University of Virginia
Department of Computer Science

CS 6501: Text Mining
Spring 2016

5:00pm-5:15pm, Monday, April 25th

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. Based on the Bayes decision theory, Naive Bayes classifier is optimal.
   ***False**, and Explain*: , conditional independence breaks the optimality.

2. Logistic regression is a linear classifier.
   ***False**, and Explain*: , it is a non-linear model, or to be more specific, a generalized linear model.

2 Multi-choice Questions (4pts×2)

1. Which of the following statement(s) is true about feature selection: (b), (c), (d)
   (a) wrapper method is able to identify the best set of features for a given classifier;
   (b) filter method assumes features are independent;
   (c) feature selection methods focus on testing the dependence between features and the response variable;
   (d) feature selection breaks down for rare features.

2. Which of the following statement(s) is true about the SVM classifier: (a), (b)
   (a) The decision hyperplane is determined by a linear combination of support vectors;
   (b) Non-linear classification can be achieved via a kernel trick;
   (c) It encourages a sparse model;
   (d) It minimizes 0/1 loss.

3 Short Questions (6 pts)

1. Name a generative classifier and a discriminative classifier, and specify at least two major differences between these two types of classification methods.

   Generative classifier: Naive Bayes, discriminative classifier: SVM. Several major differences: 1. generative classifiers model joint probability, while discriminative classifiers model conditional probability; 2. arbitrary features can be incorporated in discriminative classifiers, but not in generative classifiers; 3. unlabeled data can be easily incorporated in generative classifiers, but not in discriminative classifiers.