Advanced Machine Learning Project No. 1

Sargur N. Srihari
University at Buffalo, The State University of New York
Buffalo, New York 14260
Contact: 716-645-6162 (O), srihari@buffalo.edu

January 31, 2015

1 Project Goal

The goal of this project is to develop, implement and evaluate algorithms pertaining to learning and inference with probabilistic graphical models (PGMs). The project covers choice of PGM representations and inference based on sampling.

1.1 Data Sets

The data for testing your algorithms is drawn from multinomial distributions. They are in the form of samples drawn from a distribution of a set of D discrete random variables: $\mathbf{x} = [X_1, X_2, .., X_D]$, where values taken by the variables are $X_i \in \{x_i^0, x_i^1, .., x_i^{d_i-1}\}$, $i = 1, ..D$ and $x_i^j = j, j = 0, 1, .., d_i - 1$. The models and algorithms are to be tested from the domain of children's handwriting (Section 6) where $D = 12$ and d_i range between 3 and 5 (see Figure 2). There are two data sets, one for cursive writing and the other for hand-print. There are also three snapshots of the data corresponding to three instances of time (2011-12, 2012-13, 2013-14). Some values may be ambiguous or missing as discussed in Appendix B.

2 Representations

Construct and evaluate a family of PGMs for the same data set. At one extreme assume all the variables are independent (no edges) and at the other extreme all variables are dependent on each other (66 edges). Construct PGMs that are in-between by using appropriate learning algorithms (Section 4). Begin with Bayesian networks and then experiment with Markov networks.

3 Inference Algorithms

Develop inference algorithms for determining the mean and entropy of each distribution and the relative entropy between distributions using two methods: (i) from the samples provided (ii)

from samples obtained from the PGMs (using Monte Carlo methods such as ancestral sampling and Gibbs sampling).

Determine the number of samples needed for accurate estimation using each PGM– by comparing with direct estimation.

3.1 Mean of a Distribution $p(\mathbf{x})$

- For a distribution $p(\mathbf{x})$, its mean is:

$$E[p(\mathbf{x})] = \sum_{\mathbf{x}} \mathbf{x} p(\mathbf{x}) \tag{1}$$

- It can be estimated using N samples $\{\mathbf{x}_k\}, k = 1, .., N$ as:

$$\hat{E}[p(\mathbf{x})] = \frac{1}{N} \sum_{k=1}^{N} \mathbf{x}_k \tag{2}$$

Note that for a Bayesian Network, $p(\mathbf{x}) = \prod_{i=1}^{D} p\big(X_i | pa(X_i)\big)$, but inference using Eq. 1 is infeasible since it will involve summation over all possible values of $\mathbf{x}$ which is a very large number, e.g., $4^{12} > 16$ million.

3.2 Entropy of a Distribution

- The entropy a distribution $p(\mathbf{x})$ is:

$$H[p(\mathbf{x})] = -\sum_{\mathbf{x}} p(\mathbf{x}) \ln p(\mathbf{x}) \tag{3}$$

- It can be estimated using N samples as:

$$\hat{H}[p(\mathbf{x})] = -\frac{1}{N} \sum_{k=1}^{N} \ln p(\mathbf{x}_k) \tag{4}$$

3.3 Relative Entropy between two Distributions

- The relative entropy, or K-L divergence, between two distributions over a multinomial variable $\mathbf{x}$ is defined as:

$$KL(p||q) = -\sum_{\mathbf{x}} p(\mathbf{x}) \left[\ln q(\mathbf{x}) - \ln p(\mathbf{x})\right]. \tag{5}$$

- It can be estimated using N samples as:

$$\hat{KL}(p||q) = -\frac{1}{N}\sum_i \left[\ln q(\mathbf{x}_k) - \ln p(\mathbf{x}_k)\right]. \tag{6}$$

4 Learning

4.1 Bayesian Networks

Develop an algorithm to learn a Bayesian network for the probability distribution $p(\mathbf{x})$ from samples. Clearly indicate the assumptions made in BN construction, including: (i) the measure of deviance from independence between two variables, e.g., Pearson's χ^2 test or mutual information, (ii) the measure of goodness of the network, e.g., log-loss, and (iii) any limitations on the number of parents per node.

4.2 Markov Networks

Develop an algorithm to learn a Markov network for a probability distribution given samples from that distribution. Indicate all assumptions made and discuss the efficiency and accuracy of the MN learning algorithm.

5 Project Deliverables and Due Dates

Prepare a report addressing the following issues. Use a format such as a conference paper for submission to NIPS or ICML. Include appropriate graphs and charts addressing the following issues.

5.1 PGM Representations

Use the constructed PGMs to perform inferences such as described above. How accurate is the PGM in computing mean, entropy and K-L divergence when compared to direct computation from the data?

Discuss trade-off between accuracy and complexity of each PGM to determine the PGM that leads to the best results and least complexity.

Discuss the relative advantages of Bayesian networks and Markov networks based on your experiments.

5.2 PGM Learning Algorithms

Give detailed descriptions of the algorithms used. Discuss the efficiency (speed) of the learning algorithm.

5.3 Temporal Analysis

What do the differences in entropy and relative entropy for the same students in different years (2011-12, 2012-13, 2013-14) indicate?

5.4 Common Writing

For representative PGMs show samples in the data set that are "most similar" to the mean. Use an appropriate dissimilarity measure between two D-dimensional vectors $X = [X_1, .., X_D]$ and $Y = [Y_1, .., Y_D]$, e.g.,

$$d(X, Y) = \frac{1}{D} \sum_{i=1}^{D} \frac{|X_i - Y_i|}{\max(X_i)} \tag{7}$$

where max is over the possible values that X_i can take.

5.5 Rare Writing

Determine the most unusual (or least probability) samples in the data sets by calculating the probability of each sample in the data set.

5.6 Due Date

Hand-in a report before March 13 (last day of classes before Spring Break).

6 Appendix A: Handwriting Data set

The handwriting samples are from a large number of students as they are learning (2nd grade) or have just learned (3rd and 4th grade) how to produce cursive and printed writing (2012). This is continued through the student's primary educational career (2013, 2014). The collection occurs annually, every spring for 3 years, as the student's handwriting and hand printing skills continue to mature and as individual handwriting habits develop. It is hoped to continue for 11 years; to follow these 2nd, 3rd and 4th graders of 2011 all the way through to being high school graduates.

The students write a paragraph four times each year; two times in cursive writing and two times by hand printing.

To gather the writing samples, each student was asked to copy the following paragraph, in both hand printing and cursive writing, during class:

The brown fox went into the barn where he saw the black dog. After a second, the black dog saw the fox too. The brown fox was fast and quick. The black dog was not fast and he lost the fox. The fox hid in a hole and waited for the black dog to go home. After the black dog went home, the fox was able to go to the hole he called home and saw all the other foxes. The other foxes were glad to see him and they all asked him to tell them about his day.

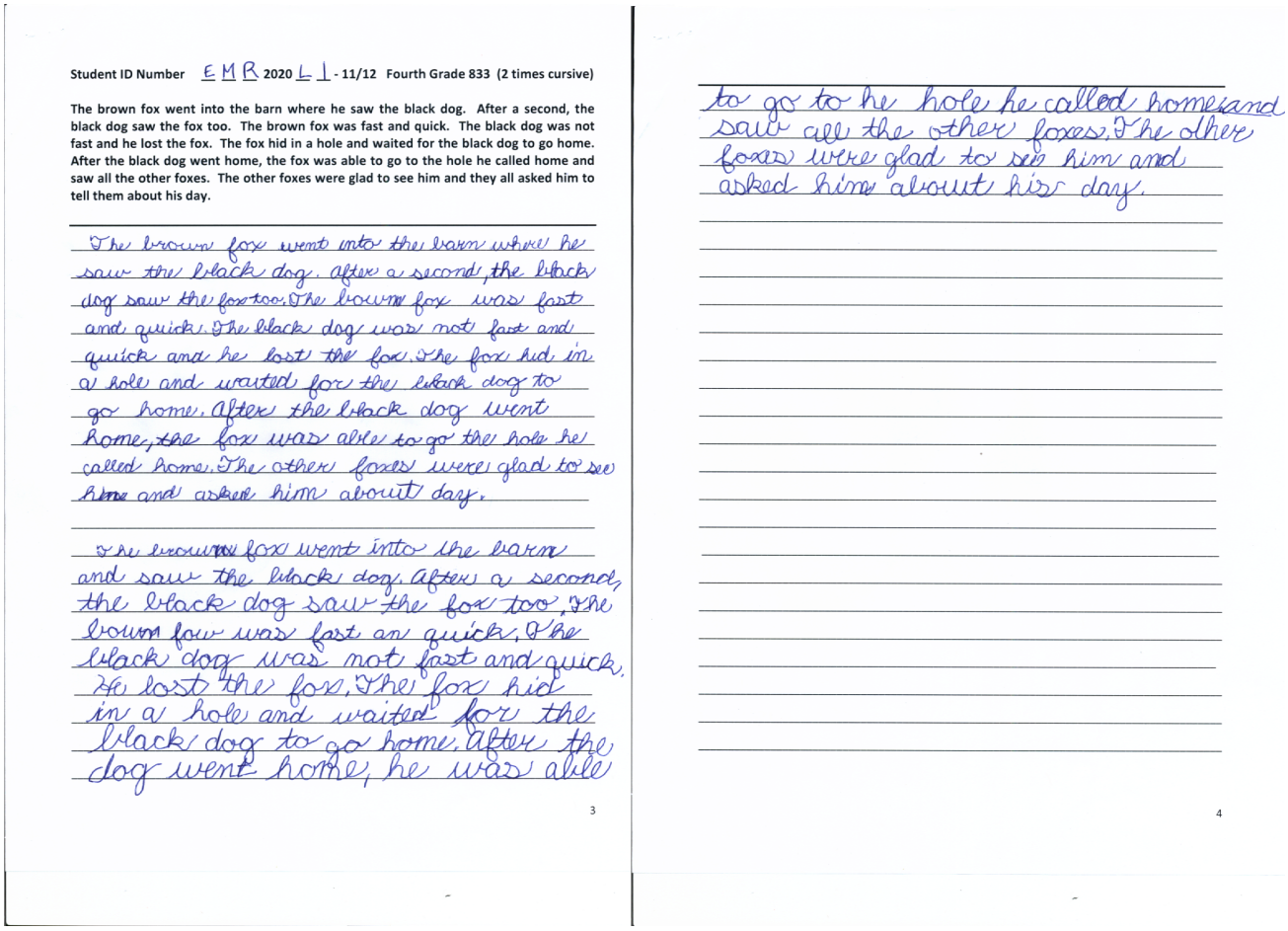
Samples of the handwritten paragraphs are shown in Figure 1.

The word *and* appears five times in the paragraph and students were asked to write the paragraph a total of four times; twice using hand printing and twice using cursive writing. This data gathering process produced an average of 5 to 10 samples of the word *and* per student in hand printing and in cursive (except for second graders) to be analyzed using a *Truthing Tool* by a human examiner. As a result the word *and*, is represented by a set of D features, $\mathbf{X} = \{X_i\}, i = 1, .., D$ where feature X_i takes one of d_i discrete values.

Samples of handwritten *and* by children of grade 2, 3 and 4 and their respective feature values are shown in Figure 3. Out of these, the samples for Handprint and Cursive writing in Grade 4 were extracted from the paragraphs in Figure 1. It is these values you will use to construct the PGMs.

7 Appendix B: Missing and Inconsistent Data

In the data set you will notice that some attributes are given a value of -1 or 99. The value -1 means that the value of that attribute was unassigned. The value of 99 means that the feature is inconsistent, e.g., "a shorter than n" in one sample, and "a taller than n" in another sample. If you directly use them as the feature value you will get incorrect results. One way to handle missing data is to determine the most likely value for that variable given the values of other variables. Thus you will need to make two passes through the data set in constructing the PGMs.



Student ID Number E M R 2020 L I - 11/12 Fourth Grade 833 (2 times cursive)

The brown fox went into the barn where he saw the black dog. After a second, the black dog saw the fox too. The brown fox was fast and quick. The black dog was not fast and he lost the fox. The fox hid in a hole and waited for the black dog to go home. After the black dog went home, the fox was able to go to the hole he called home and saw all the other foxes. The other foxes were glad to see him and they all asked him to tell them about his day.

The brown fox went into the barn where he saw the black dog. After a second, the black dog saw the fox too. The brown fox was fast and quick. The black dog was not fast and he lost the fox. The fox hid in a hole and waited for the black dog to go home. After the black dog went home, the fox was able to go the hole he called home. The other foxes were glad to see him and asked him about day.

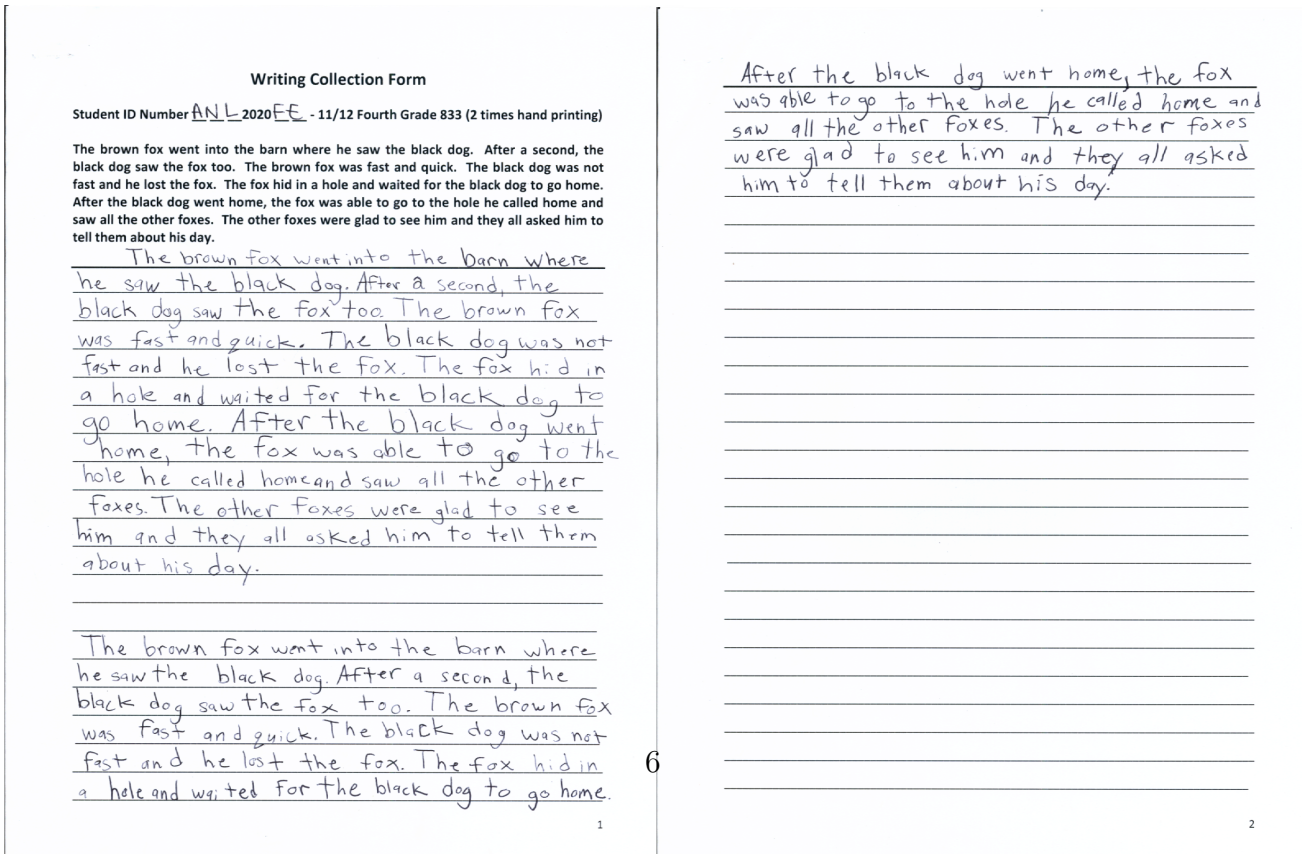
The brown fox went into the barn and saw the black dog. After a second, the black dog saw the fox too. The brown fox was fast an quick. The black dog was not fast and quick. He lost the fox. The fox hid in a hole and waited for the black dog to go home. After the dog went home, he was able

3

to go to he hole he called home and saw all the other foxes. The other foxes were glad to see him and asked him about his day.

4

(a) Cursive Writing Sample



Writing Collection Form

Student ID Number A N L 2020 E E - 11/12 Fourth Grade 833 (2 times hand printing)

The brown fox went into the barn where he saw the black dog. After a second, the black dog saw the fox too. The brown fox was fast and quick. The black dog was not fast and he lost the fox. The fox hid in a hole and waited for the black dog to go home. After the black dog went home, the fox was able to go to the hole he called home and saw all the other foxes. The other foxes were glad to see him and they all asked him to tell them about his day.

The brown fox went into the barn where he saw the black dog. After a second, the black dog saw the fox too. The brown fox was fast and quick. The black dog was not fast and he lost the fox. The fox hid in a hole and waited for the black dog to go home. After the black dog went home, the fox was able to go to the hole he called home and saw all the other foxes. The other foxes were glad to see him and they all asked him to tell them about his day.

The brown fox went into the barn where he saw the black dog. After a second, the black dog saw the fox too. The brown fox was fast and quick. The black dog was not fast and he lost the fox. The fox hid in a hole and waited for the black dog to go home.

1

After the black dog went home, the fox was able to go to the hole he called home and saw all the other foxes. The other foxes were glad to see him and they all asked him to tell them about his day.

2

(b) Printed Writing Sample

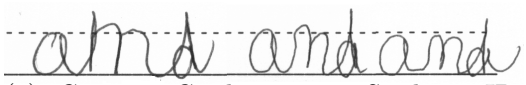
Initial stroke of “a”	staff right	staff left	staff center	
Formation of “a” staff	tented	retraced	looped	no staff
Number of “n” arches	one	two		
Shape of “n” arches	pointed	rounded	retraced	combination
Location of “n” mid	above base	below base	at base	
Formation of “d” staff	tented	retraced	looped	
Formation of “d” initial	overhand	underhand	straight across	
Formation of “d” terminal	curved up	straight	curved down	no obvious end stroke
Symbol	unusual		symbol	
a-n relationship	a taller	a equal	a smaller	
a-d relationship	a taller	a equal	a smaller	
n-d relationship	n taller	n equal	n smaller	

(a)

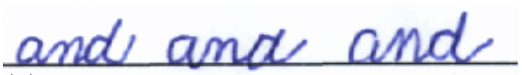
# strokes in “a”	one	two	three	uppercase	
formation of “a” staff	tented	retraced	looped	no staff	single down
# strokes in “n”	one	two	three	uppercase	
formation of “n” staff	tented	retraced	looped	no staff	single down
shape of arch of “n”	pointed	rounded			
# strokes in “d”	one	two	three	uppercase	
formation of “d” staff	tented	retraced	looped	no staff	single down
initial stroke of “d”	staff top	bulb			
Unusual formations	formation		symbol		
a-n relationship	a taller	a equal	a smaller		
a-d relationship	a taller	a equal	a smaller		
n-d relationship	n taller	n equal	n smaller		

(b)

Figure 2: Twelve features of *and* ($D = 12$) together with their possible values: (a) cursive and (b) hand-print.

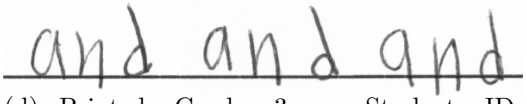


(a) Cursive Grade 3. Student ID: AlK2021WA11-12. Features: 1, 2, 1, 1, 2, 1, 1, 0, 2, 0, 0, 0

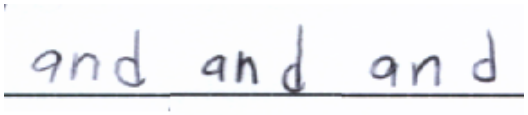


(b) Cursive Grade 4. Student ID: EmR2020Ll11-12. Features: 2, 1, 1, 1, 2, 1, 0, 0, 2, 2, 2, 0

(c) Printed Grade 2. Student ID: AlA2022RO11-12. Features: 0, 0, 0, 1, 1, 0, 3, 0, 2, 0, 0, 0



(d) Printed Grade 3. Student ID: EmL2021Za11-12. Features: 0, 0, 0, 1, 0, 0, 1, 0, 2, 0, 0, 0



(e) Printed Grade 4. Student ID: AnL2020Fe11-12. Features: 0, 1, 0, 1, 1, 0, 1, 0, 2, 0, 0, 0

Figure 3: Examples of Characteristic Values of Handwritten *and*