

1. When duplicating feature n into feature $n + 1$ and retraining the logistic regression model, the likely relationships between the new model weights w_{new_0} , w_{new_1} , w_{new_n} , and $w_{new_{n+1}}$ are as follows: - w_{new_0} and w_{new_1} : These weights are associated with different features and are unlikely to be affected significantly by the duplication. They might retain their values learned from the new training process. - w_{new_n} and $w_{new_{n+1}}$: Due to the duplication, these weights are expected to be closely related or nearly identical. The model may assign similar weights to both features n and $n + 1$ due to the redundant information.

2. In the A/B/C/D/E email template multivariate test scenario: - Option 2 is true: E is better than A with over 95

3. The approximate computational cost of each gradient descent iteration in logistic regression, considering m training examples, n features, and sparse feature vectors with an average of k non-zero entries per example ($k \ll n$), is approximately $O(m \cdot k \cdot n)$.

4. Regarding the approaches for generating additional training data for V2 classifier: - Approach 1: Getting 10k stories closest to the decision boundary of V1 might provide more challenging but informative examples, potentially improving V2's accuracy. - Approach 2: Randomly selecting labeled stories may offer less information relevant to V2's objective of distinguishing between information and entertainment. - Approach 3: Choosing examples where V1 is both wrong and farthest from the decision boundary might offer challenging but potentially misleading examples that could confuse V2.

Based solely on accuracy, the models might rank as follows: Approach 1 $\hat{>}$ Approach 3 $\hat{>}$ Approach 2. However, the effectiveness of these approaches in improving V2's accuracy depends on various factors like sample representativeness and the boundary proximity of the selected examples.

5. The estimates for p using different methods are as follows: 1. **Maximum Likelihood Estimate (MLE)**: $p = \frac{k}{n}$ 2. **Bayesian Estimate**: $p = \frac{k+1}{n+2}$ 3. **Maximum a Posteriori (MAP) Estimate**: Same as Bayesian estimate ($p = \frac{k+1}{n+2}$), considering the mode of the posterior distribution corresponds to the point estimate.

These estimates use the observed number of heads (k) from n coin tosses to approximate the probability p of getting heads.