University of Virginia
Department of Computer Science

# CS 6501: Text Mining
# Spring 2015

**9:30am-9:45am, Thursday, April 24th**

| 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. Smaller $k$ in kNN classifier will help reduce the risk of overfitting.
   ***False**, and Explain*: since in extreme case when $k$=1, the classification boundary is more likely to be affected by noise, e.g., tend to over fit the noise.

2. To make it computationally feasible, Naive Bayes assumes that features are independent from each other.
   ***False**, and Explain*: Naive Bayes assumes the features are conditionally independent rather than marginally independent.

# 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 dual form of Support Vector Machines tells us: (a), (b), (d)
   (a) The decision hyperplane is determined by a linear combination of support vectors;
   (b) Non-linear classification can be achieved via introducing kernels;
   (c) The feature weight vector $w$ in the primal form will be sparse (i.e., most of its elements are zeros);
   (d) SVM is good at dealing with high dimensional classification problems.

# 3 Short Questions (6 pts)

1. Write down the primal form of Support Vector Machines for linearly nonseparable binary classification problem. (Hint: with slack variables)

$$\begin{aligned} \arg\min_{w,\xi} \quad & w^\mathsf{T} w + C \sum_i \xi_i \\ \forall i, \quad & y_i w^\mathsf{T} x_i \geq 1 - \xi_i \\ & \xi_i \geq 0 \end{aligned}$$