Math 50 Final – Additional practice problems

November 19, 2025

Exercise 1: Suppose X_1 and X_2 are Normal random variables and

$$\Sigma = \begin{bmatrix} 2 & -1/2 \\ -1/2 & 3 \end{bmatrix} \tag{1}$$

Find the conditional distributions of $X_1|X_2$ and $X_2|X_1$. Assume X_1 and X_2 have zero mean.

Exercise 2: Gaussian process models come from marginalizing over the prior distribution in a nonlinear regression model of the form

$$\beta_j \sim \text{Normal}(0, \tau_j^2), \quad E[\beta_i\beta_j] = 0, \quad i \neq j \tag{2}$$

$$f(x) = \sum_{j=1}^{\infty} \beta_j \phi_j(x) \quad \text{(assuming the sum converges with probability 1)} \tag{3}$$

For two points x_1 and x_2, find the distribution of $f(x_1)|f(x_2)$. The answer should be written in terms of a function $K(x, x') = \sum_{j=1}^{\infty} \tau_j^2 \phi_j(x)\phi_j(x')$ called a *kernel*.

Exercise 3: Which of the following indicates that the assumptions of a linear regression model may be violated? Explain your answer

1. A low R^2
2. Large confidence intervals relative to the values of the fitted coefficients
3. Cross validation is performed and finds the model has a high variance but a low bias
4. None of the above

Exercise 4: Consider the regularized least squares estimator $\hat{\beta}_R$ which minimizes

$$\sum_{i=1}^{N} (Y_i - \beta X_i)^2 + \lambda^2(\beta - b)^2 \tag{4}$$

Write $\hat{\beta}_R$ in terms of the least squares estimator $\hat{\beta}$.

Exercise 5: You are using Laplace's Rule of Succession to estimate the outcome of an election with two candidates based on polling data. Suppose the true fraction of voters supporting candidate A is 0.7. Approximately how many people must be pooled in order to ensure the 95% interval does not overlap with the prior odds of 1/2?

Exercise 6: Consider the model with interaction term

$$Y = X_1 - X_2 + 3X_3 + 2X_1X_2 + 0.5X_1X_3 + \epsilon \tag{5}$$

If $X_1 \sim \text{Normal}(0, 2)$, what is $Y|X_2, X_3$? Is this satisfy the assumptions of a linear regression model?

Exercise 7: Consider a linear regression model with no intercept ($\beta_0 = 0$) and mean zero predictor ($E[X] = 0$). Derive the sample distribution of $\hat{\beta}_1$ in terms of σ_ϵ^2 and $\hat{\sigma}_X^2$ (assuming σ_ϵ^2 is known). By derive I mean using properties of normal distributions, expectation, conditional expectation etc. .

Exercise 8: Suppose that a certain parameter θ in a model has a sample distribution

$$\hat{\theta} \sim \text{Normal}(\theta, 0.1) \tag{6}$$

where θ is the true value. Now suppose Bayesian inference is performed with priors

$$\theta \sim \text{Normal}(0, 2) \tag{7}$$

What is the posterior distribution of θ?

Solutions

1. Using the relation $\text{cov}(X_1, X_2)/\text{var}(X_1) = \beta_{1,2}$ the regression slopes. The noise terms are found from $\sigma^2_{\epsilon_2} = \sigma^2_{X_2} - \beta_{1,2}\sigma^2_{X_1}$ (the notation should be self explanatory here). This yields $X_1|X_2 \sim \text{Normal}(-\frac{1}{6}X_2, \frac{23}{12})$ and $X_2|X_1 \sim \text{Normal}(-\frac{1}{4}X_1, \frac{23}{8})$

2. You don't need to know about Gaussian processes to solve this. Use usual regression slope covariance relations: $f(x_1)|f(x_2) \sim \text{Normal}\left(\frac{K(x_1,x_2)}{K(x_2,x_2)}f(x_2), K(x_1, x_1) - \frac{K(x_1,x_2)^2}{K(x_2,x_2)}\right)$ where $K(x, x') = \sum_{j=1}^{\infty} \tau_j^2 \phi_j(x)\phi_j(x')$

3. (4) None of the above. Low R^2 indicates poor model fit but doesn't violate assumptions. Large confidence intervals suggest uncertainty but not assumption violations. High variance/low bias from cross-validation indicates overfitting, not assumption violations.

4. If we take the derivative with respect to β as set it equal to zero, we get

$$\hat{\beta}_R = \frac{\sum_{i=1}^N X_i Y_i + \lambda^2 b}{\sum_{i=1}^N X_i^2 + \lambda^2} \tag{8}$$

$$= \frac{\sum_{i=1}^N X_i Y_i}{\sum_{i=1}^N X_i^2} \frac{1}{1 + \lambda^2/(N\hat{\sigma}_X^2)} + b\frac{\lambda^2/(N\hat{\sigma}_X^2)}{1 + \lambda^2/(N\hat{\sigma}_X^2)} \tag{9}$$

$$= \hat{\beta}\frac{1}{1 + \lambda^2/(N\hat{\sigma}_X^2)} + b\frac{\lambda^2/(N\hat{\sigma}_X^2)}{1 + \lambda^2/(N\hat{\sigma}_X^2)} \tag{10}$$

This shows $\hat{\beta}_R$ interpolates between $\hat{\beta}$ and b with weights determined by $\lambda^2/(N\hat{\sigma}_X^2)$.

5. Using Laplace's Rule estimator $\hat{q} = (Y + 1)/(N + 2)$ where $Y \sim \text{Binomial}(N, 0.7)$. We have $E[\hat{q}] = (0.7N + 1)/(N + 2)$ and $\text{var}(\hat{q}) = 0.7 \cdot 0.3 \cdot N/(N + 2)^2$. For large N, $\hat{q} \approx \text{Normal}(0.7, 0.21/N)$. The 95% interval is approximately $[0.7 - 1.96\sqrt{0.21/N}, 0.7 + 1.96\sqrt{0.21/N}]$. For this not to overlap with 0.5, we need $0.7 - 1.96\sqrt{0.21/N} > 0.5$, giving $N > (1.96\sqrt{0.21}/0.2)^2 \approx 21$.

6. Taking the mean and variance we get $Y|X_2, X_3 \sim \text{Normal}(-X_2 + 3X_3, 2(1 + 2X_2 + 0.5X_3)^2 + \sigma^2_\epsilon)$. This DOES NOT satisfy the (usual) linear regression assumptions because of the dependence on X terms in the variance.

7. See class notes.

8. The statement of the sample distribution can be recast in Bayesian framework as the distribution of the estimator $\hat{\theta}$ conditioned on θ, which combined with the prior gives

$$\theta \sim \text{Normal}(0, 2) \tag{11}$$

$$\hat{\theta}|\theta \sim \text{Normal}(\theta, 0.1). \tag{12}$$

Thus the slope of θ vs. $\hat{\theta}$ is $1 \times 2/(0.1 + 2) = 20/21$ and $\text{var}(\theta|\hat{\theta}) = 2/21$. That is,

$$\theta|\hat{\theta} \sim \text{Normal}(20/21\theta, 2/21) \tag{13}$$