COMS 4771 Machine Learning 2020
Problem Set #3

Joseph High - `jph2185`

April 13, 2020

Problem 1: Non-parametric Regression via Bayesian Modelling

Part (i)

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right) \text{ if and only if } x_i \sim \mathcal{N}\left(x_i; \mu_i, \Sigma_{ii}\right) \text{ for } i = 1, 2.\text{[1]}$$

Therefore, $x_1 \sim \mathcal{N}\left(x_1; \mu_1, \Sigma_{11}\right)$. A proof of this fact is provided below.

We can derive the marginal distribution of x_1 by integrating out x_2.

We are given that $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^d$. Then, *w.l.o.g.*, let $x_1 \in \mathbb{R}^m$ and $x_2 \in \mathbb{R}^n$ where $d = m + n$. That is, the dimensions of x_1 and x_2 should sum to the dimension of $\mathbf{x}$.

Letting $\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix}$ and $c = (2\pi)^{d/2} |\Sigma|^{1/2}$, we have the following:

$$
\begin{aligned}
f_{\mathbf{x}}(x_1) &= \frac{1}{c} \int \exp\left\{ -\frac{1}{2}\left(\begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix}^{\top} \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}^{-1} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix} \right) \right\} dx_2 \\
&= \frac{1}{c} \int \exp\left\{ -\frac{1}{2}\left(\begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix}^{\top} \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix} \right) \right\} dx_2 \\
&= \frac{1}{c} \int \exp\left\{ -\left[\frac{1}{2}(x_1 - \mu_1)^{\top}\Lambda_{11}(x_1 - \mu_1) + \frac{1}{2}(x_2 - \mu_2)^{\top}\Lambda_{21}(x_1 - \mu_1) + \right.\right. \\
&\qquad\qquad \left.\left. + \frac{1}{2}(x_1 - \mu_1)^{\top}\Lambda_{12}(x_2 - \mu_2) + \frac{1}{2}(x_2 - \mu_2)^{\top}\Lambda_{22}(x_2 - \mu_2) \right] \right\} dx_2
\end{aligned}
$$

---

[1] Larry Wasserman, *All of Statistics: A Concise Course in Statistical Inference* (New York: Springer, 2004), 40, Theorem 2.44.

$$= \frac{1}{c} \int \exp \left\{ - \left[\frac{1}{2}(x_1 - \mu_1)^\top \Lambda_{11}(x_1 - \mu_1) + (x_1 - \mu_1)^\top \Lambda_{12}(x_2 - \mu_2) + \right. \right.$$
$$\left. \left. + \frac{1}{2}(x_2 - \mu_2)^\top \Lambda_{22}(x_2 - \mu_2) \right] \right\} dx_2$$

completing the square...

$$= \frac{1}{c} \int \exp \left\{ - \left[\frac{1}{2} \left((x_2 - \mu_2) + \Lambda_{22}^{-1}\Lambda_{21}(x_1 - \mu_1) \right)^\top \Lambda_{22} \left((x_2 - \mu_2) + \Lambda_{22}^{-1}\Lambda_{21}(x_1 - \mu_1) \right) + \right. \right.$$
$$\left. \left. + \frac{1}{2}(x_1 - \mu_1)^\top \Lambda_{11}(x_1 - \mu_1) - \frac{1}{2}(x_1 - \mu_1)^\top \Lambda_{12}\Lambda_{22}^{-1}\Lambda_{21}(x_1 - \mu_1) \right] \right\} dx_2$$

$$= \exp \left\{ -\frac{1}{2}(x_1 - \mu_1)^\top \Lambda_{11}(x_1 - \mu_1) - \frac{1}{2}(x_1 - \mu_1)^\top \Lambda_{12}\Lambda_{22}^{-1}\Lambda_{21}(x_1 - \mu_1) \right\} \cdot$$

$$\cdot \underbrace{\frac{1}{c} \int \exp \left\{ - \left(\frac{1}{2} \left[(x_2 - \mu_2) + \Lambda_{22}^{-1}\Lambda_{21}(x_1 - \mu_1) \right]^\top \Lambda_{22} \left[(x_2 - \mu_2) + \Lambda_{22}^{-1}\Lambda_{21}(x_1 - \mu_1) \right] \right) \right\} dx_2}_{= (2\pi)^{n/2}|\Lambda_{22}^{-1}|^{1/2}}$$

$$= \exp \left\{ -\frac{1}{2}(x_1 - \mu_1)^\top \left(\Lambda_{11} - \Lambda_{12}\Lambda_{22}^{-1}\Lambda_{21} \right)(x_1 - \mu_1) \right\} \cdot \frac{1}{c}(2\pi)^{n/2}|\Lambda_{22}^{-1}|^{1/2}$$

$$= \frac{(2\pi)^{n/2}|\Lambda_{22}^{-1}|^{1/2}}{(2\pi)^{d/2}|\Sigma|^{1/2}} \cdot \exp \left\{ -\frac{1}{2}(x_1 - \mu_1)^\top \Sigma_{11}^{-1}(x_1 - \mu_1) \right\}$$

$$= \frac{(2\pi)^{n/2}|\Sigma_{22} - \Sigma_{12}^\top \Sigma_{11}^{-1}\Sigma_{12}|^{1/2}}{(2\pi)^{d/2} \left(|\Sigma_{11}||\Sigma_{22} - \Sigma_{12}^\top \Sigma_{11}^{-1}\Sigma_{12}| \right)^{1/2}} \cdot \exp \left\{ -\frac{1}{2}(x_1 - \mu_1)^\top \Sigma_{11}^{-1}(x_1 - \mu_1) \right\} \qquad \textit{(from part (ii))}$$

$$= \frac{1}{(2\pi)^{m/2}|\Sigma_{11}|^{1/2}} \cdot \exp \left\{ -\frac{1}{2}(x_1 - \mu_1)^\top \Sigma_{11}^{-1}(x_1 - \mu_1) \right\}$$

$$= \mathcal{N}(x_1; \mu_1, \Sigma_{11})$$

Hence, $\;x_1 \sim \mathcal{N}(x_1; \mu_1, \Sigma_{11})$

Part (ii)

Proof. The joint distribution on $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma) = \mathcal{N}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right)$, is

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})\right\} \tag{1}$$

To express (1) in the desired form, the following facts which were provided in the homework handout will be used:

a.) $\Sigma^{-1} = \begin{bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{bmatrix}$

b.) $\Sigma^{11} = \left(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^\top\right)^{-1} = \Sigma_{11}^{-1} + \Sigma_{11}^{-1}\Sigma_{12}\left(\Sigma_{22} - \Sigma_{12}^\top\Sigma_{11}^{-1}\Sigma_{12}\right)^{-1}\Sigma_{12}^\top\Sigma_{11}^{-1}$

c.) $\Sigma^{22} = \left(\Sigma_{22} - \Sigma_{12}^\top\Sigma_{11}^{-1}\Sigma_{12}\right)^{-1} = \Sigma_{22}^{-1} + \Sigma_{22}^{-1}\Sigma_{12}^\top\left(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^\top\right)^{-1}\Sigma_{12}\Sigma_{22}^{-1}$

d.) $\Sigma^{12} = -\Sigma_{11}^{-1}\Sigma_{12}\left(\Sigma_{22} - \Sigma_{12}^\top\Sigma_{11}^{-1}\Sigma_{12}\right)^{-1} = \left(\Sigma_{21}\right)^\top$

e.) $A = \Sigma_{22} - \Sigma_{12}^\top\Sigma_{11}^{-1}\Sigma_{12} = (\Sigma^{22})^{-1}$

f.) $b = \mu_2 + \Sigma_{12}^\top\Sigma_{11}^{-1}(x_1 - \mu_1)$

From (1), we have

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}\begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix}^\top \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}^{-1} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix}\right\}$$

$$= \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}\begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix}^\top \begin{bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{bmatrix} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix}\right\}$$

$$= \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}\left[(x_1 - \mu_1)^\top\Sigma^{11}(x_1 - \mu_1) + (x_2 - \mu_2)^\top\Sigma^{21}(x_1 - \mu_1)\right.\right.$$

$$\left.\left. + (x_1 - \mu_1)^\top\Sigma^{12}(x_2 - \mu_2) + (x_2 - \mu_2)^\top\Sigma^{22}(x_2 - \mu_2)\right]\right\}$$

$$= \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}\left[(x_1 - \mu_1)^\top\Sigma^{11}(x_1 - \mu_1) + 2(x_1 - \mu_1)^\top\Sigma^{12}(x_2 - \mu_2)\right.\right.$$

$$\left.\left. + (x_2 - \mu_2)^\top\Sigma^{22}(x_2 - \mu_2)\right]\right\}$$

Focusing only on the exponent and substituting in expressions (b), (c), (d), and (e) we get

$$-\frac{1}{2}(x_1 - \mu_1)^\top \Sigma^{11}(x_1 - \mu_1) \; - \; \frac{1}{2}(x_2 - \mu_2)^\top \Sigma^{21}(x_1 - \mu_1) - \frac{1}{2}(x_1 - \mu_1)^\top \Sigma^{12}(x_2 - \mu_2)$$

$$-\frac{1}{2}(x_2 - \mu_2)^\top \Sigma^{22}(x_2 - \mu_2)$$

$$= -\frac{1}{2}(x_1 - \mu_1)^\top \underbrace{\left(\Sigma_{11}^{-1} + \Sigma_{11}^{-1}\Sigma_{12}\left(\Sigma_{22} - \Sigma_{12}^\top\Sigma_{11}^{-1}\Sigma_{12}\right)^{-1}\Sigma_{12}^\top\Sigma_{11}^{-1}\right)}_{= \; \Sigma^{11}}(x_1 - \mu_1)$$

$$-\frac{1}{2}(x_2 - \mu_2)^\top \underbrace{\left(-\Sigma_{11}^{-1}\Sigma_{12}\left(\Sigma_{22} - \Sigma_{12}^\top\Sigma_{11}^{-1}\Sigma_{12}\right)^{-1}\right)^\top}_{= \; \Sigma^{21}}(x_1 - \mu_1)$$

$$-\frac{1}{2}(x_1 - \mu_1)^\top \underbrace{\left(-\Sigma_{11}^{-1}\Sigma_{12}\left(\Sigma_{22} - \Sigma_{12}^\top\Sigma_{11}^{-1}\Sigma_{12}\right)^{-1}\right)}_{= \; \Sigma^{12}}(x_2 - \mu_2) - \frac{1}{2}(x_2 - \mu_2)^\top \underbrace{A^{-1}}_{= \; \Sigma^{22}}(x_2 - \mu_2)$$

(Substituting in $A = \Sigma_{22} - \Sigma_{12}^\top\Sigma_{11}^{-1}\Sigma_{12}$)

$$= -\frac{1}{2}(x_1 - \mu_1)^\top \left(\Sigma_{11}^{-1} + \Sigma_{11}^{-1}\Sigma_{12}A^{-1}\Sigma_{12}^\top\Sigma_{11}^{-1}\right)(x_1 - \mu_1) \; + \; \frac{1}{2}(x_2 - \mu_2)^\top \left(\Sigma_{11}^{-1}\Sigma_{12}A^{-1}\right)^\top (x_1 - \mu_1)$$

$$+ \frac{1}{2}(x_1 - \mu_1)^\top \left(\Sigma_{11}^{-1}\Sigma_{12}A^{-1}\right)(x_2 - \mu_2) - \frac{1}{2}(x_2 - \mu_2)^\top A^{-1}(x_2 - \mu_2)$$

$$= -\frac{1}{2}(x_1 - \mu_1)^\top \Sigma_{11}^{-1}(x_1 - \mu_1) \; - \; \frac{1}{2}(x_1 - \mu_1)^\top \Sigma_{11}^{-1}\Sigma_{12}A^{-1}\Sigma_{12}^\top\Sigma_{11}^{-1}(x_1 - \mu_1)$$

$$+ (x_1 - \mu_1)^\top \left(\Sigma_{11}^{-1}\Sigma_{12}A^{-1}\right)(x_2 - \mu_2) \; - \; \frac{1}{2}(x_2 - \mu_2)^\top A^{-1}(x_2 - \mu_2)$$

Factor the last three summands using the fact that $(v - w)^\top S(v - w) = v^\top Sv - 2w^\top Sv + w^\top Sw$ for any vectors v and w and any symmetric matrix S.

$$= -\frac{1}{2}(x_1 - \mu_1)^\top \Sigma_{11}^{-1}(x_1 - \mu_1)$$

$$-\frac{1}{2}\left[(x_2 - \mu_2) - \Sigma_{12}^\top\Sigma_{11}^{-1}(x_1 - \mu_1)\right]^\top A^{-1}\left[(x_2 - \mu_2) - \Sigma_{12}^\top\Sigma_{11}^{-1}(x_1 - \mu_1)\right]$$

$$= -\frac{1}{2}(x_1 - \mu_1)^\top \Sigma_{11}^{-1}(x_1 - \mu_1)$$

$$-\frac{1}{2}\left[x_2 - \left(\mu_2 + \Sigma_{12}^\top\Sigma_{11}^{-1}(x_1 - \mu_1)\right)\right]^\top A^{-1}\left[x_2 - \left(\mu_2 + \Sigma_{12}^\top\Sigma_{11}^{-1}(x_1 - \mu_1)\right)\right]$$

$$= -\frac{1}{2}(x_1 - \mu_1)^\top \Sigma_{11}^{-1}(x_1 - \mu_1) + \frac{1}{2}\left[(x_2 - b)^\top A^{-1}(x_2 - b)\right]$$

Therefore, the joint distribution of $\mathbf{x}$ can be written as

$$\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}\left((x_1 - \mu_1)^\top \Sigma_{11}^{-1}(x_1 - \mu_1) + (x_2 - b)^\top A^{-1}(x_2 - b)\right)\right\}$$

$\square$

Part (iii)

Proof. We are given that $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^d$. Then *w.l.o.g.*, suppose $x_1 \in \mathbb{R}^m$ and $x_2 \in \mathbb{R}^n$ where $d = m + n$. That is, the dimensions of x_1 and x_2 should sum to the dimension of $\mathbf{x}$.

To show that the joint on $\mathbf{x}$ can be decomposed as the product $\mathcal{N}(x_1; \mu_1, \Sigma_{11})\mathcal{N}(x_2; b, A)$, we can continue from the result in part (ii):

$f(x_1, x_2) =$

$$\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}\left((x_1 - \mu_1)^\top \Sigma_{11}^{-1}(x_1 - \mu_1) + (x_2 - b)^\top A^{-1}(x_2 - b)\right)\right\}$$

$$= \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x_1 - \mu_1)^\top \Sigma_{11}^{-1}(x_1 - \mu_1) - \frac{1}{2}(x_2 - b)^\top A^{-1}(x_2 - b)\right\}$$

$$= \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x_1 - \mu_1)^\top \Sigma_{11}^{-1}(x_1 - \mu_1)\right\} \exp\left\{-\frac{1}{2}(x_2 - b)^\top A^{-1}(x_2 - b)\right\}$$

Using $|\Sigma| = |\Sigma_{11}||\Sigma_{22} - \Sigma_{12}^\top \Sigma_{11}^{-1}\Sigma_{12}|$ *and* $d = m + n$ *we have:*

$$= \frac{1}{(2\pi)^{\frac{m+n}{2}}\left(|\Sigma_{11}||\Sigma_{22} - \Sigma_{12}^\top \Sigma_{11}^{-1}\Sigma_{12}|\right)^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(x_1 - \mu_1)^\top \Sigma_{11}^{-1}(x_1 - \mu_1)\right\}$$

$$\cdot \exp\left\{-\frac{1}{2}(x_2 - b)^\top A^{-1}(x_2 - b)\right\}$$

From the supposition in part (ii), $A = \Sigma_{22} - \Sigma_{12}^\top \Sigma_{11}^{-1}\Sigma_{12}$. *Plugging in, we get:*

$$= \frac{1}{(2\pi)^{\frac{m}{2}}(2\pi)^{\frac{n}{2}}|\Sigma_{11}|^{\frac{1}{2}}|A|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(x_1 - \mu_1)^\top \Sigma_{11}^{-1}(x_1 - \mu_1)\right\} \exp\left\{-\frac{1}{2}(x_2 - b)^\top A^{-1}(x_2 - b)\right\}$$

$$= \frac{1}{(2\pi)^{m/2}|\Sigma_{11}|^{1/2}} \exp\left\{-\frac{1}{2}(x_1 - \mu_1)^\top \Sigma_{11}^{-1}(x_1 - \mu_1)\right\} \cdot \underbrace{\frac{1}{(2\pi)^{n/2}|A|^{1/2}} \exp\left\{-\frac{1}{2}(x_2 - b)^\top A^{-1}(x_2 - b)\right\}}_{= \mathcal{N}(x_2; b, A)}$$

$$\underbrace{}_{= \mathcal{N}(x_1; \mu_1, \Sigma_{11})}$$

$$= \mathcal{N}(x_1; \mu_1, \Sigma_{11})\mathcal{N}(x_2; b, A)$$

$\square$

Part (iv)

The conditional distribution of x_2 given x_1 can be computed from the joint distribution on $\mathbf{x} = [x_1 \ \ x_2]^\top$ and the marginal distribution of x_1:

$$f(x_2|x_1) = \frac{f(x_1, x_2)}{f(x_1)}$$

Thus, we can use the results from parts (i) and (iii) to compute the density of $x_2|x_1$.

In part (i) it was shown that the marginal density on x_1 is such that $f(x_1) = \mathcal{N}(x_1; \mu_1, \Sigma_{11})$.

In part (iii) it was shown that the joint density on $\mathbf{x}$ can be decomposed into the product $\mathcal{N}(x_1; \mu_1, \Sigma_{11})\mathcal{N}(x_2; b, A)$.

The conditional density of x_2 given x_1 is then

$$f(x_2|x_1) = \frac{f(x_1, x_2)}{f(x_1)} = \frac{\mathcal{N}(x_1; \mu_1, \Sigma_{11})\mathcal{N}(x_2; b, A)}{\mathcal{N}(x_1; \mu_1, \Sigma_{11})} = \mathcal{N}(x_2; b, A)$$

$$\implies \quad x_2|x_1 \sim \mathcal{N}(x_2; b, A)$$

or, equivalently (since $b = \mu_2 + \Sigma_{12}^\top \Sigma_{11}^{-1}(x_1 - \mu_1)$ and $A = \Sigma_{22} - \Sigma_{12}^\top \Sigma_{11}^{-1}\Sigma_{12}$),

$$x_2|x_1 \sim \mathcal{N}\left(x_2; \mu_2 + \Sigma_{12}^\top \Sigma_{11}^{-1}(x_1 - \mu_1), \Sigma_{22} - \Sigma_{12}^\top \Sigma_{11}^{-1}\Sigma_{12}\right)$$

Part (v)

Four random functions were generated in Python using the Gaussian distribution with $\mu_n = \vec{0}$ and $\Sigma_{n \times n} = I$ and subsequently plotted (See Figure 1). One will note that these random functions appear similar to that of a white noise process. This makes sense since the random numbers were randomly sampled from a Gaussian distribution such that the correlation between all x_i and x_j, for $i, j \in \{1, ..., 500\}$ and $i \neq j$, is zero. It's not surprising, then, that the resulting functions do not appear to be smooth at all.

Figure 1: Random Functions with $\Sigma_{n \times n} = I$ and $\mu_n = \vec{0}$
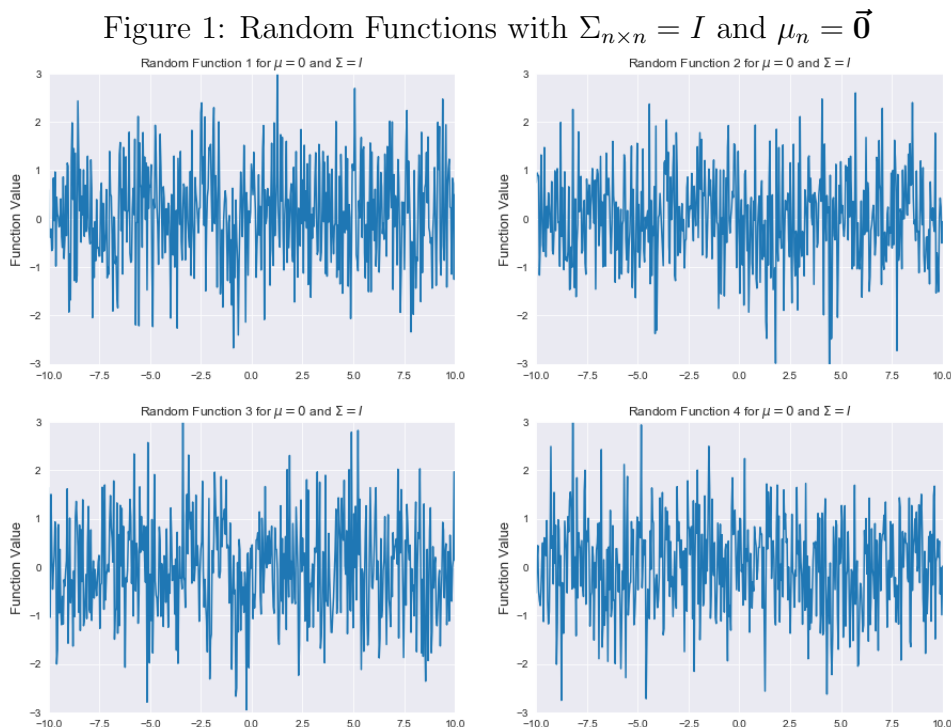
Figure 2 (below) displays random functions generated with $\Sigma_{n \times n} = \mathbb{J}_n$ (the all ones matrix) and $\mu_n = \vec{0}$. Notice that each function is a straight line at different values on the $y-$axis, indicating that the function values are perfectly correlated with one another.

Figure 2: Random Functions with $\Sigma_{n \times n} = \mathbb{J}_n$ and $\mu_n = \vec{0}$
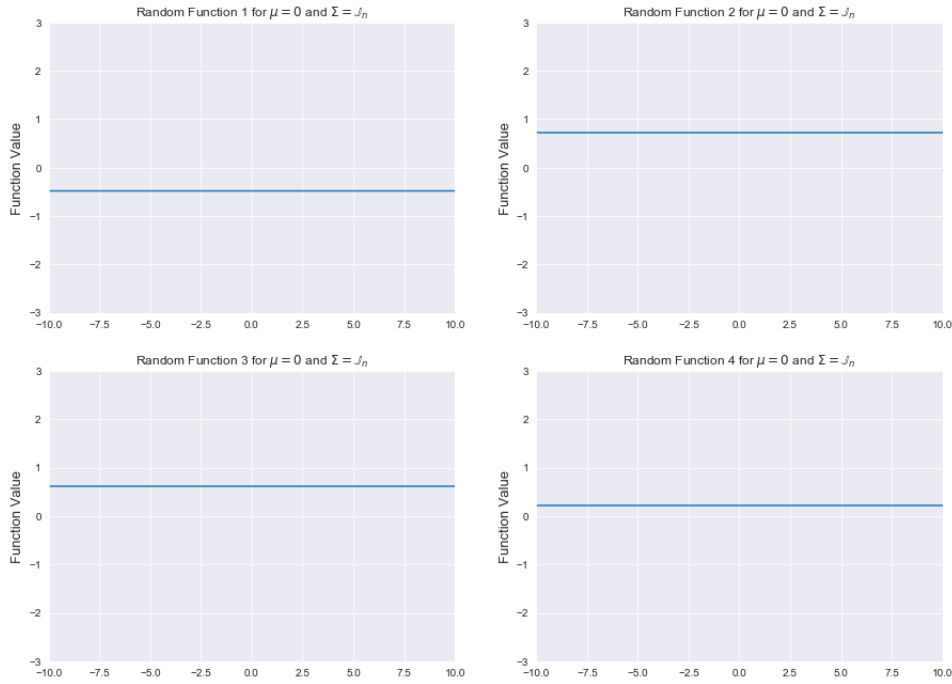
Figure 3 displays random functions generated with $[\Sigma_{i,j}]_{i \neq j} = 0.1$, $[\Sigma_{i,j}]_{i=j} = 0.2$ and $\mu_n = \vec{0}$. Here, we see that there is much less noise than that of Figure 1 where the functions were generated with zero correlation. While there is some positive correlation between each pair of values, the functions still do not appear to be smooth.
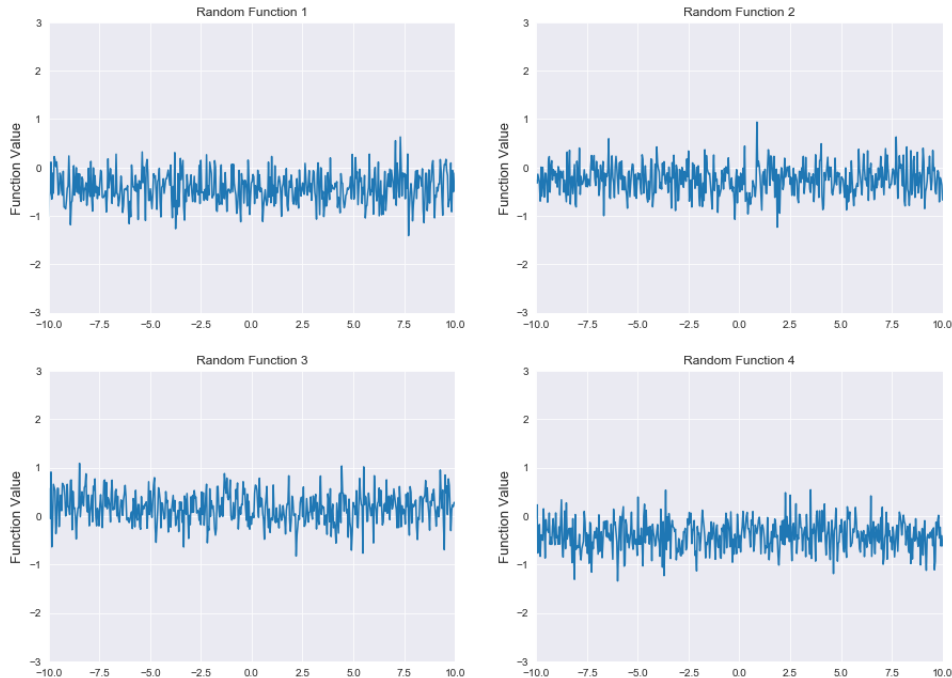
Figure 3: Random Functions with $[\Sigma_{i,j}]_{i \neq j} = 0.1$, $[\Sigma_{i,j}]_{i=j} = 0.2$ and $\mu_n = \vec{0}$

Figure 4 displays the functions whereby Σ was gradually increased. It comes as no surprise that the resulting behavior is an increase in noise as Σ is increased.

Figure 4: Varying Σ and $\mu_n = \vec{0}$
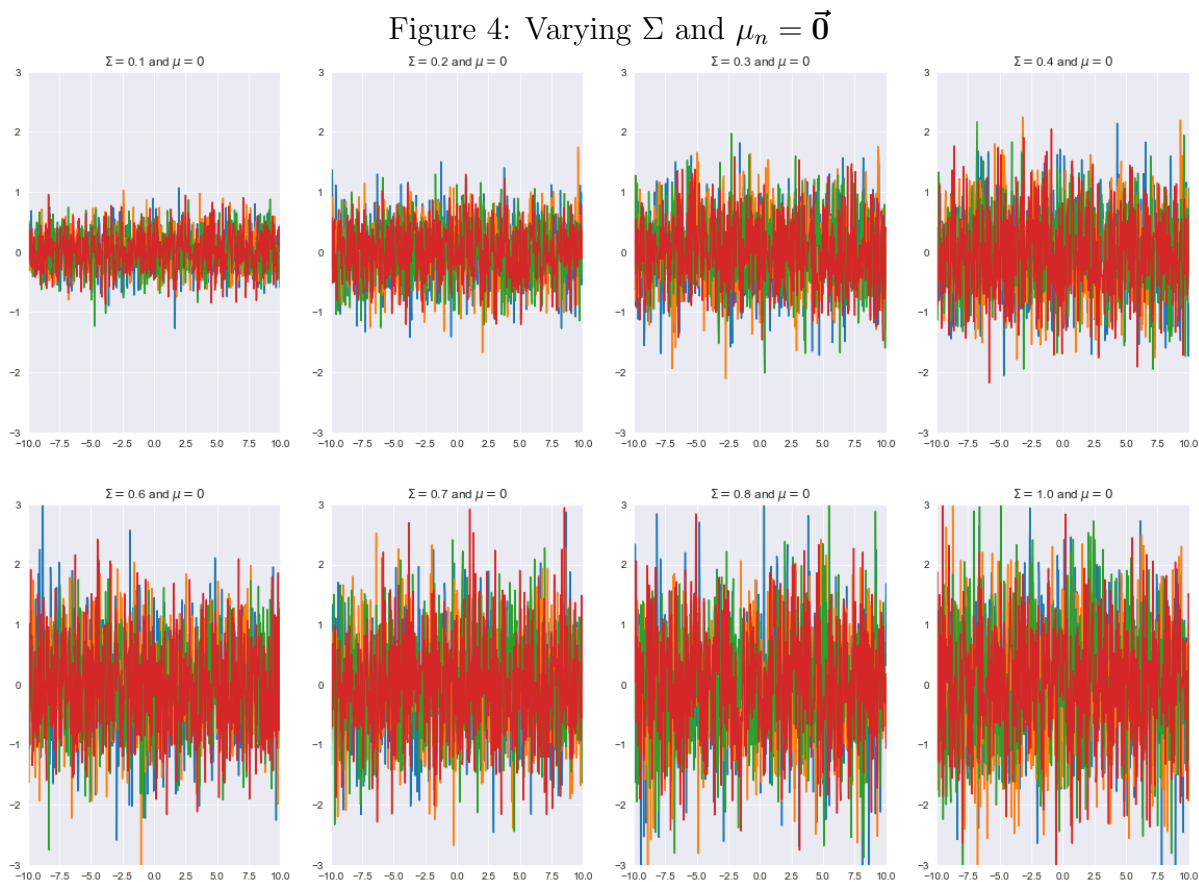
Figure 5 displays the functions whereby μ was gradually increased. Naturally, the functions are centered at wherever μ is set.
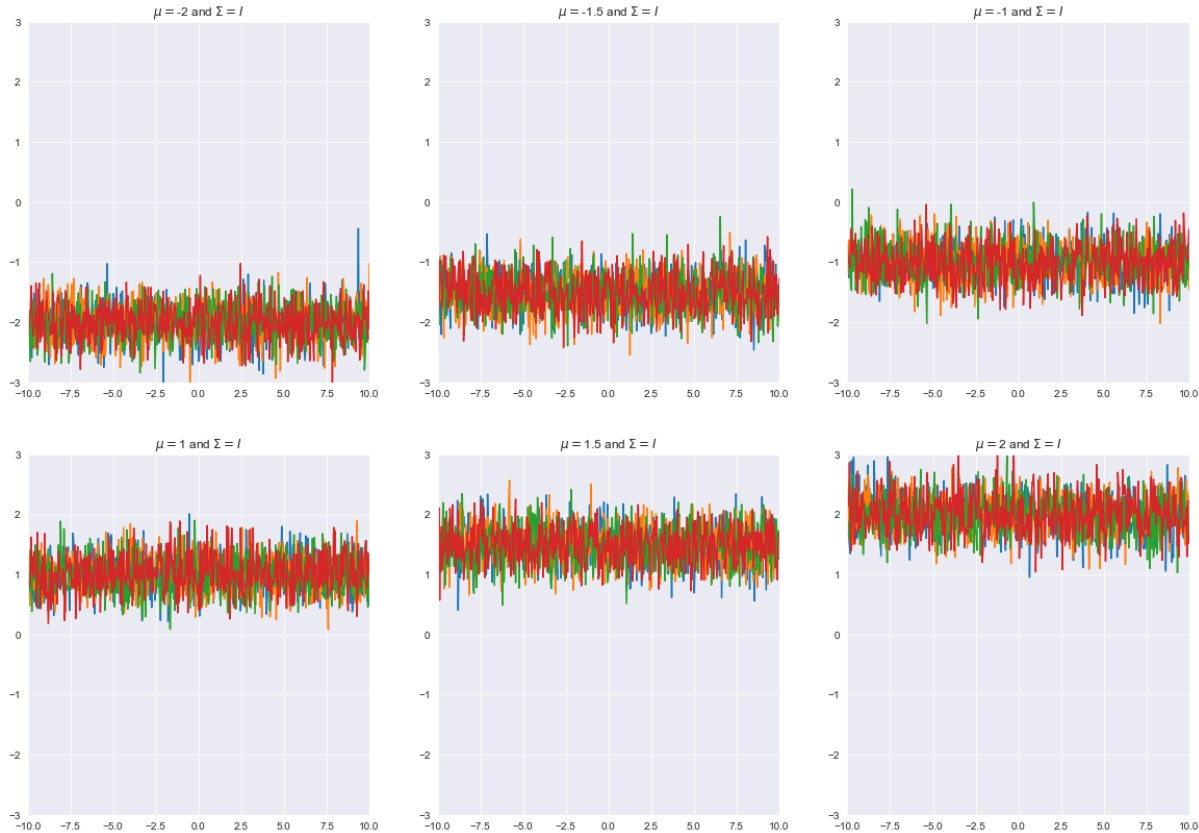
Figure 5: Varying μ and setting $[\Sigma_{i,j}]_{i=j} = 0.1$, $[\Sigma_{i,j}]_{i \neq j} = 0$

Figure 6 plots the functions whereby Σ is varied. Namely, the diagonal entries, i.e. $[\Sigma_{i,j}]_{i=j}$, were varied from 0.1 to 1, and the off-diagonal entries, i.e. $[\Sigma_{i,j}]_{i \neq j}$, were varied from 0.05 to 2. It can be seen that the function values start off relatively correlated and quickly become wildly uncorrelated. These results make sense in that while the correlation between points are increasing, the variance is also increasing rapidly.
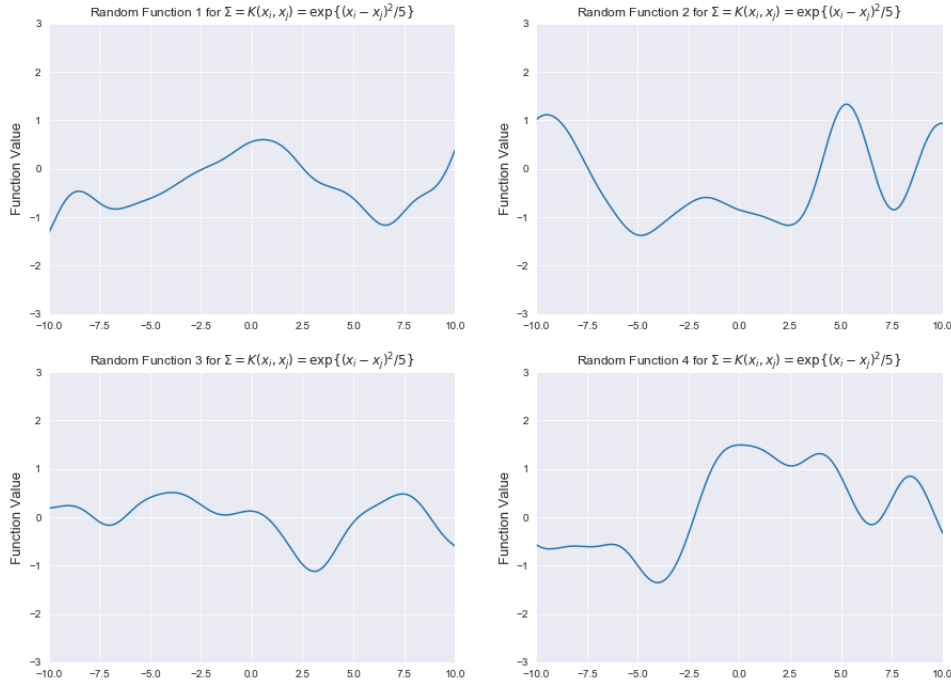
Figure 6: Varying μ and $[\Sigma_{i,j}]_{i=j} = 0.1$, $[\Sigma_{i,j}]_{i \neq j} = 0$
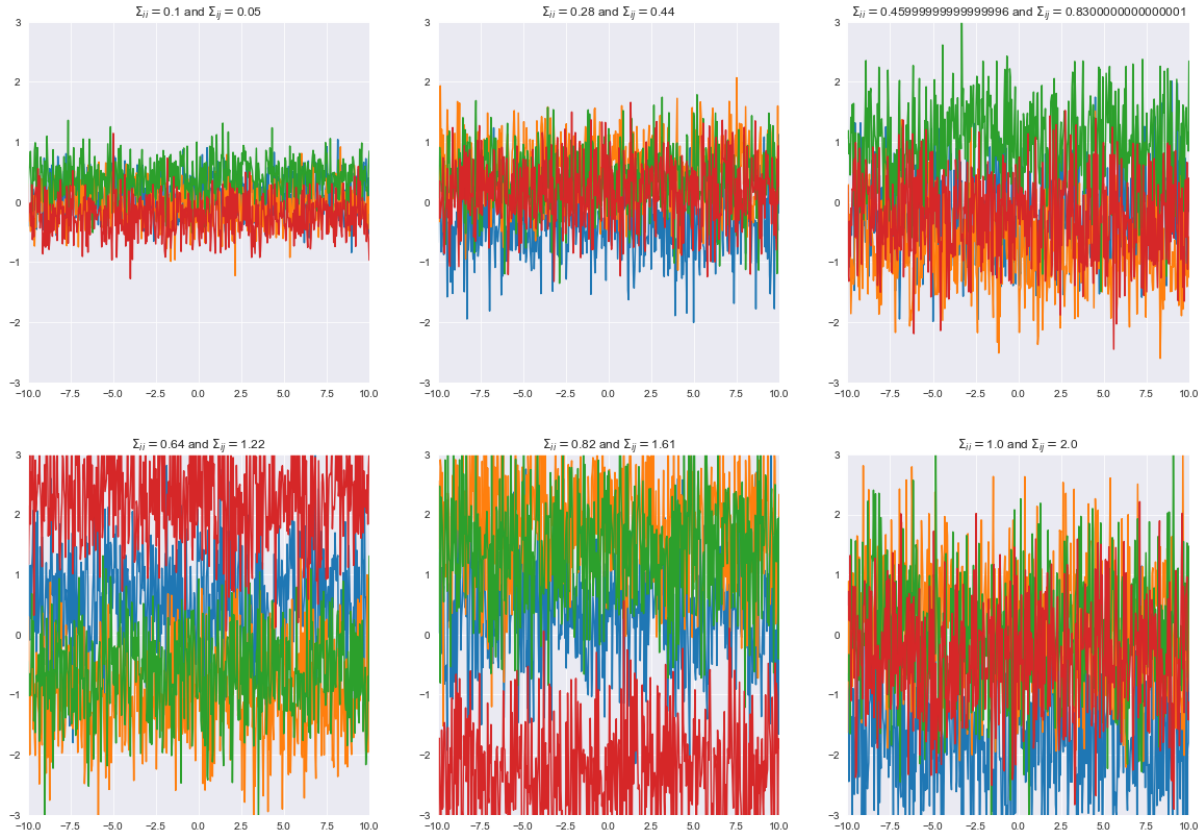
Part (vi)

Figure 7 displays the random functions generated using $\Sigma_{n \times n} = K$ and $\mu_n = \vec{0}$. In this case, each of the random functions appear to be significantly smoother than the functions plotted in part (v).

Figure 7: Random Functions with $\Sigma_{n \times n} = K$ and $\mu_n = \vec{0}$

Part (vii)

The kernel used to generate periodic functions with a periodicity of 3 units was inspired by the periodic covariance covered in *Gaussian Processes for Machine Learning*.[2] Namely, the squared-exponential kernel was implemented:

$$k(x_i, x_j) = \exp\left(-\frac{2\sin^2\left(\frac{x_i - x_j}{2}\right)}{\lambda^2}\right)$$

where λ is a hyperparameter which defines the periodicity. To achieve a periodicity of 3 units, set $\lambda = 3$.

Figure 8 displays each random function generated with $\Sigma = k(x_i, x_j)$, as defined above, and $\boldsymbol{\mu} = \vec{0}$. One will note that each function clearly exhibits periodic behavior.

---

[2]Carl Rasmussen and Christopher Williams, *Gaussian Processes for Machine Learning* (The MIT Press, 2006), 92

Figure 8: Periodic Random Functions
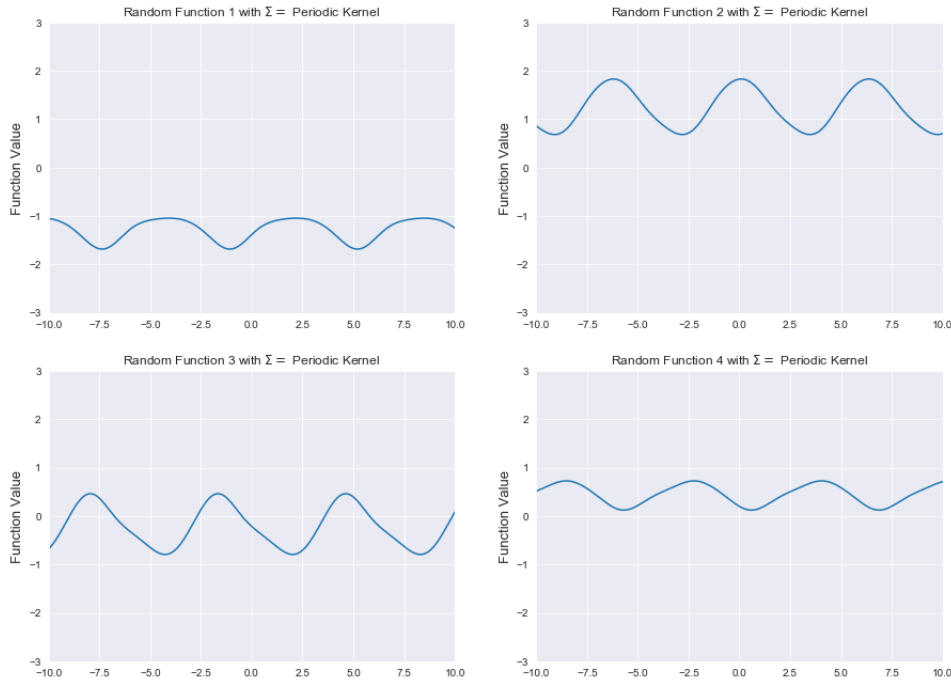
Part (viii)

Using the result from part(iv), the posterior distribution of $\mathbf{Y}|\overline{\mathbf{Y}}$ is

$$\mathcal{N}\left(\mathbf{Y};\ \mu_n + K(\mathbf{X}, \overline{\mathbf{X}})K(\overline{\mathbf{X}}, \overline{\mathbf{X}})^{-1}\left(\overline{\mathbf{Y}} - \mu_m\right),\ K(\mathbf{X}, \mathbf{X}) - K(\mathbf{X}, \overline{\mathbf{X}})K(\overline{\mathbf{X}}, \overline{\mathbf{X}})^{-1}K(\overline{\mathbf{X}}, \mathbf{X})\right)$$

Part (ix)

Figure 9 displays the plots for all four functions generated using the kernel $\exp\left\{\dfrac{(x_i - x_j)^2}{h}\right\}$ with $h = 5$.

As in parts (vi) and (vii) where a kernel was also utilized, the random functions appear to be smooth. The three training data points are plotted in blue. Notice that while all four random functions take on different values for most values of x, each passes through all three training data points.

Figure 9: Random Functions with $\Sigma_{n \times n} = K$ (with training data)
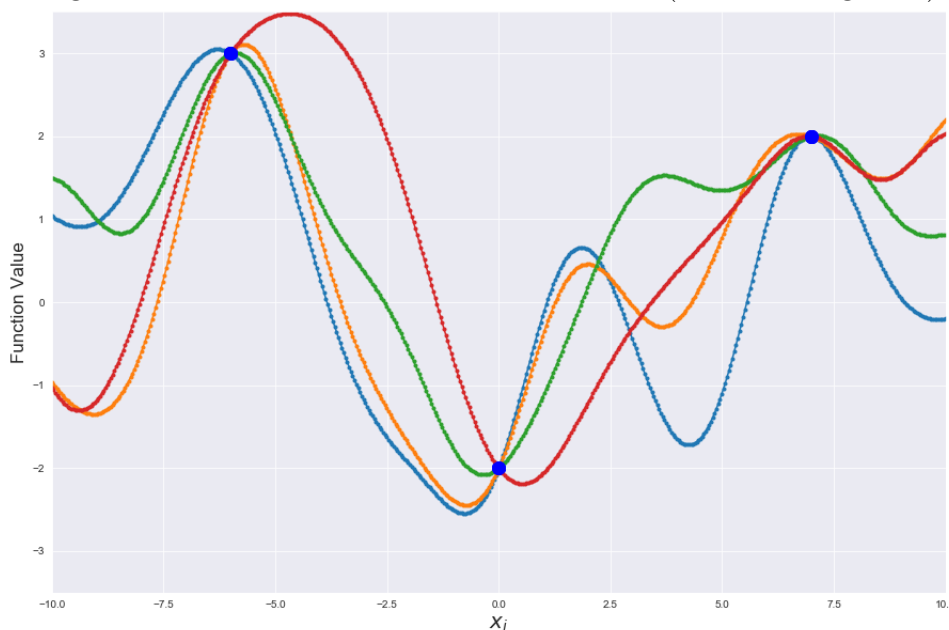
Part (x)

As discussed in part(vii) the kernel used to generate periodic functions of periodicity 3 units is:

$$k(x_i, x_j) = \exp\left(-\frac{2\sin^2\left(\frac{x_i - x_j}{2}\right)}{\lambda^2}\right)$$

where λ is a hyperparameter which defines the periodicity.[3] To achieve a periodicity of 3 units, λ was set to 3.

Figure 11 displays the plots for all four functions generated using the squared-exponential periodic kernel. The three training data points are plotted in blue. Each of the random functions are almost indistinguishable; they're practically superimposed onto one another when plotted in the same graph. Notice that while all four random functions take on different values for most values of x, all four random functions pass through all three training data points.
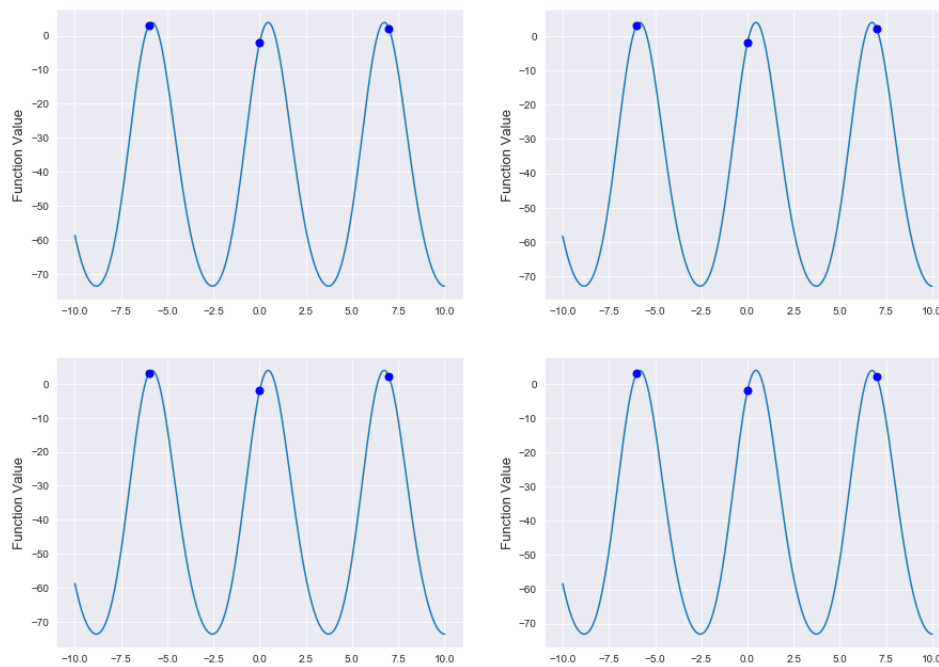
---

[3]Carl Rasmussen and Christopher Williams, *Gaussian Processes for Machine Learning* (The MIT Press, 2006), 92

Figure 10: Periodic Random Functions plotted together
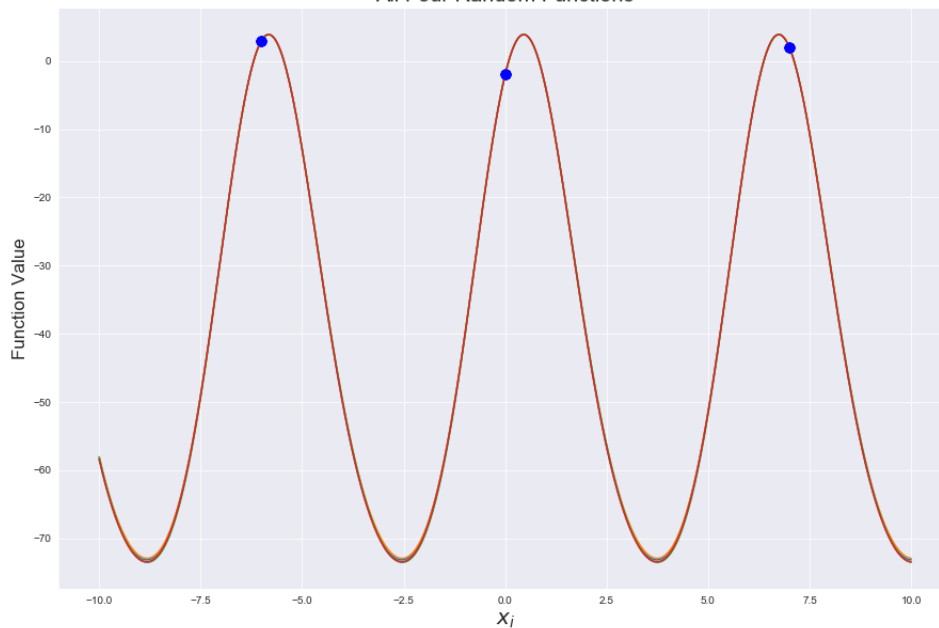
Each of the four plots were also plotted individually to confirm that all four are indeed very similar (see Figure 11).

Figure 11: Periodic Random Functions plotted individually

Part (xi)

The mean of the posterior $\mathbf{Y}|\overline{\mathbf{Y}}$ is

$$
\begin{aligned}
\mu_{\mathbf{Y}|\overline{\mathbf{Y}}} &= \mu_n + K(\mathbf{X}, \overline{\mathbf{X}}) K(\overline{\mathbf{X}}, \overline{\mathbf{X}})^{-1} \left(\overline{\mathbf{Y}} - \mu_m \right) \\
&= K(\mathbf{X}, \overline{\mathbf{X}}) K(\overline{\mathbf{X}}, \overline{\mathbf{X}})^{-1} \overline{\mathbf{Y}} \qquad \left(\text{since } \mu_n = \vec{0} \text{ and } \mu_m = \vec{0} \right)
\end{aligned}
$$

Part (xii)

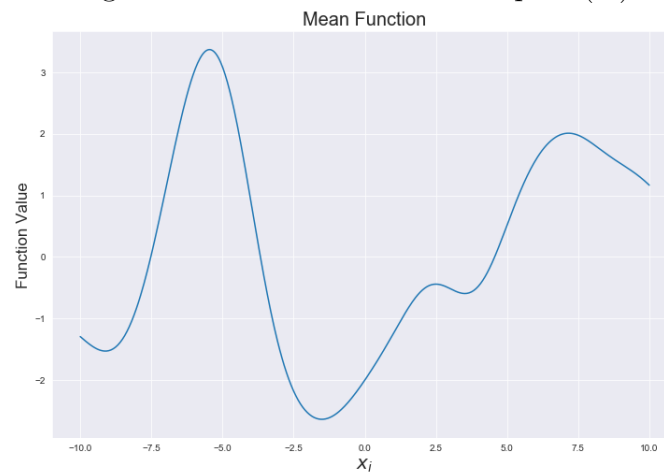Figure 12 and Figure 13 are plots of the mean functions from parts (ix) and (x), respectively.

Figure 12: Mean Function from part (ix)

Figure 13: Mean Function from part (x)