Spring 2026: Mathematical Statistics
Recitation Worksheet 12
April 24, 2026

1. The time it takes to apply a certain algorithm to a problem of size n is quadratic in n, but the coefficients of the formula for the time are unknown. Four tests were run with different n's and the results were as follows.

n	time (s)
1	3
2	6
3	15
4	22

The time measurements have some measurement error which you can assume to have uniform variance. If you build your X matrix correctly, you may take the following as given:

$$(X^TX)^{-1} = \begin{bmatrix} \frac{31}{4} & -\frac{27}{4} & \frac{5}{4} \\ -\frac{27}{4} & \frac{129}{20} & -\frac{5}{4} \\ \frac{5}{4} & -\frac{5}{4} & \frac{1}{4} \end{bmatrix} \quad \text{and} \quad (X^TX)^{-1}X^T = \begin{bmatrix} \frac{9}{4} & -\frac{3}{4} & -\frac{5}{4} & \frac{3}{4} \\ -\frac{31}{20} & \frac{23}{20} & \frac{27}{20} & -\frac{19}{20} \\ \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} \end{bmatrix}.$$

(a) Write down an equation for the time T in terms of n. Set up a least squares problem in matrix form to find the coefficients of the equation for the time.

(b) Compute the coefficients (you may want to use a calculator/phone).

(c) What is the estimated standard error of each coefficient?

(d) Think about which coefficients have a smaller standard error. Is the order you observe intuitive?

(e) If we had the same data, but tried to fit it to a cubic model, how big would our residuals be? Does this mean that even if we know our data is quadratic, it's better to fit it using a cubic model?

(f) If we were given more data points, but not told whether the data was quadratic or cubic, how would we determine which model to use?

2. Consider the standard linear model $y = X\beta + e$, where $e \sim N(0, \sigma^2 I)$. Suppose we assign a Gaussian prior to the parameters, $\beta \sim N(0, \tau^2 I)$. Use the fact that a Gaussian $N(\mu, \Sigma)$ in higher dimensions has pdf

$$f(x) \propto \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right).$$

(a) Write down the likelihood function $f_{\text{like}}(y|\beta)$, i.e., the pdf of the data y, up to a proportionality constant. Do the same for the prior.

(b) Use Bayes' rule to show that the posterior mode (argmax of the posterior $f_{\text{post}}(\beta|y)$) is the same as the least squares estimator with ridge regression, for a particular λ.

(c) Explain the expression you get for λ. When is it high? When is it low? What does it say about our confidence in the data vs. the prior?

(d) Reverse the calculation you did above to find which prior is equivalent to introducing LASSO regression (up to a proportionality constant). Roughly sketch the prior in 1 dimension.