Q 1: Assume that we have some data $x_1, \ldots, x_n \in \mathbb{R}$. Our goal is to find the parameter b such that $J = \sum_i (x_i - b)^2$ is minimized.

1. Find an analytic solution for the optimal value of b.
2. How does this problem and its solution relate to the normal distribution? Hint: consider noisy observations by adding the noise variable ϵ, and define the new loss function. There is no need for an analytical solution here.
3. What if we change the loss from $(x_i - b)^2$ to $|x_i - b|$? Find the optimal solution of b.

Q 2: Prove that the affine functions that can be expressed by $\mathbf{x}^T\mathbf{w} + b$ are equivalent to linear functions on $(\mathbf{x}, 1)$.

Q 3: Assume that you want to find a quadratic function of $\mathbf{x}$ i.e., $f(\mathbf{x}) = b + \sum_i w_i x_i + \sum_{j \leq i} w_{ij} x_i x_j$. How would you formulate this in a neural network? Can you draw a neural network that presents this quadratic function?

Q 4: What would happen if we initialized the parameters/weights to zero? Would the algorithm still work? What if we initialized the parameters/weights with a normal distribution with a variance of 1000 rather than 0.01? Please show examples that support your answers and plot the graphs of the training and validation losses versus epoch for every initialization. Check Sections 3.4, 5.2 and 5.4 in the textbook.

Q 5: Experiment using different learning rates to find out how quickly the loss function value drops. Can you reduce the error by increasing the number of epochs of training? Check Section 3.4 in the textbook.

Q 6: If the number of examples cannot be divided by the batch size, what happens to data_iter at the end of an epoch? Check Section 3.4 in the textbook.

Q 7: Why is K-fold cross-validation very expensive to compute?

Q 8: Experiment with the value of the weight decay λ in the estimation problem in Section 3.7. Plot training and validation loss per epoch for different values of the weight decay λ. You can experiment with three situations when λ is low, high and in-between. What do you observe?

Q 9: We know that $\|\mathbf{w}\|^2 = \mathbf{w}^T\mathbf{w}$. Can you write a similar formula (known as the Frobenius norm) for matrices (see the Frobenius norm in Section 2.3.11)?

Q 10: (a) Perform the house price prediction problem in Section 5.7 and complete your submission of Predicting House Prices on Kaggle. Add your code to your GitHub account as a public repository and add the link to your assignment. (b) Attach the screenshot of your submission in the PDF file.
(c) Explain in the PDF file the neural network architecture and the hyperparameters that you use. Provide the graph of the training and validation losses per epoch.

Hints: Please submit your assignment in a PDF or Word file. Please don't write answers by hand. Please don't include code scripts in your PDF. Instead, add the link to your GitHub repository, where all your codes are saved. Make sure to create a public repository for the assignment. You can include graphs that you plotted in your PDF assignment.