1. Here's a likely relationship between the weights:
w_{new0} , w_{new1} w_{newn} would be similar to w_0 , w_1, w_n
w_{newn+1} would be a randomly initialized weight

2. b

3. Cost of computing the gradient = O(k)
Cost of updating the weights = O(n)
Cost of each iteration of gradient descent = O(m*k + n)

4.
The value of three different approaches would be:

a. This approach targets difficult-to-classify stories. Incorporating these examples might help improve V2's accuracy by addressing the specific edge cases where the V1 classifier struggled. It could potentially enhance V2's performance as it focuses on the most challenging examples for the current classifier.
b. It provides diversity in the training set but may not specifically target challenging cases that could significantly impact the classifier's performance. It might help in covering a broad spectrum of topics and styles but may not significantly improve V2's accuracy in dealing with complex cases.
c. It specifically targets cases where the current classifier failed and where the decision boundary was far away, indicating potential uncertainty or confusion in classification. This approach might significantly enhance V2's accuracy by specifically addressing the weaknesses of the current classifier and focusing on challenging cases.

The ranking of approaches on the basis of accuracy would be:
c > a > b
Approach c is likely to yield the highest improvement in V2's accuracy as it specifically targets cases where the current classifier struggles the most. Approach a focuses on challenging examples, while Approach b lacks a focused strategy in dealing with complex cases.

5. The value of estimates will be

a. k/n
b. (k+1)/(n+2)
c. k/n