**CPT\_S 570 HW4**

**Yang Zhang**

**11529139**

**1.** The paper states two points: (1) The discriminative learning has lower asymptotic error than generative learning. (2) The generative learning may approach its asymptotic error faster than discriminative learning.

**2.** **a. Let us assume that the training data satisfies the Naive Bayes assumption (i.e., features are independent given the class label). As the training data approaches infinity, which classifier will produce better results, Naive Bayes or Logistic Regression? Please explain your reasoning.**

**3.(a)** **Can we compute P(X) from the learned parameters of a Naive Bayes classifier? Please explain your reasoning.**

**Yes**

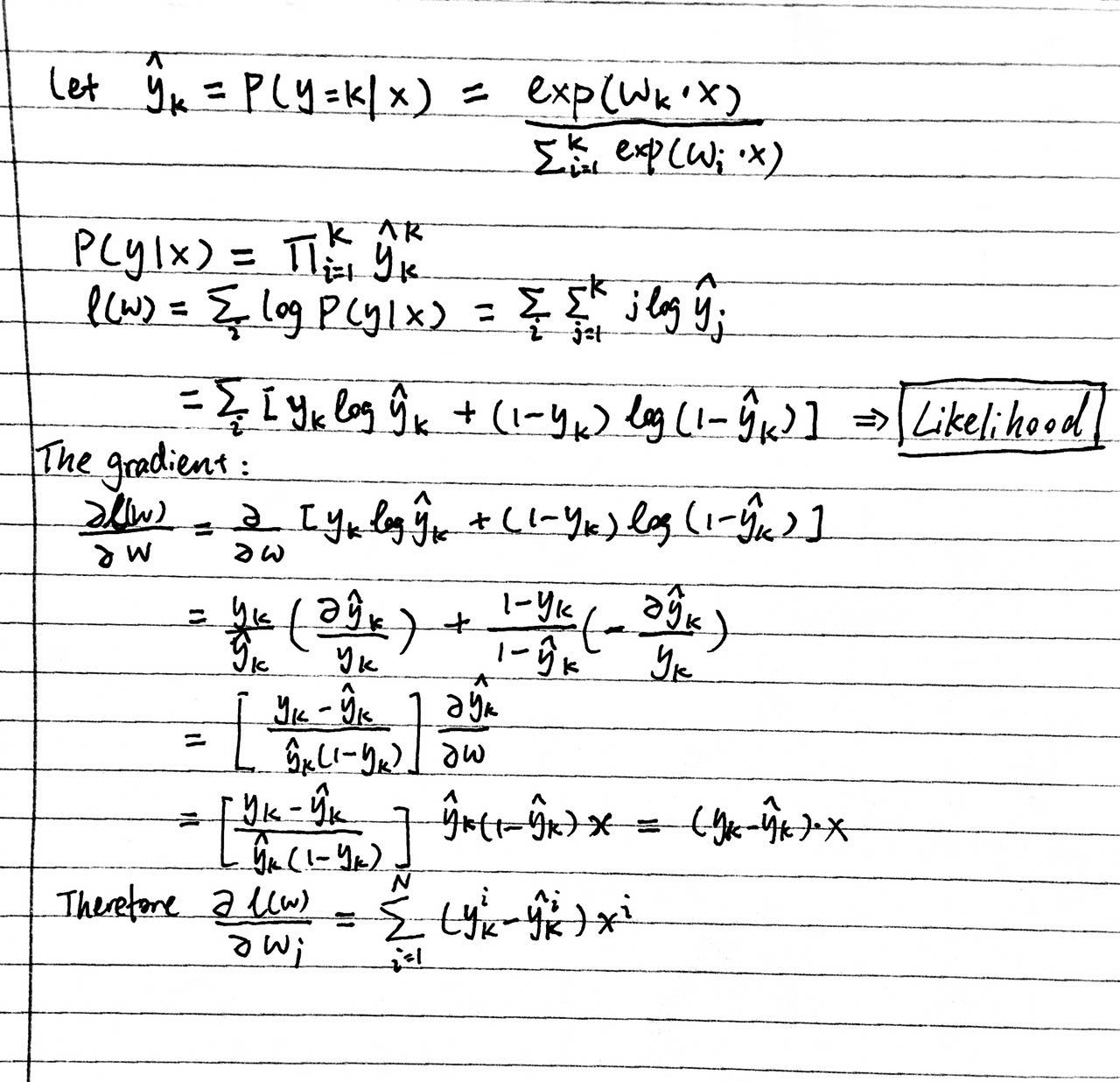
No, because Naïve Bayes doesn’t need to calculate the P(x) which is the joint distribution to estimate the p(x|y). It only need to estimate p(xi|y) for each feature i

**b. Can we compute P(X) from the learned parameters of a Logistic Regression classifier? Please explain your reasoning.**

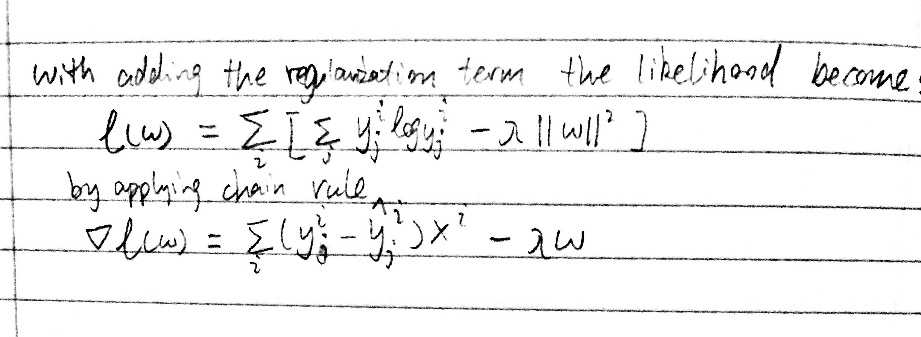
**No**

Yes, because Logistic Regression uses MLE to estimate P(x)

**4. (a)**

****

**(b)**

****

**5.** This paper compared 5 tests for the propose of picking the algorithm with better performance. As a result, the test for the difference of two proportions and paired-differences t test (based on random train/test split) should never be used, because those 2 tests are shown Type I error in certain cases. While the 10 fold cross validation test somewhat elevates the Type I error. In contrast, McNemar’s tests are shown to have low Type I error. The last test 5x2cv also provides the acceptable Type I error. As the matter of detecting differences between algorithms, the cross-validation test is the most powerful test.

**6.** This paper talked about the overfitting issue of machine learning. In most of the cases, the general objective function of a ML algorithm is to minimize the in sample error. However, by achieving this objective, the noise data has possibly been fitted as well, this is not the general solution we want. Under computing is one of the way to avoid overfitting.

**7.** This paper provided detailed introduction of popular ensemble methods (Adaboost, bootstrap .etc) and states that ensemble methods often perform better than any single classifier. The paper also explained why in 3 aspects: (1) statistical (2) computational (3) representational.

**8.** Result:

Training Accuracy: 0.9658385093167702

Testing Accuracy: 0.801980198019802

The training accuracy is higher than testing accuracy

**9**