

Linear Regression Problems

ELG 5218 – Uncertainty Evaluation in Engineering Measurements and Machine Learning

Instructor: Miodrag Bolić, University of Ottawa

Date: January 28, 2026

PART A: CONCEPTUAL QUESTIONS

A1. Classical vs Bayesian Linear Regression

Question. What is the fundamental difference between classical (frequentist) linear regression and Bayesian linear regression?

A2. Why is Bayesian Linear Regression Conjugate (with Gaussian Prior)?

Question. With a Gaussian likelihood and a Gaussian prior on w , why is Bayesian linear regression conjugate? What does conjugacy buy us?

A3. Ridge Regression as MAP in Bayesian Linear Regression

Question. Explain the relationship between ridge regression https://en.wikipedia.org/wiki/Ridge_regression and Bayesian linear regression with a zero-mean Gaussian prior on w .

A4. Aleatoric vs Epistemic Uncertainty in Linear Regression

Question. In Bayesian linear regression, distinguish aleatoric and epistemic uncertainty in the predictive distribution.

A5. Behavior of Predictive Uncertainty Far from Training Data

Question. Qualitatively, how does the predictive variance $\sigma^2 + x_*^\top \Sigma_N x_*$ behave when x_* lies far outside the span of the training inputs?

A6. Effect of a Strong Prior on Posterior and Predictions

Question. What happens to the posterior over w and the predictive distribution if the prior precision λ_0 becomes very large (strong prior), assuming the prior mean is zero?

A7. Non-IID

You observe a long time series $\{(x_t, y_t)\}_{t=1}^\infty$ where the underlying relationship slowly drifts:

$$y_t = w_t^\top x_t + \epsilon_t, \quad w_t = w_{t-1} + \eta_t,$$

with small process noise η_t .

- (a) Why is a static Bayesian linear regression model (fixed w) misspecified in this scenario?

PART B: MATHEMATICAL DERIVATIONS

Assume the standard model:

$$y \mid w \sim \mathcal{N}(Xw, \sigma^2 I_N), \quad w \sim \mathcal{N}(m_0, \lambda^{-1}).$$

B1. Posterior Predictive Distribution

Problem. For a new input $x_* \in \mathbb{R}^D$, derive the posterior predictive distribution

$$p(y_* \mid x_*, X, y).$$

B2. Gradient of the Log-Posterior (for MAP / Optimization)

Problem. Derive the gradient of the log-posterior $\nabla_w \log p(w \mid X, y)$ under the conjugate Gaussian model (with fixed σ^2, m_0, S_0).

B3. Hessian and Concavity of the Log-Posterior

Problem. Derive the Hessian $\nabla_w^2 \log p(w \mid X, y)$.

PART C: PARAMETRIC ANALYSIS (What if we change parameters?)

C1. Effect of Increasing Prior Precision λ

Question. As λ increases (stronger prior), what happens to the posterior covariance Σ_N and predictive variance?

C2. Effect of Increasing Noise Variance σ^2

Question. As σ^2 increases (more observation noise), what happens to the posterior and predictive distribution?

C3. Behavior as $N \rightarrow \infty$ (Bernstein–von Mises Intuition)

Question. Intuitively, what happens to the posterior over w and the predictive distribution as $N \rightarrow \infty$ while the model is correctly specified?

PART D: Other problems

D1. Engineering Application – Temperature Sensor Calibration

You calibrate a temperature sensor: input is a voltage x , output is temperature y . You model:

$$y_n = w_0 + w_1 x_n + \epsilon_n, \quad \epsilon_n \sim \mathcal{N}(0, \sigma^2),$$

with prior

$$w \sim \mathcal{N}(0, 10^2 I_2), \quad \sigma^2 = 1 \text{ (assumed known)}.$$

You collect $N = 20$ calibration data points spanning the range $x \in [-1, 1]$.

- (a) Explain why Bayesian linear regression is preferable to simple least squares for this calibration problem.
- (b) Your posterior summary for w_1 is approximately $\mathbb{E}[w_1 | \cdot] = 2.0$, $\text{sd}(w_1 | \cdot) = 0.2$. Interpret this physically.
- (c) For a new measurement at $x_* = 1.5$ (slightly outside the calibration range), your predictive distribution is $y_* \sim \mathcal{N}(3.1, 1.4^2)$. Comment on both the mean and the inflated variance.