CS 6362 Machine Learning, Fall 2017: Homework 5

Venkataramana Nagarajan

Question 1:

(a) **Directed Graph**
Ans : YES
Here X5 is given, so X8 and X9 are d-separated. Further X1 and X8 are d-separated because nothing in between them are given.

Since X16 is not given X9 and X12 are d-separated.

Also X1 and X12 are d-separated as nothing between them is given.

Hence, X1 and X9 are d-separated as well.

Undirected Graph
Ans : YES
To get to X9 , there are only two ways , either X5 or X14. Since both are known X1 and X9 are d-separated

(b) **Directed Graph**
Ans : NO

X13 is dependent on X12 because X8 is not given.

X12 is dependent on X11 because X15 is given.

Hence X11 and X13 are not d-separated.

Undirected Graph
Ans : YES

Since X15 is given which lies in the path between X11 and X13, they are d-separated.

(c) **Directed Graph**
Ans : YES
Here X4 and X12 are d-separated as none of the variables in between are given. Hence X4 and X5 are are also d-separated.

Undirected Graph
Ans : NO

There exists a path between X4 and X5 such that none of the variables between them are given. Eg. $X4->X6->X11->X15->X12->X8->X5$.

Hence X4 and X5 are not d-separated.

(d) **Directed Graph**
Ans : NO

Here since X15 is given, set A={x3,x4} and X12 are not d-separated.

Since X5 and X8 are not given, B={X13,X9} is not d-separated with X12. Hence A={x3,x4} and B={X13,X9} are not d-separated.

Undirected Graph
Ans : YES

Here X15 is the only link between A={x3,x4} and B={X13,X9} and X15 is given which makes A and B d-separated.

Question 2: If X5 is known X2 becomes d-separated from the graph. Hence the minimal subset required is A={X5}

Question 3: The broad approach that I will follow is as follows:
1. Train the classifier using labeled data.
2. Assign probabilistic labels to unlabeled examples.
3. Update the parameters.
4. Go back to (2) until convergence.

Step 1: Find probability of each label k = 1,2,...K

$$P(y=k) = \gamma_k = \frac{\sum_{i=1}^{n} y_i = k}{n}$$

Step 2: Calculate probability for each feature X_{ip} w.r.t to each label k , where $X_i = \{X_{i1}, ..X_{ip}.., X_{iT}\}$ and T is the number of features in each feature vector.

$$P(X_{pk}|y=k) = \theta_{kp} = \frac{\sum_i X_{ip}.y_{ik}}{\sum_i y_{ik}}$$

EM algorithm

E-Step: Calculate probability of each feature vector X_i in unlabeled dataset for each label k using distribution of labels from the previous M-step. For the 1st iteration we will use label distribution obtained from supervised learning):

$$log(p\{x_i\}_{i=1}^{m}|\gamma_k) = \sum_{i=1}^{m}\sum_{p=1}^{T} log \sum_{i=1}^{k} P(X_{wp}, y_k|\gamma_k)$$

M-Step: Maximizing probability of each label distribution. Let δ be balancing factor. The logic is to give less weight to unlabeled data.

$$\gamma_{k+1} = argmax_{\gamma_k} logp(\{x_i, y_k\}_{i=1}^{n}|\gamma_k) + \delta log(p\{x_i\}_{i=1}^{m}|\gamma_k)$$

Note : I have used a paper on Semi-Supervised Learning as reference to derive the above equations.