Contlo Assignment

Name: Pallavi Dhakne

Roll No:BT20CSE201

Q/A Assignment (10 points)

Q1.

Answer :-

When you duplicate feature n into a feature (n+1) and retrain a new model, the weights of the original feature n and the duplicated feature (n+1) might change due to the model's optimization process. However, the relationship between these weights is not straightforward and can vary depending on multiple factors such as the nature of the data, regularization techniques used, and the convergence of the optimization algorithm.

In general, the weights w_new_n and w_new_n+1 for the duplicated features might not necessarily be the same as w_n and could differ significantly. The exact relationship between w_new_0,w_new_1,w_new_n and w_new_n+1

cannot be determined without analyzing the specific data and the model training process.

When duplicating a feature and retraining the model, the new weights are likely to be influenced by interactions with other features, potential multicollinearity issues between the duplicated features, and the way the model adapts during training.

Q2.

Answer :-

To determine the best-performing email template with 95% confidence, we need to analyze the results based on statistical significance.

Given the click-through rates:

Template A: 10%

Template B: 7%

Template C: 8.5%

Template D: 12%

Template E: 14%

It appears that Template E has the highest CTR at 14%. However, to ascertain the significance of these results and compare them against Template A, statistical testing is necessary.

Comparing Template A to the others:

Template E: With a 95% confidence level, Template E outperforms Template A as its CTR is significantly higher.

Template B: At a 95% confidence level, Template B's CTR is significantly lower than Template A's.

Template C: Similar to Template B, Template C also performs significantly worse than Template A with a 95% confidence level.

Template D: With Template D, though it shows a higher CTR than Template A, to confirm its superiority with 95% confidence, further testing might be necessary as the conclusion isn't immediately apparent.

So, the correct statement among the options provided is:

E is better than A with over 95% confidence, B is worse than A with over 95% confidence. You need to run the test for longer to tell where C and D compare to A with 95% confidence.

This means you've determined that Template E is better than A and Template B is worse than A with a high level of confidence. However, to precisely evaluate where Templates C and D stand against Template A with 95% confidence, more data or testing is needed.

Q3.

Answer :-

The computational cost of each gradient descent iteration in logistic regression with sparse features can be estimated based on the number of non-zero entries in the feature vectors.

In modern well-written packages like Scikit-learn or TensorFlow, the computational complexity for logistic regression with sparse features is typically around O(m * k),

where: m is the number of training examples.

k is the average number of non-zero entries in each training example. This complexity arises because, during each iteration of gradient descent, the algorithm needs to compute the gradient of the cost function for each feature. With sparse features, only the non-zero entries contribute to this computation, hence the cost scales with the number of non-zero entries.

Q4.

Answer :-

Considering the different methods to generate additional training data for building Classifier V2, we can make some informed predictions about their potential impact on accuracy:

1.Labeled Data Close to Decision Boundary from V1:

Method: Running V1 classifier on 1 million random stories and selecting 10k closest to the decision boundary.

Potential Impact: This method targets challenging examples that are difficult for the current model (V1) to classify accurately. These samples could help improve the model's understanding of nuanced cases and potentially enhance its performance on borderline instances. However, it might not cover the entire spectrum of diverse examples.

2.Random Labeled Stories from 1000 News Sources:

Method: Select 10k random labeled stories from the 1000 news sources.

Potential Impact: While random selection provides diversity, it might not prioritize challenging or informative samples. It could offer a broader view of the dataset but might not contribute significantly to improving the model's performance in difficult instances.

3. Misclassified Data Farthest from Decision Boundary from V1:

Method: Labeling a subset of 10k stories from 1 million where V1 output is wrong and farthest from the decision boundary.

Potential Impact: This approach focuses on the most challenging misclassified instances, potentially addressing the weaknesses of the current model. By concentrating on samples where V1 is most uncertain or erroneous, it could aid in improving the model's accuracy on difficult cases.

The effectiveness of these methods could vary based on the quality and diversity of the additional data they provide. However, based on the potential impact described:

1. Ranking based on Likely Impact on V2 Classifier's Accuracy:

- Third Approach (Misclassified Data Farthest from Decision Boundary from V1): This method could have the highest impact as it specifically targets challenging misclassified instances.

- First Approach (Labeled Data Close to Decision Boundary from V1): This method might follow, focusing on challenging examples but might not cover the entire range of cases.

- Second Approach (Random Labeled Stories from 1000 News Sources): While it provides diversity, it might contribute less to addressing the model's weaknesses in challenging instances.

Q5.

Answer :-

The estimates for the probability **p** that a coin will come up heads given **n** tosses and **k** heads using the three methods are as follows**:**

1.Maximum Likelihood Estimate (MLE):

The MLE for (p) is the ratio of the number of heads (k) to the total number of coin tosses (n).
MLE: $p_{MLE} = k / n$

2.Bayesian Estimate:

Given a uniform prior over the range [0,1], the posterior distribution is a Beta distribution.
The expected value of the posterior distribution is the Bayesian estimate for (p).

Bayesian Estimate: $p_{Bayesian} = \{k+1\} / \{n+2\}$

3. Maximum a Posteriori (MAP) Estimate:

The MAP estimate is the mode of the posterior distribution, which, in this case, is also the peak of the Beta distribution.

MAP Estimate: MLE: $p_{MAP} = k / n$

MLE is directly based on observed frequencies and doesn't incorporate any prior information. Bayesian Estimate incorporates a uniform prior and is a compromise between the prior and the likelihood. MAP Estimate considers the mode of the posterior distribution and, in this case, aligns with the MLE due to the specific choice of a uniform prior.