Exam 1

Benny Chen

October 13, 2023

Problem 1

1. Logistic regression can be used for classification.
2. When solving a convex optimization problem using gradient descent, the algorithm will always converge since all local minima are also global minima.
3. If $X \sim N(H, \sigma^2)$ and $Y = aX + b$, then the variance of Y is $a\sigma^2$?.

Solutions

1. True. Logistic regression is a classification algorithm that is used to predict the probability of a categorical variable. In logistic regression, the variable is binary so either 1 (yes, success, etc.) or 0 (no, failure, etc.). It does this by using a sigmoid function to map predicted values to probabilities.
2. True. Due to the convexity of the function, the gradient descent algorithm will always converge to one global minimum. With that, gradient descent will always converge to the global minimum.
3. False. If $X \sim N(H, \sigma^2)$ and $Y = aX + b$, then the variance of Y is $a^2\sigma^2$ and not $a\sigma^2$.

Problem 2

Given N training data points $\{(x_k, y_k)\}$, $k = 1, 2, \ldots, N$, x_k in R^d, and labels as y_k in $\{0, 1\}$ (either -1 or 1), we seek a linear discriminant function $f(x) = w \cdot x_k = \sum_{j=1}^{d} w_j x_{k,j}$ (where $x_{k,j}$ is the feature value of attribute j of a data points x_k) optimizing a special loss function $L(z) = e^{-z}$ where $z = yf(x)$.
Let $\eta > 0$ be the learning rate, please derive the gradient update Δw_k for a randomly elected data point k in the stochastic gradient descent (SGD) method.

$$\Delta w_k = -\eta \frac{\partial L(z)}{\partial w_k} \tag{1}$$

$$\Delta w_k = -\eta \frac{\partial L(z)}{\partial z} \frac{\partial z}{\partial w_k} \tag{2}$$

$$\Delta w_k = \eta e^{-z} \frac{\partial z}{\partial w_k} = \eta e^{-z} y x_k \tag{3}$$

Problem 3

In the linear regression taught in the lecture, all the data points are considered to be of "equal" weight. In reality, we could assign weight for each of them based on different "importance" to reduce influence of some potentially noisy data points and focus on the important ones.
Consider the weighted least squares problem in which you are given a dataset $\{x_i, y_i, w_i\}$, $i = 1, \ldots, N$, where w_i is an importance weight attached to the i-th data point. The loss is defined as $L(\beta) = \sum_i^N w_i (y_i - \beta^T x_i)^2$.
Please provide the derivation process to find an expression to estimate the coeffcients $\hat{\beta}$ in closed form.

$$Y = [y_1, y_2, \ldots, y_N]^T \tag{4}$$

$$X = \begin{bmatrix} X_1^T \\ X_2^T \\ \vdots \\ X_N^T \end{bmatrix} \tag{5}$$

Then.

$$L(\beta) = (Y - X\beta)^T W (Y - X\beta) \tag{6}$$

$$L(\beta) = (Y^T - \beta^T X^T) W (Y - X\beta) \tag{7}$$

$$L(\beta) = Y^T W Y - Y^T W X \beta - \beta^T X^T W Y + \beta^T X^T W X \beta \tag{8}$$

$$\frac{\partial L(\beta)}{\partial \beta} = -2X^T W Y + 2X^T W X \beta = 0 \tag{9}$$

$$X^T W X \beta = X^T W Y \tag{10}$$

$$\beta = (X^T W X)^{-1} X^T W Y \tag{11}$$