SuperAGI Assignment

QnA

Ans 1. The likely relation between w_newn and w_newn+1 is that they are equal with their sum adding up to the older weight of the pre duplication feature i.e. w_n. No relation between w_new0 and w_new1 since these features were not duplicated and per the assumption of logistic regression these must be independent features, hence these weights shall not be affected by duplication of a third feature.

Ans 3. What is the approximate computational cost of each gradient descent iteration of logistic regression in modern well written packages?

When the feature matrix is sparse, meaning that each sample has, on average, only k non-zero entries, the computational cost can be calculated like so:

1. Forward Propagation:

In the case of sparse matrices, the dot product between the feature vector and the parameter vector can be computed more efficiently by considering only the non-zero entries. Therefore, the time complexity for each sample becomes O(k)O(k) instead of O(n)O(n). With mm samples, the overall complexity is O(m·k).

2. Loss Calculation:

Calculating the logistic loss for each sample still involves simple arithmetic operations, so the time complexity remains O(m)

3. Backward Propagation:

Computing the gradient involves another dot product operation for each parameter, and in the case of sparse matrices, this can also be done more efficiently by considering only the non-zero entries. Therefore, the time complexity becomes O(k) instead of O(n). Since there are n parameters, the overall complexity is O(n*k).

4. Parameter Update:

Updating each parameter still involves a constant time complexity, so it remains O(n).

Summing up, the final complexity is O(m*k + n*k)