[X_1]) + (X_2 - E[X_2]))((X_1 - E[X_1]) + (X_2 - E[X_2]))^\top]$$

$$= E[(X_1-E[X_1])(X_1-E[X_1])^\top]+E[(X_2-E[X_2])(X_2-E[X_2])^\top]+E[(X_1-E[X_1])(X_2-E[X_2])^\top]+E[(X_2-E[X_2$$

$$= Var[X_1] + Var[X_2] + Cov[X_1, X_2] + Cov[X_2, X_1]$$

Lemma 3: Variance of Random Vector After Linear Transform

$$Var[AX + B] = E[(AX + B - E[AX + B])(AX + B - E[AX + B])^\top]$$

$$= E[(A(X - E[X]))(A(X - E[X]))^\top]$$

$$= E[(A(X - E[X]))((X - E[X])^\top A^\top)]$$

$$= AE[(X - E[X]))((X - E[X])^\top]A^\top$$

$$= AVar[X]A^\top$$

Lemma 4: Consider the covariance of random vector, when the linear transformation apply on the later one.

$$Cov[X_1, AX_2 + B] = E[(X_1 - E[X_1])(AX_2 + B - E[AX_2 + B])^\top]$$

$$= E[(X_1 - E[X_1])(X_2 - E[X_2])^\top A^\top]$$

$$= Cov[X_1, X_2]A^\top$$

Lemma 5: Consider the covariance of random vector, when the linear transformation apply on the former one.

$$Cov[AX_1 + B, X_2] = E[(AX_1 + B - E[AX_1 + B])(X_2 - E[X_2])^\top]$$

$$= E[A(X_1 - E[X_1])(X_2 - E[X_2])^\top]$$

$$= ACov[X_1, X_2]$$

Lemma 6: Consider the covariance of sum of 2 random vectors. Given a new random vector Z.

$$Cov[X_1 + W, X_2] = E[(X_1 + W - E[X_1 + W])(X_2 - E[X_2])]$$

$$= E[((X_1 - E[X_1]) + (W - E[W]))(X_2 - E[X_2])]$$

$$= E[(X_1 - E[X_1])(X_2 - E[X_2]) + (W - E[W])(X_2 - E[X_2])]$$

$$= Cov[X_1, X_2] + Cov[W, X_2]$$

With the above lemmas, we can start to derive the mean and the covariance of the conditional distribution.

The mean of the conditional distribution

$$E[X_1|X_2] = E[Z - AX_2|X_2] = E[Z|X_2] - E[AX_2|X_2]$$

$$= E[Z] - AE[X_2|X_2] = E[X_1 + AX_2] - AX_2 = \mu_1 + A\mu_2 - AX_2$$

$$= \mu_1 + A(\mu_2 - X_2) = \mu_1 + K_{12}K_{22}^{-1}(X_2 - \mu_2)$$

The variance of the conditional distribution

$$Var[X_1|X_2] = Var[Z] = Var[X_1 + AX_2]$$

$$= Var[X_1] + Var[AX_2] + Cov[X_1, AX_2] + Cov[AX_2, X_1]$$

$$= Var[X_1] + AVar[X_2]A^\top + Cov[X_1, X_2]A^\top + ACov[X_2, X_1]$$

$$= K_{11} + (-K_{12}K_{22}^{-1})K_{22}(-K_{12}K_{22}^{-1})^\top + K_{12}(-K_{12}K_{22}^{-1})^\top + (-K_{12}K_{22}^{-1})K_{21}$$

$$= K_{11} + (K_{12}K_{22}^{-1})K_{22}(K_{22}^{-1}K_{21}) - K_{12}(K_{22}^{-1}K_{21}) - (K_{12}K_{22}^{-1})K_{21}$$

$$= K_{11} + K_{12}K_{22}^{-1}K_{21} - 2K_{12}(K_{22}^{-1}K_{21}$$

$$= K_{11} - K_{12}K_{22}^{-1}K_{21}$$

Thus

$$X_1|X_2 \sim \mathcal{N}(\mu_1 + K_{12}K_{22}^{-1}(X_2 - \mu_2), K_{11} - K_{12}K_{22}^{-1}K_{21})$$

Reference

- Cross Validated - Deriving the conditional distributions of a multivariate normal distribution
- Stanford CS229 - Gaussian processes
- UT - Lecture 10 Conditional Expectation