University at Buffalo The State University of New York | REACHING OTHERS

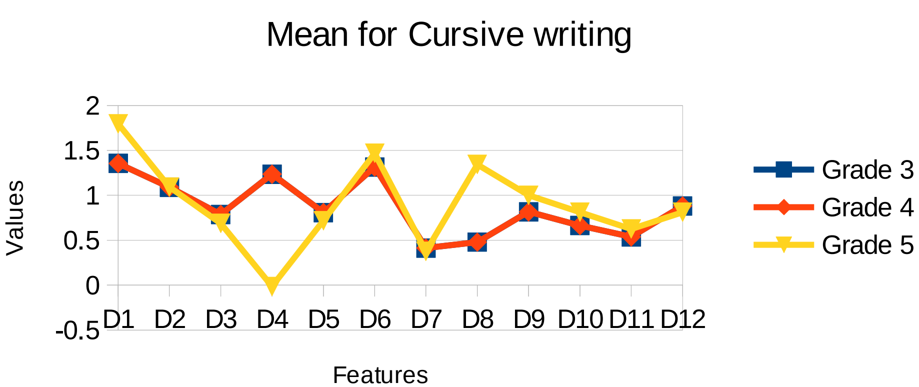
Probabilistic Graphical Models Bayesian Network and Markov Network construction for Children's Handwriting analysis (Implemented in R)

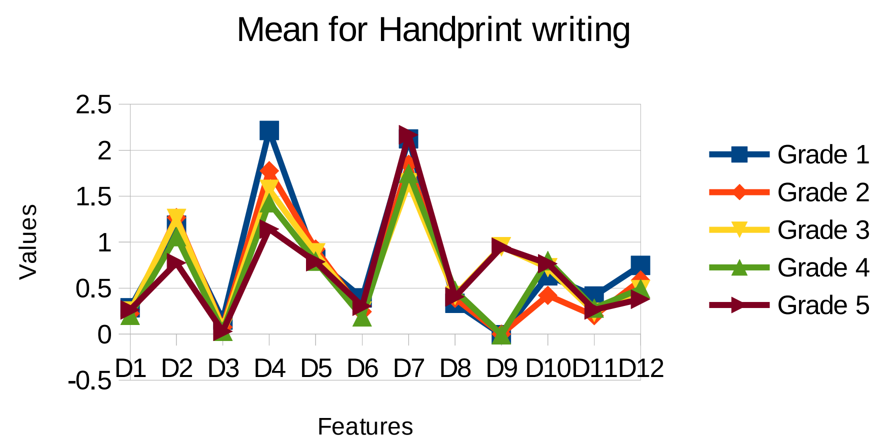
Team:
Karthik Kiran(kkiran@buffalo.edu) UB# 50133145
Harshith Kumar Ramadev(harshith@buffalo.edu) UB# 50134007

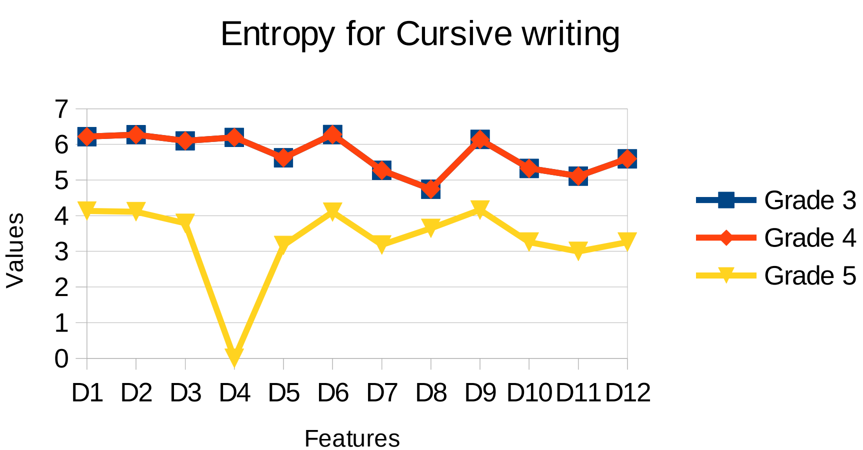
Under the Guidance of:
Prof.Sargur N. Srihari

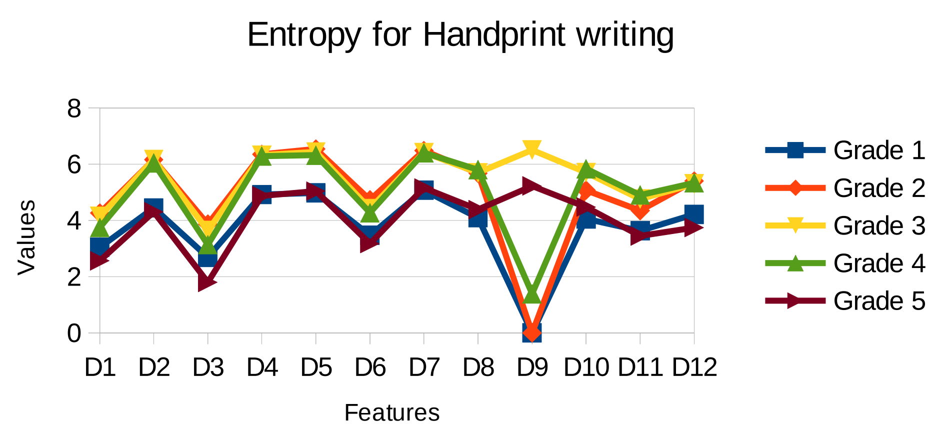
Dataset Cleaning and Inference setup:

- The invalid entries such as “99” and “-1” are converted to legitimate entries spanning in the range from “0” to “5” in all the 12 features
- Compute Mean, Entropy and Relative entropy of cleaned dataset, which are used for inference purposes from the constructed graphical models
- From the observation of mean samples of cursive and hand-printing data, **we can infer that there is a slight variation in their handwriting styles over the years. In other words, consistent writing styles are displayed.**









Bayesian Network Construction:

- Apply **Pearson's Chi-square test** on the features (pairwise) to obtain a dependency matrix between the features
- **Sort the matrix in descending order and pick the top 10 pairs**, indicating pairs which have the highest dependency between them
- **Log-loss approach** on the identified pairs to determine the directionality of dependency(parent and child relationship) which is shown:

  a) Find the **frequency of each combination of values that a pair of features** takes in the cleaned dataset. To find the frequency of a pair, count the number of times each combination of values are present in dataset and divide this figure with total sum of count of all possible combinations
  Eg: If features (D3,D4) are selected from Chi-square resultant matrix, find the frequency of (D3,D4)={(0,0),(0,1),(0,2),(0,3),(0,4),(1,0),....,(5,5)}, select only those value combinations that are present in cleaned dataset for this pair.

  b) Find the **frequency of the second feature** in the identified pair taking the value as taken by it in the pair.
  Eg: If (D3,D4) pair takes a value (0,1), then calculate the frequency for D4 taking the value 1.

  c)**Divide frequency** obtained in a) and b) to obtain the **Conditional probability distribution (CPD**) for the identified feature pair

  d) Repeat steps a), b) and c) until it covers all possible value combinations for a identified feature pair.

  e) Calculate the log-loss for the pair by taking the **sum of logarithm to base 10** outcome of all the different values obtained in c) = **Log-Loss1**

  f) **Swap the features in the identified pair**, (Eg: say (D3,D4) becomes (D4,D3) and second feature becomes D3 instead of D4) and repeat the procedure from a) to e) = **Log-Loss2**

  g**) If Log-Loss1<Log-Loss2, then feature 2 is the child of feature 1(Eg: D3->D4) Else, feature 1 is the child of feature 2 (Eg:D4->D3).**

  Assumption: Any node in the Bayesian network will have a maximum of two children.

- **Cursive data Chi-square test output, Log-Loss1, Log-Loss2 and maximum 10 pairs in Chi-square are shown in next slide respectively: (Clockwise)**

  Note that, the Chi-Square output is 12*12 matrix , where 144 possibilities of feature pairs are considered. Secondly, the Log-loss tables contain top 10 feature pairs which have the maximum dependency in the Chi-square output.

	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12
1	0.000000e+00	8.420599e-02	1.314271e-02	7.828903e-02	6.302874e-07	5.252668e-01	2.902278e-03	5.064977e-13	1.997354e-16	5.514208e-03	4.732948e-01	8.556749e-01
2	8.420599e-02	0.000000e+00	5.560520e-01	4.625042e-01	1.199527e-01	6.974525e-05	1.043933e-05	1.817290e-02	5.035703e-03	5.063550e-02	5.666434e-02	5.760110e-01
3	1.314271e-02	5.560520e-01	1.977752e-246	9.605542e-06	2.021123e-23	3.414269e-01	3.509873e-59	4.601840e-52	4.730208e-31	2.693317e-02	3.238531e-01	6.936474e-02
4	7.828903e-02	4.625042e-01	9.605542e-06	0.000000e+00	6.428209e-02	9.155768e-01	6.673982e-01	4.374931e-02	1.396587e-01	7.185698e-01	3.887232e-01	7.752951e-01
5	6.302874e-07	1.199527e-01	2.021123e-23	6.428209e-02	0.000000e+00	1.408937e-03	7.517706e-11	1.621054e-32	1.132474e-35	1.191076e-01	2.581297e-01	7.750194e-03
6	5.252668e-01	6.974525e-05	3.414269e-01	9.155768e-01	1.408937e-03	0.000000e+00	7.309334e-02	5.311160e-06	4.069559e-01	1.571098e-01	4.229680e-03	2.842781e-03
7	2.902278e-03	1.043933e-05	3.509873e-59	6.673982e-01	7.517706e-11	7.309334e-02	0.000000e+00	2.694926e-43	1.046420e-14	3.707025e-02	4.479505e-01	1.076157e-02
8	5.064977e-13	1.817290e-02	4.601840e-52	4.374931e-02	1.621054e-32	5.311160e-06	2.694926e-43	0.000000e+00	1.455024e-41	1.035224e-03	2.321880e-01	6.185613e-02
9	1.997354e-16	5.035703e-03	4.730208e-31	1.396587e-01	1.132474e-35	4.069559e-01	1.046420e-14	1.455024e-41	4.376448e-246	1.800788e-01	4.935696e-01	1.619884e-01
10	5.514208e-03	5.063550e-02	2.693317e-02	7.185698e-01	1.191076e-01	1.571098e-01	3.707025e-02	1.035224e-03	1.800788e-01	0.000000e+00	4.984147e-172	3.161566e-124
11	4.732948e-01	5.666434e-02	3.238531e-01	3.887232e-01	2.581297e-01	4.229680e-03	4.479505e-01	2.321880e-01	4.935696e-01	4.984147e-172	0.000000e+00	5.174183e-147
12	8.556749e-01	5.760110e-01	6.936474e-02	7.752951e-01	7.750194e-03	2.842781e-03	1.076157e-02	6.185613e-02	1.619884e-01	3.161566e-124	5.174183e-147	0.000000e+00

	row.names	nodepairs	logloss1
1	Node1	2	-4.1235665
2	Node2	3	-4.1235665
3	Node1	1	-6.2256328
4	Node2	6	-6.2256328
5	Node1	4	-3.6055970
6	Node2	6	-3.6055970
7	Node1	4	-1.6055825
8	Node2	7	-1.6055825
9	Node1	4	-4.0539744
10	Node2	10	-4.0539744
11	Node1	1	-2.7477833
12	Node2	11	-2.7477833
13	Node1	9	-0.9323562
14	Node2	11	-0.9323562
15	Node1	1	-3.4075541
16	Node2	12	-3.4075541
17	Node1	2	-4.5378209
18	Node2	12	-4.5378209
19	Node1	4	-4.3173071
20	Node2	12	-4.3173071

	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12
1	0	0.08420599	0.01314271	7.828903e-02	6.302874e-07	5.252668e-01	2.902278e-03	5.064977e-13	1.997354e-16	0.005514208	4.732948e-01	8.556749e-01
2	0	0.00000000	0.55605201	4.625042e-01	1.199527e-01	6.974525e-05	1.043933e-05	1.817290e-02	5.035703e-03	0.050635499	5.666434e-02	5.760110e-01
3	0	0.00000000	0.0000000	9.605542e-06	2.021123e-23	3.414269e-01	3.509873e-59	4.601840e-52	4.730208e-31	0.026933171	3.238531e-01	6.936474e-02
4	0	0.00000000	0.0000000	0.000000e+00	6.428209e-02	9.155768e-01	6.673982e-01	4.374931e-02	1.396587e-01	0.718569768	3.887232e-01	7.752951e-01
5	0	0.00000000	0.0000000	0.000000e+00	0.000000e+00	1.408937e-03	7.517706e-11	1.621054e-32	1.132474e-35	0.119107611	2.581297e-01	7.750194e-03
6	0	0.00000000	0.0000000	0.000000e+00	0.000000e+00	0.000000e+00	7.309334e-02	5.311160e-06	4.069559e-01	0.157109844	4.229680e-03	2.842781e-03
7	0	0.00000000	0.0000000	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	2.694926e-43	1.046420e-14	0.037070247	4.479505e-01	1.076157e-02
8	0	0.00000000	0.0000000	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	1.455024e-41	0.001035224	2.321880e-01	6.185613e-02
9	0	0.00000000	0.0000000	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.180078780	4.935696e-01	1.619884e-01
10	0	0.00000000	0.0000000	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000000	4.984147e-172	3.161566e-124
11	0	0.00000000	0.0000000	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000000	0.000000e+00	5.174183e-147
12	0	0.00000000	0.0000000	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000000	0.000000e+00	0.000000e+00

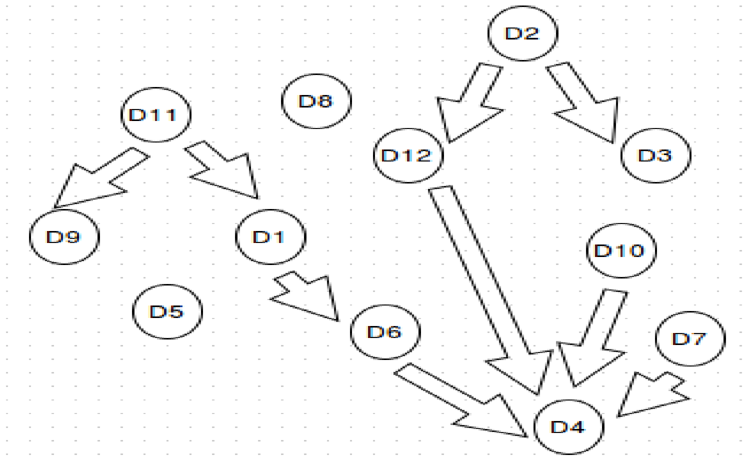
	row.names	nodepairs	logloss2
1	Nodee1	3	-2.335282
2	Nodee2	2	-2.335282
3	Nodee1	6	-5.476111
4	Nodee2	1	-5.476111
5	Nodee1	6	-3.959986
6	Nodee2	4	-3.959986
7	Nodee1	7	-7.961990
8	Nodee2	4	-7.961990
9	Nodee1	10	-7.250261
10	Nodee2	4	-7.250261
11	Nodee1	11	-6.434708
12	Nodee2	1	-6.434708
13	Nodee1	11	-3.999560
14	Nodee2	9	-3.999560
15	Nodee1	12	-6.106383
16	Nodee2	1	-6.106383
17	Nodee1	12	-4.425804
18	Nodee2	2	-4.425804
19	Nodee1	12	-7.301414
20	Nodee2	4	-7.301414

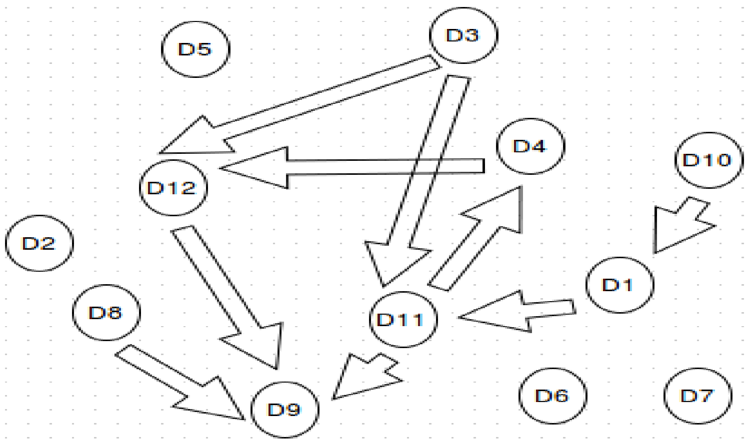
- **Cursive writing Bayesian network (Grade 3+Grade 4+Grade 5) Inference: (Left figure)**

We can observe the direct dependency between feature 12(n-d relationship) and feature 4(Shape of “n” arches) and another direct dependency between feature 12 and feature 1(Initial stroke of “a”). Also, there exists a dependency between feature 1 to feature 6(Formation of “d” staff) and from feature 6 to feature 4. Hence, the relationships between these 4 features constitute that important specifications of each “a”, “n” and “d” letters are important to analyze the whole word “and”. On a contrary, less significant features like feature 5(Location of “n” mid) and feature 8(Formation of “d” terminal) are independent.

- **Hand-print writing Bayesian network (Grade 1+ Grade 2+ Grade 3+Grade 4+Grade 5) Inference: (Right figure)**

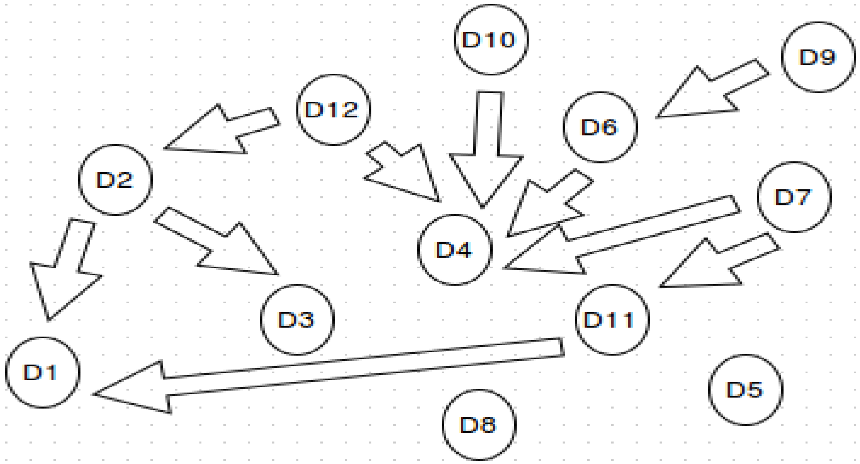
It is observant from the figure that the feature 11(a-d relationship) is dependent on feature 4(formation of “n” staff) which is in turn dependent on feature 12(n-d relationship). Hence, a-d relationship is determined through n-d relationship



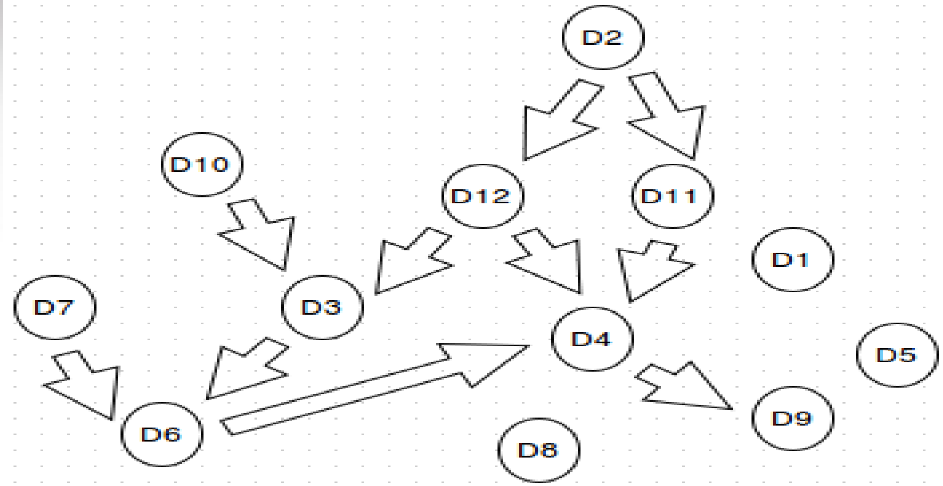


- **Dynamic Bayesian Network:**

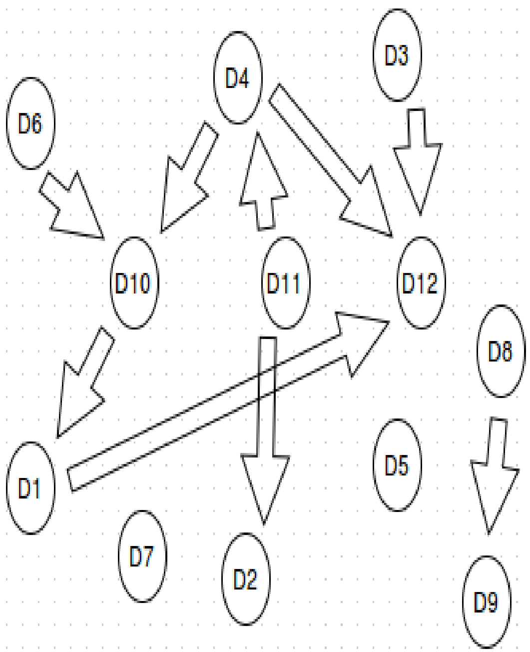
As the dataset is progressive over the years, we can construct the Bayesian networks for Grade 3, Grade 4 with respect to Cursive writing and for Grade 3, Grade 4 and Grade 5 with respect to Hand-print writing. The transition in dependency between the features can be noted from these figures which are shown in the following slide:



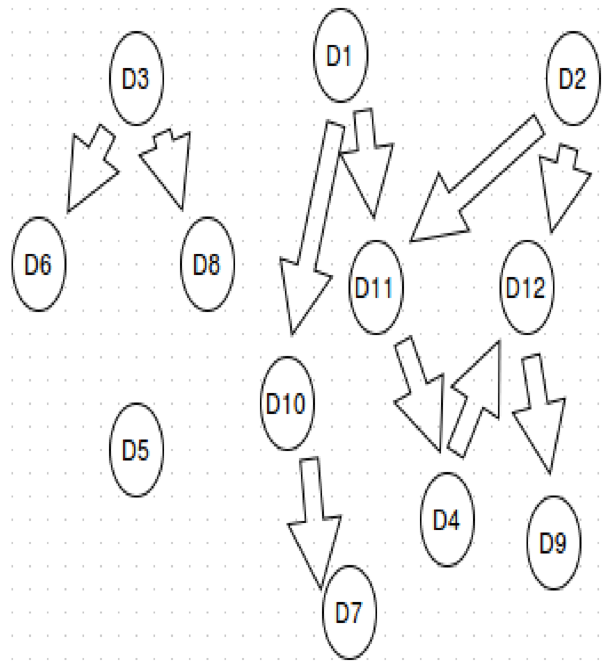
Cursive Grade 3



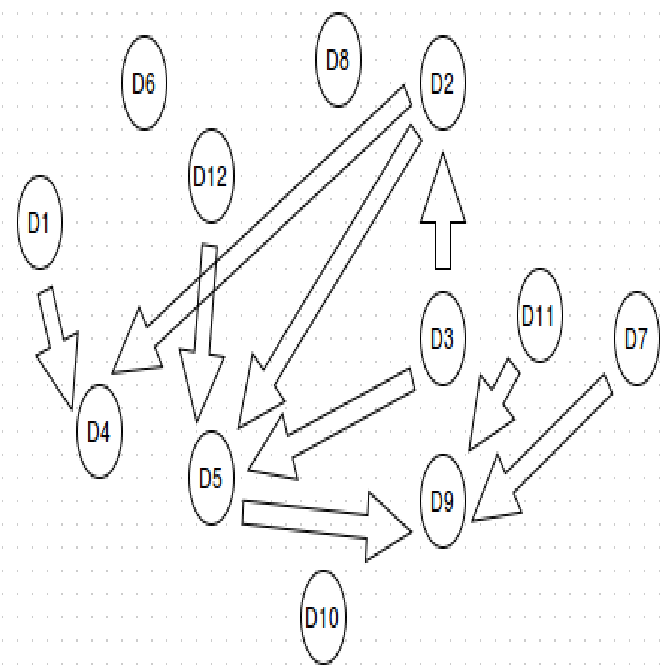
Cursive Grade 4



Handprint Grade 3



Handprint Grade 4



Handprint Grade 5

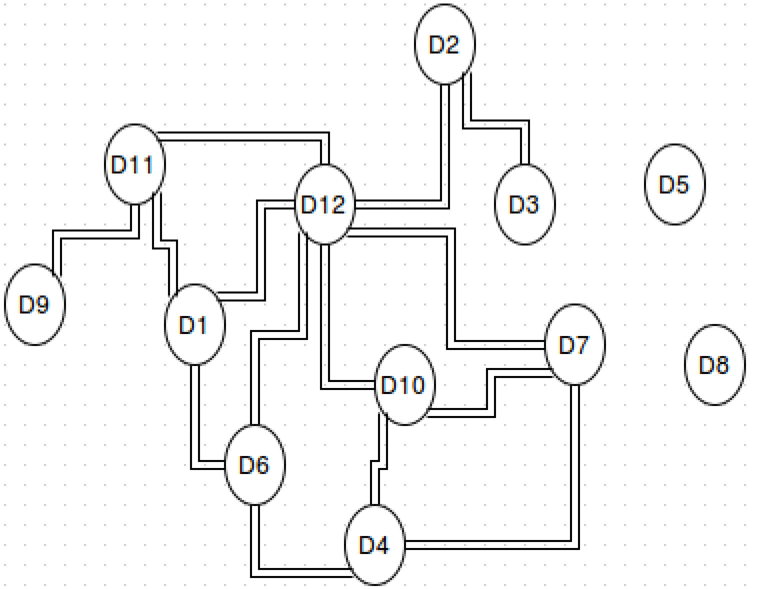
Markov Network construction:

Moralization: The existent Bayesian network is transformed into a Markov network except for the fact that directed edges are converted to non-directed edges and an extra edge is added between those pair of nodes which are having a common child.

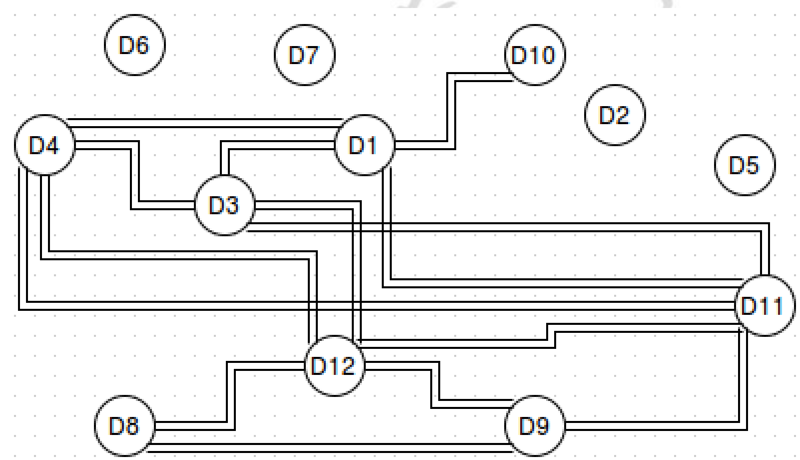
KL Divergence threshold approach: Here, we take the inverse of the Co-variance matrix to get the matrix containing values that indicate the divergence between the features. Now we set a threshold value and then consider all the feature pairs which are having a value greater than this threshold for Markov network construction

Initially, we assume all the variables to be independent, i.e. there are no edges between any nodes. Edges are added to the graph based on the pairwise chi-square values on the condition that the resulting hyper-graph is a hyper-tree At each stage the entropy is computed and the graph corresponding to the minimum entropy is selected for adding edges in the next step. This process is repeated till a threshold is reached with respect to the decrease in entropy between two successive stages. Setting the threshold is important and generally set according to the size of the dataset. Smaller the dataset larger the threshold, this is to handle the erroneous samples in the small dataset, so using a higher threshold can suppress false dependencies.

The below graphs demonstrate the Markov networks constructed by applying the Moralization technique to Cursive Bayesian and Hand-print Bayesian networks:



Cursive Markov network



Handprint Markov network