Here is a quiz on Linear Regression, covering weighted least squares, its probabilistic interpretations, and a coding problem for Locally Weighted Linear Regression (LWLR).

Linear Regression: Weighted and Probabilistic Interpretations

This assignment explores advanced concepts in Linear Regression, including weighted least squares and its probabilistic foundations. Please answer the theoretical questions and complete the coding task.

Part (a): Theoretical Questions

Question 1: Weighted Least Squares Cost Function in Matrix Form [2 points]

The weighted least squares cost function is defined as $J(\theta) = \sum_{i=1}^{m} w^{(i)} (\theta^T x^{(i)} - y^{(i)})^2$.

Show that $J(\theta)$ can also be written in matrix form as $J(\theta) = (X\theta - y)^T W (X\theta - y)$, where X is the design matrix, y is the vector of target values, and W is an appropriate matrix. Clearly specify the value of each element of the matrix W.

Question 2: Generalized Normal Equation for Weighted Linear Regression [4 points]

In standard (unweighted) linear regression, the normal equation is $X^T X \theta = X^T y$, and the closed-form solution for θ is $(X^T X)^{-1} X^T y$.

By finding the derivative $\nabla_\theta J(\theta)$ of the weighted cost function $J(\theta) = (X\theta - y)^T W (X\theta - y)$ and setting it to zero, generalize the normal equation to this weighted setting. Provide the new value of θ that minimizes $J(\theta)$ in closed form as a function of X, W, and y.

Question 3: Probabilistic Interpretation of Weighted Linear Regression [4 points]

Suppose we have a dataset ${(x^{(i)}, y^{(i)}); i = 1, \dots, m}$ of m independent examples. We model the $y^{(i)}$'s as drawn from conditional distributions with different levels of variance $(\sigma^{(i)})^2$. Specifically, assume the model:

$p(y^{(i)}|x^{(i)}; \theta) = \frac{1}{\sqrt{2\pi(\sigma^{(i)})^2}} \exp\left(-\frac{(y^{(i)} - \theta^T x^{(i)})^2}{2(\sigma^{(i)})^2}\right)$

That is, each $y^{(i)}$ is drawn from a Gaussian distribution with mean $\theta^T x^{(i)}$ and variance $(\sigma^{(i)})^2$ (where the $\sigma^{(i)}$'s are fixed, known constants).

Show that finding the maximum likelihood estimate of θ for this model reduces to solving a weighted linear regression problem. State clearly what the weights $w^{(i)}$ are in terms of the $\sigma^{(i)}$'s.

Question 4: Probabilistic Assumptions for Standard Least Squares [Bonus/Conceptual - 2 points]

Briefly describe the key probabilistic assumptions about the error term $\epsilon^{(i)}$ in the model $y^{(i)} = \theta^T x^{(i)} + \epsilon^{(i)}$ that lead to the standard (unweighted) least squares cost function being equivalent to finding the maximum likelihood estimate of θ.

Part (b): Coding Problem - Locally Weighted Linear Regression (LWLR) [10 points]

In this part, you will implement Locally Weighted Linear Regression (LWLR).

Task:

Complete the implementation of Locally Weighted Linear Regression in the file src/p05b_lwr.py.

Your implementation should use the normal equations you derived in Part (a) for the weighted setting.

Weight Function:

The weight $w^{(i)}$ for each training example $(x^{(i)}, y^{(i)})$ when making a prediction at a query point x should be calculated using the following formula:

$w^{(i)}(x) = \exp\left(-\frac{\|x^{(i)} - x\|^2}{2\tau^2}\right)$

where τ is a bandwidth parameter that controls how quickly the weights fall off with distance.

Dataset:

You will be working with the dataset provided in data/ds5_{train,valid,test}.csv. Note that $x^{(i)}$ in this dataset is 1-dimensional.

Requirements:

1. Implement the predict method within the LWLR class (or equivalent function) that takes a query point x and returns its predicted y value. This method should internally compute the weights $w^{(i)}(x)$ for all training examples relative to the query point x, construct the W matrix, and then use the generalized normal equation to find θ for that specific query point.

2. Ensure your code is well-commented and follows good programming practices.

3. (Optional, but recommended for testing) Include functionality to evaluate your LWLR model on the validation/test sets using an appropriate metric (e.g., Mean Squared Error).