

University of Virginia  
Department of Computer Science

**CS 6501: Text Mining**  
**Spring 2019**

**3:30pm-3:50pm, Tuesday, April 23rd**

Name:
ComputingID:

- This is a **closed book** and **closed notes** quiz. No electronic aids or cheat sheets are allowed.
- There are 2 pages, 3 parts of questions, and 20 total points in this quiz.
- The questions are printed on the **back** of this paper!
- Please carefully read the instructions and questions before you answer them.
- Please pay special attention on your handwriting; if the answers are not recognizable by the instructor, the grading might be inaccurate (*NO* argument about this after the grading is done).
- Try to keep your answers as concise as possible; grading is *not* by keyword matching.

Total	/20
-------	-----

## 1 True/False Questions (3pts×2)

For the statement you believe it is *False*, please give your brief explanation of it (you do not need to explain anything when you believe it is *True*). *Note the credit can only be granted if your explanation is correct.*

1. L2 regularization introduces sparsity in parameter estimation, which helps avoid overfitting issue in model learning.

*False, and Explain:* L2 regularization avoids overfitting by enhancing numerical stability of optimization, but it cannot really introduce sparsity.

2. To make it computationally feasible, Naive Bayes assumes that features are independent from each other.

*False, and Explain:* It only assumes conditional independence.

## 2 Multi-choice Questions (4pts×2)

1. Which of the following models can be estimated by maximum likelihood estimator:  
(b)(d)

(a) Support Vector Machines; (b) Maximum Entropy Model;  
(c) k Nearest Neighbor; (d) Naive Bayes.

2. The difference(s) between generative models and discriminative models include(s):  
(c)(d)

(a) Discriminative models capture the joint distribution between features and class labels;  
(b) Generative models assume conditional independence among features;  
(c) Generative models can effectively explore unlabeled data;  
(d) Discriminative models provide more flexibility in introducing features.

## 3 Short Answer Question (6 pts)

1. Briefly explain the idea of maximum margin classifier, and write down the realization of it in the primal form of Support Vector Machines for linearly nonseparable binary classification problem. (*Hint: use slack variables to relax the margin.*)

A maximum margin classifier aims to find the classification hyperplane that can not only minimize the misclassification error rate, but also maximize its distance to the closest individual instances. Mathematically, the primal form of SVM can be written as follows,

$$\begin{aligned} \min_{w, \xi} \quad & \frac{w^T w}{2} + \sum_i \xi_i \\ \text{s.t. } \forall i, \quad & y_i w^T x_i \geq 1 - \xi_i \\ & \xi_i \geq 0 \end{aligned}$$