

## HW07: Probabilistic modeling

Hand in via moodle at: <https://moodle.umass.edu/course/view.php?id=20836>. Remember that only PDF submissions are accepted. We encourage using L<sup>A</sup>T<sub>E</sub>X to produce your writeups. See `hw00.tex` for an example of how to do so. You can make a `.pdf` out of the `.tex` by running “`pdflatex hw00.tex`”. You’ll need `mydefs.sty` and `notes.sty` which can be downloaded from the course page.

1. Answer the following questions about Naive Bayes:

- If we train a Naive Bayes classifier using infinite training data that satisfies all of its modeling assumptions (e.g., conditional independence), then it will achieve zero training error over these training examples. True or False? If True, give a short justification. If False a counter-example or a convincing one-sentence explanation.
- Suppose  $X$  is a vector of  $n$  boolean attributes and  $Y$  is a single discrete-valued variable that can take on  $J$  possible values. Let  $\theta_{ij} = P(X_i|Y = y_j)$ . What is the number of independent  $\theta_{ij}$  parameters?
- Consider the same problem, but now suppose  $X$  is a vector of  $n$  real-valued attributes, where each of these  $X_i$  follows a Normal (Gaussian) distribution:  $P(X_i = x_i|Y = y_j) \sim N(x_i|\mu_{ij}, \sigma_{ij})$ . How many distinct  $\mu_{ij}, \sigma_{ij}$  are there?
- We can write the classification rule for the Naive Bayes as:

$$\hat{y} = \arg \max_{y_k} \frac{P(Y = y_k) \prod_i P(X_i|Y = y_k)}{\sum_j P(Y = y_j) \prod_i P(X_i|Y = y_j)} \quad (1)$$

- We often do not compute the denominator when estimating  $\hat{y}$ . Explain why.
- Is it possible to calculate  $P(X)$  from the parameters estimated by Naive Bayes?

2. Answer the following questions about Logistic regression:

- Suppose the data satisfies the conditional independence assumption of Naive Bayes. As the number of training examples approaches infinity, which classifier produces better results, NB or LR? Justify your answer in one sentence.
- Suppose the data does not satisfy the conditional independence assumption of Naive Bayes. As the number of training examples approaches infinity, which classifier produces better results, NB or LR? Justify your answer in one sentence.
- Is it possible to calculate  $P(X)$  from the parameters estimated by Logistic Regression?