1. It depends on how it will interact with the original feature (W0new to Wn-1new) will remain the same mainly. (Wn new and Wn+1 new) can be half or equal to the original Wn.
2. 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.
3. In case of sparse data where k<<n, the number of features does not directly factor into the computational cost of each gradient descent iteration of logistic regression . b'coz the cost is dominated by the number of non-zero entries k in each training example and the number of training examples m. so, order will be O(k*m).
4. In terms of accuracy for classifier V2, the methods are likely to rank as follows:
    1. **Method 3**: Highest accuracy. Targets V1's weaknesses.
    2. **Method 1**: Second highest accuracy. Focuses on ambiguous cases.
    3. **Method 2**: Lowest accuracy. Increases data diversity but doesn't target V1's weaknesses.
5. MLE = k/n, bayesian Estimate = k+1/n+2, MAP = k/n