AIE6001 Assignment 3: Basics of Algorithms, Optimization

Due date: 11:59 PM, Monday, November 17, 2025.

Remark: The Maximum point is 100.

Question 1 (Algorithms)**. Simple conceptual questions (16 points/4 points for each question.)**

- *What does algorithm efficiency mean? What are two types of algorithm efficiency measures?*
- *What does algorithm robustness mean? Given one example of robust algorithm.*
- *What does algorithm stability mean? What's the difference of algorithm stability and robustness?*
- *Given two commonly seen definition of algorithm accuracy. Why do we sometimes prefer approximate algorithms?*

Question 2 (Quicksort)**.** *What's the complexity (exact number of comparisons) for Quicksort to sort the array [1, 3, 2, 4, 5, 7, 6] and [1, 2, 3, 4, 5, 6, 7], respectively?* *(4 points)*

Question 3 (Worst-case complexity of quicksort)**.** *Show that the worst case of quick sort takes $\mathcal{O}(n^2)$ operations. (5 points)*

Question 4 (Fourier transform of a delayed signal)**.** *Show that*

$$\mathcal{F}(x(t-\tau)) = e^{-i2\pi f\tau}X(f).$$

(5 points)

Question 5 (Steps for deriving FFT)**.** *Let x_n be a signal that is 0 outside the interval $0 \le n \le N-1$. Suppose N is even. Let $e_n = x_{2n}$ represent the even-indexed samples, and let $o_n = x_{2n+1}$ represent the odd-indexed samples*

1) *Show that e_n and o_n are zero outside the interval $0 \le n \le (N/2)-1$.* *(5 points)*
2) *Show that*

$$\tilde{x}_k = \frac{1}{2}\tilde{E}_k + \frac{1}{2}W_N^k\tilde{O}_k, \quad k = 0, 1, \ldots, N-1,$$

where $W_N = e^{-i\frac{2\pi}{N}}$, and

$$\tilde{E}_k = 2\sum_{n=0}^{N/2-1} e_n W_{N/2}^{nk}, \quad \tilde{O}_k = 2\sum_{n=0}^{N/2-1} o_n W_{N/2}^{nk}.$$

(5 points)

3) *Show that*

$$\tilde{E}_{k+N/2} = \tilde{E}_k, \quad \tilde{O}_{k+N/2} = \tilde{O}_k.$$

(5 points)

Question 6 (Application: Medical image estimation). *The positron emission tomography (PET) imaging can be modeled as using Poisson observations to recover an image inside patient's area-of-interest. The image is represented by $\{\mu_i\}$, $i = 1, ..., n$. The data collection involves m measurements on the patient. If location i emits a positron, it is detected in the jth measurement with probability p_{ji}. We assume the probabilities p_{ji} are given (with $p_{ji} > 0$ and $\sum_{j=1}^{m} p_{ji} \leq 1$). The total number of events recorded in the jth measurement is denoted by y_j,*

$$y_j = \sum_{i=1}^{n} y_{ji}, \quad j = 1, \ldots, m,$$

where the variables y_{ji} have Poisson distribution with means $p_{ji}\mu_i$. (A random variable $Z \sim$ Poisson(λ) measures $\mathbb{P}(Z = k) = \lambda^k e^{-\lambda}/k!$, $k = 0, 1, 2, \ldots$. Another useful property is that the sum of Poisson random variables also follows the Poisson distribution, whose parameter is the sum of individual parameters.)

1) *Formulate the maximum likelihood estimation problem of estimating the means μ_i, based on observed values of y_j, $j = 1, \ldots, m$.* *(5 points)*
2) *Will the maximum likelihood function returns a unique maximizer?* *(5 points)*
3) *Discuss how can we introduce regularizer to improve this algorithm?* *(5 points)*

Question 7 (Implementing bisection). *Write your own code to implement bisection algorithm to find the $\alpha = 0.9$-quantile of a t-distribution with $n = 5$ degrees of freedom. Start with initial interval $[1.291, 2.582]$. Stop when the length of the interval is less than 10^{-4}.* *(10 points)*

Question 8 (Logistic regression). *Given n observations (x_i, y_i), $i = 1, \ldots, n$, $x_i \in \mathbb{R}^p$, $y_i \in \{0, 1\}$, parameters $a \in \mathbb{R}^p$ and $b \in \mathbb{R}$. Consider the log-likelihood function for logistic regression:*

$$\ell(a, b) = \sum_{i=1}^{n} \{y_i \log h(x_i; a, b) + (1 - y_i) \log(1 - h(x_i; a, b))\}$$

1) *Derive the Hessian H of this function and show that H is negative semi-definite (this implies that ℓ is concave and has no local maxima other than the global one.)* *(5 points)*
2) *Use data* logit-x.dat *and* logit-y.dat*, which contain the predictors $x_i \in \mathbb{R}^2$ and response $y_i \in \{0, 1\}$ respectively for logistic regression problem. Implement Newton's method for optimizing $\ell(a, b)$ and apply it to fit a logistic regression model to the data. Initialize Newton's method with $a = 0$, $b = 0$. Plot the value of the log likelihood function versus iterations. What are the coefficients a and b from your fit?* *(10 points)*

Question 9 (Locally weighted linear regression). *Consider a linear regression problem in which we want to weight different training examples differently. Specifically, suppose we want to minimize*

$$J(\theta) = \frac{1}{2} \sum_{i=1}^{n} w_i (\theta^T x_i - y_i)^2.$$

In class, we have worked out what happens for the case where all the weights are the same. In this problem, we will generalize some of those ideas to the weighted setting, and also implement the locally weighted linear regression algorithm.

1) *Show that $J(\theta)$ can also be written as*

$$J(\theta) = (X\theta - y)^T W (X\theta - y)$$

for an appropriate diagonal matrix W, matrix X and vector y. State clearly what these matrices and vectors are. *(2 points)*

2) *Suppose we have samples (x_i, y_i), $i = 1, \ldots, n$ of n independent examples, but in which the y_i's were observed with different variances, and*

$$p(y_i|x_i, \theta) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp(-\frac{(y_i - \theta^T x_i)^2}{2\sigma_i^2})$$

i.e. y_i has mean $\theta^T x_i$ and variance σ_i^2 (where σ_i^2 are fixed, known, constants). Show that finding the maximum likelihood estimate of θ reduces to solving a weighted linear regression problem. State clearly what the w_is are in terms of σ_i^2's. *(3 points)*

3) *Use data* rx.dat *and* ry.dat*, which contain the predictors x_i and response y_i respectively for our problem. Implement gradient descent for (unweighted) linear regression that we derived in class on this dataset, and plot on the same figure the data and the straight line resulting from your fit. (Remember to include the intercept term.)* *(5 points)*

4) *Implement locally weighted linear regression on this dataset, using gradient descent, and plot on the same figure the data and the line resulting from your fit. Using the following weights*

$$w_i = \exp(-x_i^2/(20)).$$

Plot the $J(\theta)$ versus iterations. *(5 points)*