In groups of 3, work out answers to the following questions on paper. No computers unless stated.

1. Linear regression minimizes \( J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (\theta^T x^{(i)} - y^{(i)})^2 \)

   Consider a “featureless” minimization problem with \( n=0 \), i.e., you have no inputs \( x \), just outputs \( y \). What is the equation for \( J \)? (Hint: you still have \( \theta_0 \)).

2. Write down a random set of \( m=5 \) or \( m=6 \) integers \( y^{(i)} \) between 0 and 10. Using graph paper (I’ll provide), plot \( J \) for \( \theta_0 = 0, 1, 2, ..., 9, 10 \). (Divide the work among the group members.) For simplicity, you can ignore the \( \frac{1}{2m} \) term. You may use a calculator, but you may not write a program. Which value minimizes \( J \)? How could you compute this value from the \( y \) values directly?

3. You may have wondered why the cost function uses the \textit{squared} errors. While this works well when the data is fairly clean, it doesn’t work well in the presence of outliers (e.g., if one of the \( y \) values from the previous problem was 1000). In the presence of outliers, a \textit{robust} cost function works better (but may be harder to minimize). One such robust function is the sum of the absolute (rather than squared) errors:

   \[ A(\theta) = \frac{1}{2m} \sum_{i=1}^{m} |\theta^T x^{(i)} - y^{(i)}| \]

   Using your data points from the previous question, plot \( A \) on a new piece of graph paper (again, ignore the \( \frac{1}{2m} \) term), and find the value that minimizes it. Is that value unique? How could you compute this value from the \( y \) values directly?

4. Normal equation

   Consider the (single-variable) linear regression problem shown above. Write down the 5x2 input matrix \( X \) and the 5x1 output vector \( y \).

   Compute \( X^T X \) as well as \( X^T y \) by hand and verify your results in Octave.

   Recall that you can solve for \( \theta \) via the normal equation

   \[ \theta = (X^T X)^{-1} X^T y \]

   Compute \( \theta \) in Octave using the “\text{inv}” function. What is 57*\( \theta \)?

   Also compute the hypothesis \( \hat{y} = X \hat{\theta} \), i.e. the predicted \( y \) values, and sketch the line in the above image (draw on this worksheet).