Q/A Assignment

1.

For i in {0, 1, 2, ….., n-2, n-1}
    w_new_i = w_i

For i in {n, n+1}
    w_new_i = (w_n / 2)

w_new_0	=	w_0
w_new_1	=	w_1
w_new_n	=	w_n / 2
w_new_n+1	=	w_n / 2

2.

c. Both D and E are better than A with 95% confidence. Both B and C are worse than A with over 95% confidence

3.

Modern well-written packages, such as scikit-learn or TensorFlow, often leverage efficient sparse matrix operations to handle sparse data.

The cost of each iteration can be roughly estimated as **O(m * k)**, where m is the number of training examples and k is the average number of non-zero entries in each example.

The sparsity of the feature vectors is a key factor. In sparse datasets, where most entries are zero, operations involving zero entries can be skipped, leading to computational savings.

4.

Running V1 on 1 Million Random Stories:
This method selects stories where V1 is close to the decision boundary. These examples are likely to be challenging for V1, and getting them labeled could help the model understand ambiguous cases better. However, it might introduce noise if the boundary is inherently fuzzy.

Getting 10k Random Labeled Stories:
This approach involves randomly selecting labeled stories. While it provides diversity, it may not focus on the challenging cases that the model struggles with. It might contribute to a more balanced and generalizable dataset but may not significantly improve performance on the borderline cases.

Random Sample of 1 Million Stories with V1 Misclassifications:
Selecting stories where V1 is both wrong and far from the decision boundary could provide valuable information on the types of errors V1 makes. These examples are likely to be significant outliers and might help the model correct specific types of mistakes made by V1.

<u>In terms of likely value for improving the accuracy of V2:</u>

Approach 3 may provide the most value because it specifically targets cases where V1 is not only wrong but also far from the decision boundary. These are likely to be the most challenging and informative examples for improving the model's performance.

Approach 1 is also valuable as it targets ambiguous cases, but it may introduce some noise if the decision boundary is inherently fuzzy.

Approach 2 contributes to dataset diversity but may not focus on the areas where the model needs improvement the most.

<u>In terms of ranking based on accuracy improvement:</u>

Approach 3
Approach 1
Approach 2

5.

Maximum Likelihood Estimate (MLE):

The MLE for the probability of heads (p) is simply the ratio of the number of heads (k) to the total number of coin tosses (n).
MLE: **p = k / n**

Bayesian Estimate:

Assuming a uniform prior over the range [0,1], the posterior distribution follows a Beta distribution. The expected value of this distribution is a Bayesian estimate.
Bayesian Estimate: **p = (k+1) / (n+2)**

Maximum a Posteriori (MAP) Estimate:

Similar to the Bayesian estimate, but the MAP estimate corresponds to the mode (peak) of the posterior distribution, which, in the case of a Beta distribution, has a closed-form solution.
MAP Estimate: **p = k / n**

So, to summarize:

MLE: **p = k / n**

Bayesian Estimate: **p = (k+1) / (n+2)**

MAP Estimate: **p = k / n**