Exercise of Supervised Learning: Gaussian Processes Part 2

Yawei Li

yawei.li@stat.uni-muenchen.de

February 4, 2025

Exercise 1: Gaussian Posterior Process

Assume your data follows the following law:

$$\mathbf{y} = \mathbf{f} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{I}),$$

with $\mathbf{f} = f(\mathbf{x}) \in \mathbb{R}^n$ being a realization of a Gaussian process (GP), for which we a priori assume

$$f(\mathbf{x}) \sim \mathcal{GP}\left(m(\mathbf{x}), k\left(\mathbf{x}, \mathbf{x}'\right)\right).$$

$\mathbf{x}$ here only consists of 1 feature that is observed for n data points.

(a) Derive / define the prior distribution of $\mathbf{f}$.

1 (a): Prior Distribution of f

- $\mathbf{f} \sim \mathcal{N}(\mathbf{m}, \mathbf{K})$.
- $\mathbf{m} = m(\mathbf{x})$.
- $\boldsymbol{K}_{ij} = k(x^{(i)}, x^{(j)})$.
- NB: Note the (in-)finite Gaussian property of a GP: no matter which finite collection of points you choose from the domain of the process, the corresponding values of the processes are jointly Gaussian.

1 (a): Prior Distribution of f

- $\mathbf{f} \sim \mathcal{N}(\mathbf{m}, \mathbf{K})$.
- $\mathbf{m} = m(\mathbf{x})$.
- $\boldsymbol{K}_{ij} = k(x^{(i)}, x^{(j)})$.
- NB: Note the (in-)finite Gaussian property of a GP: no matter which finite collection of points you choose from the domain of the process, the corresponding values of the processes are jointly Gaussian.

1 (a): Prior Distribution of f

- $\mathbf{f} \sim \mathcal{N}(\mathbf{m}, \mathbf{K})$.
- $\mathbf{m} = m(\mathbf{x})$.
- $\boldsymbol{K}_{ij} = k(x^{(i)}, x^{(j)})$.
- NB: Note the (in-)finite Gaussian property of a GP: no matter which finite collection of points you choose from the domain of the process, the corresponding values of the processes are jointly Gaussian.

1 (a): Prior Distribution of f

- $\mathbf{f} \sim \mathcal{N}(\mathbf{m}, \mathbf{K})$.
- $\mathbf{m} = m(\mathbf{x})$.
- $\boldsymbol{K}_{ij} = k(x^{(i)}, x^{(j)})$.
- NB: Note the (in-)finite Gaussian property of a GP: no matter which finite collection of points you choose from the domain of the process, the corresponding values of the processes are jointly Gaussian.

Exercise 1 (b)

(b) Derive the posterior distribution of $\mathbf{f}|\mathbf{y}$.

1 (b): Likelihood and Prior

The posterior can be derived from the likelihood and prior using Bayes rule:

$$\mathbf{y} = \mathbf{f} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{I})$$

The likelihood is:

$$p(\mathbf{y}|\mathbf{f}) \propto \exp\left(-\frac{1}{2}(\mathbf{y}-\mathbf{f})^\top (\sigma^2 \boldsymbol{I})^{-1}(\mathbf{y}-\mathbf{f})\right).$$

The prior is:

$$p(\mathbf{f}) \propto \exp\left(-\frac{1}{2}(\mathbf{f}-\mathbf{m})^\top \mathbf{K}^{-1}(\mathbf{f}-\mathbf{m})\right).$$

1 (b): Likelihood and Prior

The posterior can be derived from the likelihood and prior using Bayes rule:

$$\mathbf{y} = \mathbf{f} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{I})$$

The likelihood is:

$$p(\mathbf{y}|\mathbf{f}) \propto \exp\left(-\frac{1}{2}(\mathbf{y}-\mathbf{f})^\top (\sigma^2 \boldsymbol{I})^{-1} (\mathbf{y}-\mathbf{f})\right).$$

The prior is:

$$p(\mathbf{f}) \propto \exp\left(-\frac{1}{2}(\mathbf{f}-\mathbf{m})^\top \mathbf{K}^{-1} (\mathbf{f}-\mathbf{m})\right).$$

1 (b): Likelihood and Prior

The posterior can be derived from the likelihood and prior using Bayes rule:

$$\mathbf{y} = \mathbf{f} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{I})$$

The likelihood is:

$$p(\mathbf{y}|\mathbf{f}) \propto \exp\left(-\frac{1}{2}(\mathbf{y}-\mathbf{f})^\top (\sigma^2 \boldsymbol{I})^{-1}(\mathbf{y}-\mathbf{f})\right).$$

The prior is:

$$p(\mathbf{f}) \propto \exp\left(-\frac{1}{2}(\mathbf{f}-\mathbf{m})^\top \mathbf{K}^{-1}(\mathbf{f}-\mathbf{m})\right).$$

1 (b): Posterior

So the posterior is:

$$p(\mathbf{f}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{f}) \cdot p(\mathbf{f})$$

1 (b): Posterior

So the posterior is:

$$
\begin{aligned}
p(\mathbf{f}|\mathbf{y}) &\propto p(\mathbf{y}|\mathbf{f}) \cdot p(\mathbf{f}) \\
&\propto \exp\left(-\frac{1}{2}(\mathbf{y}-\mathbf{f})^\top(\sigma^2 \boldsymbol{I})^{-1}(\mathbf{y}-\mathbf{f})\right) \cdot \exp\left(-\frac{1}{2}(\mathbf{f}-\mathbf{m})^\top \mathbf{K}^{-1}(\mathbf{f}-\mathbf{m})\right)
\end{aligned}
$$

1 (b): Posterior

So the posterior is:

$$
\begin{aligned}
p(\mathbf{f}|\mathbf{y}) &\propto p(\mathbf{y}|\mathbf{f}) \cdot p(\mathbf{f}) \\
&\propto \exp\left(-\frac{1}{2}(\mathbf{y}-\mathbf{f})^\top(\sigma^2 \boldsymbol{I})^{-1}(\mathbf{y}-\mathbf{f})\right) \cdot \exp\left(-\frac{1}{2}(\mathbf{f}-\mathbf{m})^\top \mathbf{K}^{-1}(\mathbf{f}-\mathbf{m})\right) \\
&\propto \exp\left(-\frac{1}{2}\left(\mathbf{f}^\top(\sigma^2 \boldsymbol{I})^{-1}\mathbf{f} - 2\mathbf{f}^\top(\sigma^2 \boldsymbol{I})^{-1}\mathbf{y} + \mathbf{f}^\top \mathbf{K}^{-1}\mathbf{f} - 2\mathbf{f}^\top \mathbf{K}^{-1}\mathbf{m}\right)\right) \\
&\propto \exp\left(-\frac{1}{2}\left(\mathbf{f}^\top((\sigma^2 \boldsymbol{I})^{-1} + \mathbf{K}^{-1})\mathbf{f} \; - 2 \cdot \mathbf{f}^\top((\sigma^2 \boldsymbol{I})^{-1}\mathbf{y} + \mathbf{K}^{-1}\mathbf{m})\right)\right)
\end{aligned}
$$

1 (b): Posterior

So the posterior is:

$$
\begin{aligned}
p(\mathbf{f}|\mathbf{y}) &\propto p(\mathbf{y}|\mathbf{f}) \cdot p(\mathbf{f}) \\
&\propto \exp\left(-\frac{1}{2}(\mathbf{y}-\mathbf{f})^\top(\sigma^2 \boldsymbol{I})^{-1}(\mathbf{y}-\mathbf{f})\right) \cdot \exp\left(-\frac{1}{2}(\mathbf{f}-\mathbf{m})^\top \mathbf{K}^{-1}(\mathbf{f}-\mathbf{m})\right) \\
&\propto \exp\left(-\frac{1}{2}\left(\mathbf{f}^\top(\sigma^2 \boldsymbol{I})^{-1}\mathbf{f} - 2\mathbf{f}^\top(\sigma^2 \boldsymbol{I})^{-1}\mathbf{y} + \mathbf{f}^\top \mathbf{K}^{-1}\mathbf{f} - 2\mathbf{f}^\top \mathbf{K}^{-1}\mathbf{m}\right)\right) \\
&\propto \exp\left(-\frac{1}{2}\left(\mathbf{f}^\top((\sigma^2 \boldsymbol{I})^{-1} + \mathbf{K}^{-1})\mathbf{f} \; - 2 \cdot \mathbf{f}^\top((\sigma^2 \boldsymbol{I})^{-1}\mathbf{y} + \mathbf{K}^{-1}\mathbf{m})\right)\right) \\
&\propto \exp\left(-\frac{1}{2}\mathbf{f}^\top \underbrace{((\sigma^2 \boldsymbol{I})^{-1} + \mathbf{K}^{-1})}_{=:\mathbf{K}_{\text{post}}^{-1}}\mathbf{f} \; + \; \mathbf{f}^\top \underbrace{((\sigma^2 \boldsymbol{I})^{-1}\mathbf{y} + \mathbf{K}^{-1}\mathbf{m})}_{=:\tilde{\boldsymbol{f}}}\right)
\end{aligned}
$$

1 (b): Posterior

So the posterior is:

$$
\begin{aligned}
p(\mathbf{f}|\mathbf{y}) &\propto p(\mathbf{y}|\mathbf{f}) \cdot p(\mathbf{f}) \\
&\propto \exp\left(-\frac{1}{2}(\mathbf{y}-\mathbf{f})^\top (\sigma^2 \boldsymbol{I})^{-1}(\mathbf{y}-\mathbf{f})\right) \cdot \exp\left(-\frac{1}{2}(\mathbf{f}-\mathbf{m})^\top \mathbf{K}^{-1}(\mathbf{f}-\mathbf{m})\right) \\
&\propto \exp\left(-\frac{1}{2}\left(\mathbf{f}^\top (\sigma^2 \boldsymbol{I})^{-1}\mathbf{f} - 2\mathbf{f}^\top (\sigma^2 \boldsymbol{I})^{-1}\mathbf{y} + \mathbf{f}^\top \mathbf{K}^{-1}\mathbf{f} - 2\mathbf{f}^\top \mathbf{K}^{-1}\mathbf{m}\right)\right) \\
&\propto \exp\left(-\frac{1}{2}\left(\mathbf{f}^\top ((\sigma^2 \boldsymbol{I})^{-1} + \mathbf{K}^{-1})\mathbf{f} \; - 2 \cdot \mathbf{f}^\top ((\sigma^2 \boldsymbol{I})^{-1}\mathbf{y} + \mathbf{K}^{-1}\mathbf{m})\right)\right) \\
&\propto \exp\left(-\frac{1}{2}\mathbf{f}^\top \underbrace{((\sigma^2 \boldsymbol{I})^{-1} + \mathbf{K}^{-1})}_{=:\mathbf{K}_{\text{post}}^{-1}}\mathbf{f} \; + \; \mathbf{f}^\top \underbrace{((\sigma^2 \boldsymbol{I})^{-1}\mathbf{y} + \mathbf{K}^{-1}\mathbf{m})}_{=:\tilde{\boldsymbol{f}}}\right) \\
&\propto \exp\left(-\frac{1}{2}\mathbf{f}^\top \mathbf{K}_{\text{post}}^{-1}\mathbf{f} + \mathbf{f}^\top \tilde{\boldsymbol{f}}\right)
\end{aligned}
$$

1 (b): Complete the Square

$$p(\mathbf{f}|\mathbf{y}) \propto \exp\left(-\frac{1}{2}\mathbf{f}^\top \mathbf{K}_{\text{post}}^{-1}\mathbf{f} + \mathbf{f}^\top \tilde{\boldsymbol{f}}\right)$$

Recall the technique of **completing the square**:

- Scalar: $ax^2 + bx + c = a(x + \frac{b}{2a})^2 + (c - \frac{b^2}{4a})$
- Matrix / vector: $\mathbf{x}^\top \mathbf{A}\mathbf{x} + \mathbf{x}^\top \boldsymbol{b} + c = (\mathbf{x} + \frac{1}{2}\mathbf{A}^{-1}\boldsymbol{b})^\top \mathbf{A}(\mathbf{x} + \frac{1}{2}\mathbf{A}^{-1}\boldsymbol{b}) + (c - \frac{1}{4}\boldsymbol{b}^\top \mathbf{A}^{-1}\boldsymbol{b})$ for a **symmetric** matrix $\mathbf{A}$.

In our case: $\mathbf{A} = -\frac{1}{2}\mathbf{K}_{\text{post}}^{-1}$, and $\boldsymbol{b} = \tilde{\boldsymbol{f}}$, and $c = 0$ or an arbitary const. So,

$$\mathbf{x} + \frac{1}{2}\mathbf{A}^{-1}\boldsymbol{b} = \mathbf{f} + \frac{1}{2}(-\frac{1}{2}\mathbf{K}_{\text{post}}^{-1})^{-1}\tilde{\boldsymbol{f}} = \mathbf{f} - K_{\text{post}}\tilde{\boldsymbol{f}}$$

and we omit $c - \frac{1}{4}\boldsymbol{b}^\top \mathbf{A}^{-1}\mathbf{A}\boldsymbol{b}$ because

- It is a constant w.r.t. $\mathbf{f}$.
- the exp and $\propto$ operators.

1 (b): Complete the Square

$$p(\mathbf{f}|\mathbf{y}) \propto \exp\left(-\frac{1}{2}\mathbf{f}^\top \mathbf{K}_{\text{post}}^{-1}\mathbf{f} + \mathbf{f}^\top \tilde{\boldsymbol{f}}\right)$$

Recall the technique of **completing the square**:

- Scalar: $ax^2 + bx + c = a(x + \frac{b}{2a})^2 + (c - \frac{b^2}{4a})$
- Matrix / vector: $\mathbf{x}^\top \mathbf{A}\mathbf{x} + \mathbf{x}^\top \boldsymbol{b} + c = (\mathbf{x} + \frac{1}{2}\mathbf{A}^{-1}\boldsymbol{b})^\top \mathbf{A}(\mathbf{x} + \frac{1}{2}\mathbf{A}^{-1}\boldsymbol{b}) + (c - \frac{1}{4}\boldsymbol{b}^\top \mathbf{A}^{-1}\boldsymbol{b})$ for a **symmetric** matrix **A**.

In our case: $\mathbf{A} = -\frac{1}{2}\mathbf{K}_{\text{post}}^{-1}$, and $\boldsymbol{b} = \tilde{\boldsymbol{f}}$, and $c = 0$ or an arbitary const. So,

$$\mathbf{x} + \frac{1}{2}\mathbf{A}^{-1}\boldsymbol{b} = \mathbf{f} + \frac{1}{2}(-\frac{1}{2}\mathbf{K}_{\text{post}}^{-1})^{-1}\tilde{\boldsymbol{f}} = \mathbf{f} - \boldsymbol{K}_{\text{post}}\tilde{\boldsymbol{f}}$$

and we omit $c - \frac{1}{4}\boldsymbol{b}^\top \mathbf{A}^{-1}\mathbf{A}\boldsymbol{b}$ because

- It is a constant w.r.t. **f**.
- the exp and $\propto$ operators.

1 (b): Complete the Square

$$p(\mathbf{f}|\mathbf{y}) \propto \exp\left((\mathbf{f} - \boldsymbol{K}_{\text{post}}\tilde{\boldsymbol{f}})^\top \left(-\frac{1}{2}\mathbf{K}_{\text{post}}^{-1}\right)^{-1} (\mathbf{f} - \underbrace{\boldsymbol{K}_{\text{post}}\tilde{\boldsymbol{f}}}_{:=\mathbf{f}_{\text{post}}}) \right)$$

$$\propto \exp\left((\mathbf{f} - \mathbf{f}_{\text{post}})^\top \left(-\frac{1}{2}\mathbf{K}_{\text{post}}^{-1}\right)^{-1} (\mathbf{f} - \mathbf{f}_{\text{post}}) \right)$$

Hence,

$$\mathbf{f}|\mathbf{y} \sim \mathcal{N}(\mathbf{f}_{\text{post}}, \boldsymbol{K}_{\text{post}}).$$

1 (b): Complete the Square

$$p(\mathbf{f}|\mathbf{y}) \propto \exp\left((\mathbf{f} - \boldsymbol{K}_{\text{post}}\tilde{\boldsymbol{f}})^\top \left(-\frac{1}{2}\mathbf{K}_{\text{post}}^{-1} \right)^{-1} (\mathbf{f} - \underbrace{\boldsymbol{K}_{\text{post}}\tilde{\boldsymbol{f}})}_{:=\mathbf{f}_{\text{post}}} \right)$$

$$\propto \exp\left((\mathbf{f} - \mathbf{f}_{\text{post}})^\top \left(-\frac{1}{2}\mathbf{K}_{\text{post}}^{-1} \right)^{-1} (\mathbf{f} - \mathbf{f}_{\text{post}}) \right)$$

Hence,

$$\mathbf{f}|\mathbf{y} \sim \mathcal{N}(\mathbf{f}_{\text{post}}, \boldsymbol{K}_{\text{post}}).$$

Exercise 1 (c)

(c) Derive the posterior predictive distribution $y_* | x_*, \mathbf{x}, \mathbf{y}$ for a new sample x_* from the sample data-generating process.

1 (c): Derive Predictive Posterior from Joint Distribution

Naïvely, we can compute

$$p(y_*|x_*, \mathbf{x}, \mathbf{y}) = \int p(y_*|x_*, \mathbf{x}, \mathbf{y}, \mathbf{f}) \cdot p(\mathbf{f}|\mathbf{y}, \mathbf{x}) \mathrm{d}\mathbf{f}.$$

This is cumbersome. Alternative, we can use the fact that

$$\text{if} \begin{pmatrix} \boldsymbol{a} \\ \boldsymbol{b} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \end{pmatrix}, \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ab}^\top & \Sigma_{bb} \end{pmatrix}\right), \text{ then } \quad \boldsymbol{a}|\boldsymbol{b} \sim \mathcal{N}(\boldsymbol{\mu}_{a|b}, \Sigma_{a|b})$$

where

$$\boldsymbol{\mu}_{a|b} = \boldsymbol{\mu}_a + \Sigma_{ab}\Sigma_{bb}^{-1}(\boldsymbol{b} - \boldsymbol{\mu}_b)$$
$$\Sigma_{a|b} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}$$

1 (c): Derive Predictive Posterior from Joint Distribution

Naïvely, we can compute

$$p(y_*|x_*, \mathbf{x}, \mathbf{y}) = \int p(y_*|x_*, \mathbf{x}, \mathbf{y}, \mathbf{f}) \cdot p(\mathbf{f}|\mathbf{y}, \mathbf{x}) d\mathbf{f}.$$

This is cumbersome. Alternative, we can use the fact that

$$\text{if } \begin{pmatrix} \boldsymbol{a} \\ \boldsymbol{b} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \end{pmatrix}, \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ab}^\top & \Sigma_{bb} \end{pmatrix} \right), \text{ then } \quad \boldsymbol{a}|\boldsymbol{b} \sim \mathcal{N}(\boldsymbol{\mu}_{a|b}, \Sigma_{a|b})$$

where

$$\mu_{a|b} = \mu_a + \Sigma_{ab}\Sigma_{bb}^{-1}(\boldsymbol{b} - \mu_b)$$
$$\Sigma_{a|b} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}$$

1 (c): Derive Predictive Posterior from Joint Distribution

Naïvely, we can compute

$$p(y_*|x_*, \mathbf{x}, \mathbf{y}) = \int p(y_*|x_*, \mathbf{x}, \mathbf{y}, \mathbf{f}) \cdot p(\mathbf{f}|\mathbf{y}, \mathbf{x}) d\mathbf{f}.$$

This is cumbersome. Alternative, we can use the fact that

$$\text{if } \begin{pmatrix} \boldsymbol{a} \\ \boldsymbol{b} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \end{pmatrix}, \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ab}^\top & \Sigma_{bb} \end{pmatrix} \right), \text{ then } \quad \boldsymbol{a}|\boldsymbol{b} \sim \mathcal{N}(\boldsymbol{\mu}_{a|b}, \Sigma_{a|b})$$

where

$$\boldsymbol{\mu}_{a|b} = \boldsymbol{\mu}_a + \Sigma_{ab}\Sigma_{bb}^{-1}(\boldsymbol{b} - \boldsymbol{\mu}_b)$$
$$\Sigma_{a|b} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}$$

1 (c) Derive Predictive Posterior from Joint Distribution

The joint distribution of $(\mathbf{y}, y_*)$: **Note: here is y instead of f, and $\mathbf{y} = \mathbf{f} + \epsilon$. So we have σ^2 in the in the cov. matrix.**

$$\begin{pmatrix} \mathbf{y} \\ y_* \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mathbf{m} \\ m_* \end{pmatrix}, \begin{pmatrix} \mathbf{K} + \sigma^2 \boldsymbol{I} & \mathbf{K}_* \\ \mathbf{K}_*^\top & \mathbf{K}_{**} + \sigma^2 \end{pmatrix}\right),$$

Therefore, the conditional distribution $y_* | x_*, \mathbf{x}, \mathbf{y}$ is also a Gaussian:

$$y_* | x_*, \mathbf{x}, \mathbf{y} \sim \mathcal{N}(m_* + \mathbf{K}_*^\top (\mathbf{K} + \sigma^2 \boldsymbol{I})^{-1} (\mathbf{y} - \mathbf{m}), \mathbf{K}_{**} + \sigma^2 - \mathbf{K}_*^\top (\mathbf{K} + \sigma^2 \boldsymbol{I})^{-1} \mathbf{K}_*).$$

1 (c) Derive Predictive Posterior from Joint Distribution

The joint distribution of $(\mathbf{y}, y_*)$: **Note: here is y instead of f, and $\mathbf{y} = \mathbf{f} + \epsilon$. So we have σ^2 in the in the cov. matrix.**

$$\begin{pmatrix} \mathbf{y} \\ y_* \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mathbf{m} \\ m_* \end{pmatrix}, \begin{pmatrix} \mathbf{K} + \sigma^2 \boldsymbol{I} & \mathbf{K}_* \\ \mathbf{K}_*^\top & \mathbf{K}_{**} + \sigma^2 \end{pmatrix}\right),$$

Therefore, the conditional distribution $y_* | x_*, \mathbf{x}, \mathbf{y}$ is also a Gaussian:

$$y_* | x_*, \mathbf{x}, \mathbf{y} \sim \mathcal{N}(m_* + \mathbf{K}_*^\top (\mathbf{K} + \sigma^2 \boldsymbol{I})^{-1} (\mathbf{y} - \mathbf{m}), \mathbf{K}_{**} + \sigma^2 - \mathbf{K}_*^\top (\mathbf{K} + \sigma^2 \boldsymbol{I})^{-1} \mathbf{K}_*).$$

Exercise 1 (d)

Implement the GP with squared exponential kernel, zero mean function and $\ell = 1$ from scratch for $n = 2$ observations $(\mathbf{y}, \mathbf{x})$. Do this as efficiently as possible by explicitly calculating all expensive computations by hand. Do the same for the posterior predictive distribution of y_*. Test your implementation using simulated data.
Show the standard solution.