PRACTICE FINAL EXAM – WITH CORRECTIONS

STAD91 WINTER 2026
University of Toronto Scarborough

Exam duration: 3H
No calculators will be allowed during the final exam.
A formula sheet with the usual families of distributions will be provided during the final midterm.

Read the following instructions carefully:

1. Exam is closed book and internet.
2. If a question asks you to do some calculations, you must show your work for full credit. Feel free to admit the result of a question and use it in your answer to the next ones if you cannot prove it.
3. Conceptual questions do not require long answers.
4. You will write your answers to each question in the space provided on the exam sheet.
5. After solving each question, you should write your answers immediately. Do not wait last minute to write them all at once.
6. Do not share the exam with anyone or in any platform!
7. Lastly, enjoy the problems!!!

Exercise 1 (15 points)

Select the correct answer.

1. According to the Bernstein-von Mises theorem, under certain regularity conditions, the posterior distribution asymptotically approaches a Gaussian distribution centered at the maximum likelihood estimator, effectively diminishing the influence of the prior.

   True **False**

2. In Bayesian decision theory, the Bayes estimator under a zero-one (0-1) loss function is always the posterior mean.

   True **False**

3. Gibbs sampling always achieves an acceptance probability of exactly 1 because it acts as a special case of the Metropolis-Hastings algorithm where the proposal distribution is the marginal posterior distribution of the target variable.

   True **False**

4. While the frequentist LASSO can shrink coefficients to exactly zero, the posterior draws from a Bayesian LASSO (using a Laplace prior) are never exactly zero with probability 1, meaning it does not perform exact variable selection without an ad-hoc thresholding step.

   True **False**

5. In high-dimensional Bayesian inference, as the number of predictors (p) grows much larger than the sample size (n), the data likelihood heavily dominates the posterior, making the choice of prior relatively unimportant.

   True **False**

6. The Horseshoe prior is a global-local shrinkage prior that strictly forces small coefficients to be exactly zero, performing variable selection in the same discrete manner as the frequentist LASSO.

   True **False**

7. If a Markov Chain generated by the Metropolis-Hastings algorithm has reached its stationary distribution, the subsequent samples are completely independent of one another.

   True **False**

Exercise 3 (40 points)

Let $\sigma^2 > 0$ and let $x_1, \dots, x_n$ be **fixed and known** real numbers. Consider the following Bayesian framework:

$$\theta \sim \Pi_{\sigma^2} = \mathcal{N}(0, \sigma^2)$$

$$\mathbf{Y} = (Y_1, \dots, Y_n) | \theta = \theta \sim \mathcal{N}(\theta x_1, 1) \otimes \dots \otimes \mathcal{N}(\theta x_n, 1).$$

In other words, the distribution of $\mathbf{Y}$ is that of $\theta x + \varepsilon$, where $x = (x_1, \dots, x_n)$, and where $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n) \sim \mathcal{N}(0,1)^{\otimes n}$, with θ and ε independent.

1. Show that the posterior distribution is given by

$$\Pi_{\sigma^2}(\cdot|\mathbf{Y}) = \mathcal{N}\left(\frac{\langle x, \mathbf{Y}\rangle}{\sigma^{-2} + \|x\|^2}, \frac{1}{\sigma^{-2} + \|x\|^2}\right),$$

   where $\langle x, \mathbf{Y}\rangle = \sum_{i=1}^n x_i Y_i$ and $\|x\|^2 = \langle x, x\rangle = \sum_{i=1}^n x_i^2$.

   Correction: For all $\theta \in \mathbb{R}$, we have

$$\begin{aligned}\pi(\theta|\mathbf{Y}) &\propto e^{-\frac{\theta^2}{2\sigma^2}} \prod_{i=1}^n e^{-\frac{(Y_i - \theta x_i)^2}{2}} \\ &\propto e^{-\frac{1}{2}\left(\theta^2 \sigma^{-2} + \sum_{i=1}^n (\theta^2 x_i^2 - 2\theta x_i Y_i)\right)} \\ &\propto e^{-\frac{\sigma^{-2} + \|x\|^2}{2}\left(\theta^2 - 2\theta \frac{\langle x, \mathbf{Y}\rangle}{\sigma^{-2} + \|x\|^2}\right)} \\ &\propto e^{-\frac{\sigma^{-2} + \|x\|^2}{2}\left(\theta - \frac{\langle x, \mathbf{Y}\rangle}{\sigma^{-2} + \|x\|^2}\right)^2}\end{aligned}$$

   We thus indeed have

$$\Pi_{\sigma^2}(\cdot|\mathbf{Y}) = \mathcal{N}\left(\frac{\langle x, \mathbf{Y}\rangle}{\sigma^{-2} + \|x\|^2}, \frac{1}{\sigma^{-2} + \|x\|^2}\right).$$

2. What is the posterior mean $m_{\mathbf{Y}}$? And the posterior variance $v_{\mathbf{Y}}$?

   Correction: We have

$$m_{\mathbf{Y}} = \frac{\langle x, \mathbf{Y}\rangle}{\sigma^{-2} + \|x\|^2} \quad \text{and} \quad v_{\mathbf{Y}} = \frac{1}{\sigma^{-2} + \|x\|^2}.$$

3. We consider the quadratic loss $\ell(\theta, t) = (\theta - t)^2$.

   (a) Give a Bayes estimator. We will denote it $T^\star(\mathbf{Y})$.
   Correction: It is the posterior mean: $T^\star(\mathbf{Y}) = m_{\mathbf{Y}}$.

   (b) What is the posterior risk of $T^\star(\mathbf{Y})$?
   Correction: We have

$$\rho(\Pi, T^\star(\mathbf{Y})|\mathbf{Y}) = \mathbb{E}\left[(\theta - m_{\mathbf{Y}})^2 | \mathbf{Y}\right] = v_{\mathbf{Y}} = \frac{1}{\sigma^{-2} + \|x\|^2}.$$

(c) Deduce the value of the Bayes risk $R_B(\Pi_{\sigma^2})$.

Correction: We have $R_B(\Pi_{\sigma^2}) = \mathbb{E}\left[\rho(\Pi, T^\star(\mathbf{Y})|\mathbf{Y})\right] = \frac{1}{\sigma^{-2}+\|x\|^2}$.

(d) Suppose that $\|x\|^2 > 0$ and consider the estimator $T(\mathbf{Y}) = \frac{\langle x,\mathbf{Y}\rangle}{\|x\|^2}$.

i. Determine the risk function of T (still for the quadratic loss). One can use the fact that under $\mathbb{P}_\theta$, $\mathbf{Y} \sim x\theta + \varepsilon$ with $\varepsilon \sim \mathcal{N}(0,1)^{\otimes n}$.

Correction: For all $\theta \in \mathbb{R}$, we have

$$\begin{aligned}
\mathbf{R}(\theta, T) &= \mathbb{E}_\theta\left[\left(\frac{\langle x, \mathbf{Y}\rangle}{\|x\|^2} - \theta\right)^2\right] \\
&= \frac{1}{\|x\|^4}\mathbb{E}_\theta\left[\left(\langle x, x\theta + \varepsilon\rangle - \|x\|^2\theta\right)^2\right] \\
&= \frac{1}{\|x\|^4}\mathbb{E}\left[\left(\sum_{i=1}^n x_i\varepsilon_i\right)^2\right] \\
&= \frac{1}{\|x\|^4}\sum_{i=1}^n x_i^2 \\
&= \frac{1}{\|x\|^2},
\end{aligned}$$

where we used the fact that, under $\mathbb{P}_\theta$, $\mathbf{Y} - \theta x \sim \varepsilon \sim \mathcal{N}(0,1)^{\otimes n}$.

ii. Is the estimator T minimax?

Correction: Since the risk function of T is constant, we have $\mathbf{R}_{\max}(T) = \frac{1}{\|x\|^2}$. Moreover, by considering the sequence of prior distributions $(\Pi_k = \mathcal{N}(0,k))_{k\geqslant 1}$, we have

$$\lim_{k\to+\infty} R_B(\Pi_k) = \lim_{k\to+\infty}\frac{1}{\frac{1}{k} + \|x\|^2} = \frac{1}{\|x\|^2} = \mathbf{R}_{\max}(T).$$

By a result from the course, T is minimax.

4. Let $\theta_0 \in \mathbb{R}$. In this question, suppose that the real numbers x_i are such that $\|x\|^2$ tends to $+\infty$ as $n \to +\infty$, and we are interested in the consistency of the posterior distribution at θ_0.

(a) Show that for all $\delta > 0$, we have

$$\mathbb{P}\left(|\theta - \theta_0| \geqslant \delta|\mathbf{Y}\right) \leqslant \frac{1}{\delta^2}\left(v_{\mathbf{Y}} + (m_{\mathbf{Y}} - \theta_0)^2\right).$$

Correction: By Markov's inequality applied to the square, we have

$$\begin{aligned}
\mathbb{P}\left(|\theta - \theta_0| \geqslant \delta|\mathbf{Y}\right) &\leqslant \frac{1}{\delta^2}\mathbb{E}\left[(\theta-\theta_0)^2|\mathbf{Y}\right] \\
&= \frac{1}{\delta^2}\mathbb{E}\left[(\theta - m_{\mathbf{Y}})^2 + (m_{\mathbf{Y}} - \theta_0)^2 + 2(\theta - m_{\mathbf{Y}})(m_{\mathbf{Y}} - \theta_0)|\mathbf{Y}\right] \\
&= \frac{1}{\delta^2}\left(v_{\mathbf{Y}} + (m_{\mathbf{Y}} - \theta_0)^2\right),
\end{aligned}$$

where we used the fact that $\mathbb{E}\left[(\theta - m_{\mathbf{Y}})(m_{\mathbf{Y}} - \theta_0)|\mathbf{Y}\right] = 0$.

(b) Using Chebyshev's inequality, show that

$$\frac{\langle x, \varepsilon \rangle}{\|x\|^2} \xrightarrow{\mathbb{P}} 0.$$

Correction: Let $\delta > 0$. Since the variable $\langle x, \varepsilon \rangle$ is centered, we have, by Chebyshev's inequality,

$$\mathbb{P}\left(\left|\frac{\langle x, \varepsilon \rangle}{\|x\|^2}\right| \geqslant \delta\right) \leqslant \frac{1}{\delta^2 \|x\|^4} \mathrm{Var}\left(\sum_{i=1}^n x_i \varepsilon_i\right) = \frac{1}{\delta^2 \|x\|^2} \xrightarrow[n \to \infty]{} 0,$$

since $\|x\|^2 \to +\infty$ by hypothesis.

(c) Deduce that $m_{\mathbf{Y}} \xrightarrow{\mathbb{P}_{\theta_0}} \theta_0$, and conclude.

Correction: Under $\mathbb{P}_{\theta_0}$, we have $\mathbf{Y} \sim x\theta_0 + \varepsilon$. Thus,

$$m_{\mathbf{Y}} - \theta_0 \sim \frac{\langle x, x\theta_0 + \varepsilon \rangle}{\sigma^{-2} + \|x\|^2} - \theta_0 = \frac{\langle x, \varepsilon \rangle - \sigma^{-2}\theta_0}{\sigma^{-2} + \|x\|^2}.$$

Using the previous question, the fact that $\|x\|^2 \to +\infty$, and continuity, we indeed obtain $m_{\mathbf{Y}} - \theta_0 \xrightarrow{\mathbb{P}} 0$. Moreover, $v_{\mathbf{Y}} = \frac{1}{\sigma^{-2} + \|x\|^2} \xrightarrow[n \to \infty]{} 0$. Thus, the inequality from question 4(a) gives

$$\mathbb{P}\left(|\theta - \theta_0| \geqslant \delta | \mathbf{Y}\right) \xrightarrow{\mathbb{P}_{\theta_0}} 0.$$

The posterior distribution is therefore consistent at θ_0.

Exercise 2 (10 points)

Consider the Poisson model $\mathcal{P} = (\mathcal{P}(\theta)^{\otimes n})_{\theta > 0}$, and we wish to test

$$H_0 : \theta = 1 \quad \text{against} \quad H_1 : \theta \neq 1.$$

To do this, we consider the balanced loss function and the prior distribution

$$\Pi = \alpha\delta_1 + (1-\alpha)\mathrm{Exp}(1),$$

with $\alpha \in]0, 1[$.

1. Calculate $\Pi(\{1\}|\mathbf{X})$.

   Correction:

$$\begin{aligned}\Pi(\{1\}|\mathbf{X}) &= \frac{\alpha\prod_{i=1}^n\left(\frac{e^{-1}1^{X_i}}{X_i!}\right)}{\alpha\prod_{i=1}^n\left(\frac{e^{-1}1^{X_i}}{X_i!}\right) + (1-\alpha)\int_{\mathbb{R}_+^*\setminus\{1\}} e^{-\theta}\prod_{i=1}^n\left(\frac{e^{-\theta}\theta^{X_i}}{X_i!}\right)d\theta} \\ &= \frac{\alpha e^{-n}}{\alpha e^{-n} + (1-\alpha)\int_0^{+\infty}\theta^{n\overline{X}_n}e^{-(n+1)\theta}d\theta} \\ &= \frac{\alpha e^{-n}}{\alpha e^{-n} + (1-\alpha)\frac{\Gamma(n\overline{X}_n+1)}{(n+1)^{n\overline{X}_n+1}}}.\end{aligned}$$

2. Show that a Bayes test is given by

$$\varphi^\star(\mathbf{X}) = \mathbb{1}_{\left\{\frac{(n\overline{X}_n)!e^n}{(n+1)^{n\overline{X}_n+1}} \geqslant c_\alpha\right\}},$$

   where c_α is a constant depending only on α that you will specify.

   Correction: We know from the course that a Bayes test is given by $\varphi^\star(\mathbf{X}) = \mathbb{1}_{\Pi(\{1\}|\mathbf{X})\leqslant 1/2}$. However

$$\begin{aligned}\Pi(\{1\}|\mathbf{X}) \leqslant \frac{1}{2} &\iff \alpha e^{-n} \leqslant (1-\alpha)\frac{\Gamma(n\overline{X}_n+1)}{(n+1)^{n\overline{X}_n+1}} \\ &\iff \frac{(n\overline{X}_n)!e^n}{(n+1)^{n\overline{X}_n+1}} \geqslant \frac{\alpha}{1-\alpha}.\end{aligned}$$

   We therefore indeed have the desired form with $c_\alpha = \frac{\alpha}{1-\alpha}$.

Exercise 4 (35 points)

Let $n \geqslant 1$. Consider the following Bayesian framework for classification:

$$\mathbf{f} = (f_1, \dots, f_n) \sim \mathcal{N}(0, K)$$
$$\mathbf{y} = (y_1, \dots, y_n) \mid \mathbf{f} \sim \bigotimes_{i=1}^{n} \mathcal{B}(\Phi(f_i))$$

where K is the covariance matrix evaluated at the inputs x_i, $\mathcal{B}$ denotes the Bernoulli distribution, and $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution.

Because the likelihood is not Gaussian, the posterior distribution is intractable. To formulate a Gibbs sampler, we introduce continuous latent variables $\mathbf{z} = (z_1, \dots, z_N)$ such that, for $i = 1, \dots, N$:

$$z_i \mid f_i \sim \mathcal{N}(f_i, 1)$$
$$y_i = \mathbb{1}_{\{z_i > 0\}}$$

1. Show that marginalizing out the latent variable z_i recovers the original probit likelihood $p(y_i = 1 \mid f_i) = \Phi(f_i)$. Deduce that the posterior distribution $p(\mathbf{f} \mid \mathbf{y})$ is the same in both models.

   Correction: By the definition of the indicator function and the law of total probability:

$$\begin{aligned} p(y_i = 1 \mid f_i) &= \int_0^{\infty} p(z_i \mid f_i)\, dz_i \\ &= \int_0^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(z_i - f_i)^2\right) dz_i \end{aligned}$$

   Applying the substitution $u = z_i - f_i$, the limits of integration change from $[0, \infty)$ to $[-f_i, \infty)$:

$$p(y_i = 1 \mid f_i) = \int_{-f_i}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right) du$$

   Because the standard normal distribution is symmetric around zero, the probability mass from $-f_i$ to ∞ is exactly equal to the cumulative probability up to f_i, which is $\Phi(f_i)$.

2. Derive the full conditional distribution $p(z_i \mid f_i, y_i)$ up to a normalizing constant. How would you sample from this distribution?

   Correction: Using Bayes' theorem, the conditional (as a function of z_i) is proportional to the joint distribution of z_i, y_i and f_i:

$$\begin{aligned} p(z_i \mid f_i, y_i) &\propto p(z_i, f_i, y_i) \\ &= p(y_i \mid z_i, f_i) p(z_i \mid f_i) p(f_i) \\ &\propto p(y_i \mid z_i) p(z_i \mid f_i) \\ &= \left(\mathbb{1}_{\{z_i > 0\}}\right)^{y_i} \left(\mathbb{1}_{\{z_i \leqslant 0\}}\right)^{1-y_i} \frac{1}{\sqrt{2\pi}} e^{-\frac{(z_i - f_i)^2}{2}}. \end{aligned}$$

   Therefore, $p(z_i \mid f_i, y_i)$ is a Truncated Normal distribution:

   - If $y_i = 1$, $z_i \sim \mathcal{N}(f_i, 1)$ truncated to the interval $(0, \infty)$, with density

$$p(z_i \mid f_i, y_i) = \mathbb{1}_{\{z_i > 0\}} \frac{1}{\Phi(f_i)} \frac{1}{\sqrt{2\pi}} e^{-\frac{(z_i - f_i)^2}{2}}$$

- If $y_i = 0$, $z_i \sim \mathcal{N}(f_i, 1)$ truncated to the interval $(-\infty, 0]$, with density

$$p(z_i \mid f_i, y_i) = \mathbb{1}_{\{z_i \leqslant 0\}} \frac{1}{\Phi(-f_i)} \frac{1}{\sqrt{2\pi}} e^{-\frac{(z_i - f_i)^2}{2}}$$

To sample from these distributions, we can use a rejection sampling algorithm with $\mathcal{N}(f_i, 1)$ as a proposal distribution, and only accept samples falling in the relevant interval.

3. Derive the full conditional distribution of the latent function values, $p(\mathbf{f} \mid \mathbf{z}, \mathbf{y})$.

Correction: Once $\mathbf{z}$ is conditioned upon, the observed labels $\mathbf{y}$ provide no additional information about $\mathbf{f}$, meaning $\mathbf{f}$ and $\mathbf{y}$ are conditionally independent given $\mathbf{z}$ ($\mathbf{f} \perp\!\!\!\perp \mathbf{y} \mid \mathbf{z}$):

$$\begin{aligned} p(\mathbf{f} \mid \mathbf{z}, \mathbf{y}) &\propto p(\mathbf{f}, \mathbf{z}, \mathbf{y}) \\ &= p(\mathbf{y} \mid \mathbf{z}, \mathbf{f}) p(\mathbf{z} \mid \mathbf{f}) p(\mathbf{f}) \\ &= p(\mathbf{y} \mid \mathbf{z}) p(\mathbf{z} \mid \mathbf{f}) p(\mathbf{f}) \\ &\propto p(\mathbf{z} \mid \mathbf{f}) p(\mathbf{f}). \end{aligned}$$

The conditional posterior is derived using the standard Bayesian Gaussian regression formulas:

$$p(\mathbf{z} \mid \mathbf{f}) p(\mathbf{f}) = \mathcal{N}(\mathbf{z} \mid \mathbf{f}, I) \mathcal{N}(\mathbf{f} \mid 0, K)$$

By completing the square in the exponent for the product of these two Gaussians, we find:

$$p(\mathbf{f} \mid \mathbf{z}) = \mathcal{N}(\mathbf{f} \mid \mu_{\text{post}}, \Sigma_{\text{post}})$$

where:

$$\begin{aligned} \Sigma_{\text{post}} &= (K^{-1} + I)^{-1} \\ \mu_{\text{post}} &= \Sigma_{\text{post}} z \end{aligned}$$

4. Describe the steps of a Gibbs sampling algorithm to sample from the posterior distribution $p(\mathbf{f} \mid \mathbf{y})$. Specify the distributions used at each step.

Correction: To sample from the intractable marginal posterior $p(\mathbf{f} \mid \mathbf{y})$, the Gibbs sampler constructs a Markov chain that alternates between sampling from the different conditionals of the joint posterior $p(\mathbf{f}, \mathbf{z} \mid \mathbf{y})$. By simply ignoring the generated samples for $\mathbf{z}$, we obtain valid samples from the marginal distribution $p(\mathbf{f} \mid \mathbf{y})$.

After initializing $\mathbf{f}^{(0)}$, we repeat the following steps for iterations $s = 1, \ldots, S$:

i. **Sample the latent variables $\mathbf{z}^{(s)}$:**
Conditioned on $\mathbf{f}^{(s-1)}$ and $\mathbf{y}$, the variables z_i are independent. For each $i = 1, \ldots, N$, we sample from the Truncated Normal distribution derived in Part 2 (with $\mathcal{TN}$ the truncated normal distribution):

$$z_i^{(s)} \sim \begin{cases} \mathcal{TN}(f_i^{(s-1)}, 1, (0, \infty)) & \text{if } y_i = 1 \\ \mathcal{TN}(f_i^{(s-1)}, 1, (-\infty, 0]) & \text{if } y_i = 0 \end{cases}$$

ii. **Sample the latent function values $\mathbf{f}^{(s)}$:**
Conditioned on the newly sampled $\mathbf{z}^{(s)}$, we sample the entire vector $\mathbf{f}$ from the Multivariate Normal distribution derived in Part 3:

$$\mathbf{f}^{(s)} \sim \mathcal{N}\left(\mu_{\text{post}}^{(s)}, \Sigma_{\text{post}}\right)$$

where $\Sigma_{\text{post}} = (K^{-1} + I)^{-1}$ and $\mu_{\text{post}}^{(s)} = \Sigma_{\text{post}} \mathbf{z}^{(s)}$.

Note that Σ_{post} does not depend on $\mathbf{z}$ or $\mathbf{y}$, so it only needs to be computed once before the loop begins.

In the following, we consider the predictive extension of our model. Let the latent function f be endowed with a Gaussian process prior with mean zero and covariance kernel k. By the definition of a Gaussian process, any finite collection of its evaluations is jointly Gaussian. Therefore, for n training inputs $x_1, \ldots, x_n$ and any new test input x_*, the corresponding function values $\mathbf{f} = (f(x_1), \ldots, f(x_n)) = (f_1, \ldots, f_n)$ and $f_* = f(x_*)$ follow a joint multivariate normal distribution.

5. Suppose we have collected S posterior samples of the latent function values at the training points from our Gibbs sampler, denoted as $\{\mathbf{f}^{(1)}, \ldots, \mathbf{f}^{(S)}\}$. We receive a new test input x_*. Let K_* be the $1 \times n$ cross-covariance vector between x_* and the training inputs, and $K_{**} = k(x_*, x_*)$ be the prior variance at x_*.

   Describe the step-by-step procedure to generate a single sample of the predicted class label $y_*^{(s)}$, given a single MCMC sample $f^{(s)}$. Provide the distribution for each step.

   Correction:

   i. **Sample the test latent function value, $f_*^{(s)}$:**
   By the properties of Gaussian processes, f_* and $\mathbf{f}$ are jointly Gaussian. Conditioning on the noise-free sample $\mathbf{f}^{(s)}$ yields:

   $$f_*^{(s)} \sim \mathcal{N}(\mu_*, \Sigma_*)$$

   where $\mu_* = K_* K^{-1} \mathbf{f}^{(s)}$ and $\Sigma_* = K_{**} - K_* K^{-1} K_*^T$.

   ii. **Sample the test auxiliary variable, $z_*^{(s)}$:**
   Following the data augmentation setup:

   $$z_*^{(s)} \sim \mathcal{N}(f_*^{(s)}, 1)$$

   iii. **Determine the test class label, $y_*^{(s)}$:**
   Apply the deterministic threshold:

   $$y_*^{(s)} = \begin{cases} 1 & \text{if } z_*^{(s)} > 0 \\ 0 & \text{if } z_*^{(s)} \leqslant 0 \end{cases}$$

   We can also directly sample $y_*^{(s)} \sim \mathcal{B}(\Phi(f_*^{(s)}))$

6. The covariance matrix K depends on the choice of kernel function (e.g., the Squared Exponential kernel) and its hyperparameters, such as the lengthscale l. If you were to use an extremely small lengthscale, what do you expect will happen to the true posterior distribution of f_*?

Correction: As the lengthscale $l \to 0$, the correlation between any two distinct points drops to zero. Consequently, the cross-covariance vector K_* between a new test point x_* and the training inputs X approaches the zero vector ($\mathbf{0}$).

Looking at the posterior predictive parameters for a Gaussian process:

$$\mu_* = K_* K^{-1} \mathbf{f} \xrightarrow{l \to 0} 0$$
$$\Sigma_* = K_{**} - K_* K^{-1} K_*^T \xrightarrow{l \to 0} K_{**}$$

Therefore, the true posterior distribution of f_* reverts entirely to the prior distribution, $\mathcal{N}(0, K_{**})$. The model fails to generalize, as the training data provides no information about unobserved locations.