Question 1

Please submit Question1.pdf on Moodle using the Final Exam - Question 1 object. You must submit a single PDF. You may submit multiple .py files (placed in a single zip file) if you wish. **Do not put your pdf in the zip file.** The parts are worth (2 + 4+ 4 + 3) + (3+3+6) = 25.

(a) (**Bias, Variance & MSE**) Let $X_1, \dots, X_n$ be i.i.d. random variables with mean μ and variance σ^2. Define the estimator $T = \sum_{i=1}^n a_i X_i$ for some constants $a_1, \dots, a_n$.

(i) What condition must the a_i's satisfy to ensure that T is an unbiased estimator of μ? Unbiased means that $\mathbb{E}T = \mu$.
What to submit: your working out, either typed or handwritten.

(ii) Under the condition identified in the previous part, which choice of the a_i's will minimize the MSE of T? Does this choice of a_i's surprise you? Provide some brief discussion. **Hint: this is a constrained minimization problem.**
What to submit: your working out, either typed or handwritten, and some commentary.

(iii) Suppose that instead, we let $a_i = \frac{1+b}{n}$ for $i = 1, \dots, n$, where b is some constant. Find the value of b (in terms of μ and σ^2) which minimizes the MSE of T. How does the answer here compare to the estimator found in the previous part? How do the two compare as the sample size n increases? Are there any obvious issues with using the result of this question in practice?
What to submit: your working out, either typed or handwritten, and some commentary.

(iv) Suppose now that you are told that $\sigma^2 = \mu^2$. For the choice of b identified in the previous part, find the MSE of the estimator in this setting. How does the MSE compare to the MSE of the sample average $\overline{X}$? (Recall that MSE$(\overline{X}) = \text{var}(\overline{X}) = \sigma^2/n = \mu^2/n$). Further, explain whether or not you can use this choice of b in practice.
What to submit: your working out, either typed or handwritten, and some commentary.

(b) **(kNN Regression)** Consider the usual data generating process $y = f(x) + \epsilon$, where f is some unknown function, and ϵ is a noise variable with mean zero and variance σ^2. Recall that in kNN regression, we look at the k nearest neighbours (in our dataset) of an input point x_0, we then consider their corresponding response values and average them to get a prediction for x_0. Given a dataset $D = \{(x_i, y_i)\}_{i=1}^n$ we can write down the kNN prediction as

$$\hat{m}(x_0) = \frac{1}{k} \sum_{i \in \mathcal{N}_k(x_0)} y_i,$$

where $\mathcal{N}_k(x_0)$ is the set of indices of the k nearest neighbours of x_0. Without loss of generality, label the k nearest neighbours of x_0 as $z_1, \dots, z_k$ and their corresponding response values by $t_1, \dots, t_k$.

(i) Show that

$$[\text{bias}(\hat{m}(x_0))]^2 = \left(f(x_0) - \frac{1}{k} \sum_{i=1}^k f(z_i) \right)^2.$$

Throughout, you should treat x_0 as a fixed point (not a random variable). You may use any results from tutorials, labs or lectures without proof[1].
What to submit: your working out, either typed or handwritten.

(ii) Derive an expression for the variance $\text{var}(\hat{m}(x_0))$.
What to submit: your working out, either typed or handwritten.

[1] Please reference any results that you use, e.g. by stating that a particular result follows from Tutorial A, question B, part C.

(iii) Using the results so far, write down an expression for the MSE of $\hat{m}(x_0)$. Describe what happens to the bias of the kNN estimator at x_0 when k is very small (1NN), and what happens when k is very large ($k \to \infty$). Similarly, what happens to the variance? What does this tell you about the relationship between bias and variance and choice of k?
What to submit: your working out, either typed or handwritten, and some commentary.