TDA 231 Machine Learning 2012: Final Exam

Due: 10 AM, March 5, 2012

General Instructions

1. All datasets can be downloaded from http://www.cse.chalmers.se/edu/year/2012/course/TDA231_Machine_Learning/final/
2. All code and matlab files must be submitted in a single zip file *code.zip*. This has to be submitted through FIRE.
3. All figures and tables must be included in the answersheet. The answersheet can be submitted either through FIRE using file name *solution.pdf* or handed in between $9-10$ am on Monday at EDIT6453.
4. If you have any doubts, please email to the mailing group. Doubts and queries will be answered at Saturday $4-5$ pm and Sunday $4-5$ pm to the mailing group by email.

Questions

1. (10 points) Consider dataset `f1.mat` having $D_1 = \{(x_i, y_i)|i = 1,\dots,n\}$ where each $x_i \in \mathbb{R}^2, y_i \in \{1,-1\}$. Assume that the class conditional densities are spherical gaussian with means μ_1 and μ_2. Design a classifier of the form
$$y = 1 \text{ if } t_1\|x-\mu_1\| \leq t_2\|x-\mu_2\|$$
$$y = 2 \text{ otherwise}$$
Find t_1, t_2 such that the given Classifier is a Bayes Classifier. Give reasons for your answer. Submit your prediction Y_pred on the test set X_test in the form of a file `f1pred.mat`.

2. (10 points) Consider dataset `f2.mat` consisting of observations of the form $X = [x_1, x_2, \dots, x_d]$ where $x_i \in \{1,-1\}$; and labels $Y_i \in \{+1,-1\}$. Design a Naive Bayes Classifier for this kind of data when number of classes is 2.
   (a) (2 points) State the parameters of the Naive Bayes Classifier.
   (b) (5 points) Briefly describe and implement a method to estimate the parameters. Discuss any problems you might face with the estimates and what steps you will take to correct them.
   (c) (2 points) Show that the Naive Bayes Classifier can be written as a linear classifier.
   (d) (1 point) Submit your prediction Y_pred on the test set X_test in file `f2pred.mat`.

3. (10 points) Given a dataset $D = \{(x_i, y_i)|i = 1,\dots,n\}$ where each $x_i \in \mathbb{R}^d$ $y_i \in \{1,-1\}$ consider the SVM formulation.
$$min_{w,b,\xi} \ \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i^2$$
$$s.t. y_i(w^\top x_i + b) \geq 1 - \xi_i$$
$$\xi_i \geq 0 \ i = 1,\dots,n$$
   (a) (2 points) State the Lagrangian of the problem.

(b) (4 points) State the KKT conditions.

(c) (4 points) Derive the dual.

4. (10 points) Consider the mixture distribution

$$p(x|\theta) = \sum_{i=1}^{k} \pi_i f_i(x|\theta)$$

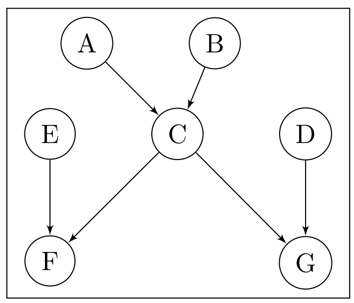
where $f_i(x|\theta) = e^{-\mu_i} \frac{\mu_i^x}{x!} \quad x = 0, 1, \ldots$

(a) (5 points) Derive the EM updates for this problem?

(b) (3 points) The dataset `f4.mat` has been generated according to the mixture distribution described above. The value of k is not known but it is assumed to be an integer between $\{1, \ldots, 5\}$. Implement the EM algorithm on the given dataset using your update equations. Based on your results can you say what is the value of k. Give resasons.

(c) (2 points) For your choice of k plot the likelihood of data after every EM iteration. Submit your plot. Also report the estimated values of μ_i and π_i.

5. (6 points) Consider the Bayesian Network shown below where all variables are binary e.g. $A \in \{0, 1\}$.



(a) (2 points) Write the probability distribution corresponding to above Bayesian Network.

(b) (4 points) Write whether the following conditional independence statements are true or false, and why.

1. $A \perp\!\!\!\perp C | B, F, G$
2. $A \perp\!\!\!\perp E | F$
3. $E \perp\!\!\!\perp D | F, G$
4. $F \perp\!\!\!\perp G | C, D$

6. (14 points) Consider the probability distribution

$$P(x_1, x_2, x_3, x_4, x_5) = \frac{1}{Z} \phi_a(x_1, x_2) \phi_b(x_4, x_5) \phi_c(x_3, x_1, x_4) \phi_d(x_5)$$

where $x_i \in \{0, 1\} \forall i$ and Z is the normalization constant.

(a) (1 point) Write the expression for the normalization constant Z.

(b) (2 points) Draw the graphical model for the probability distribution.

(c) (3 points) Write the sum-product update equations for the following messages

$$m_{\phi_c \to x_1}(0), \qquad m_{x_4 \to \phi_b}(1)$$

as well as the belief update $b_4(0)$ $(= P(x_4 = 0))$.

(d) (4 points) Consider dataset `f5.mat` having multiple observations $X = [x_1, x_2, x_3, x_4, x_5]$. Write the formula for mutual information between variables x_i and x_j. Compute and report in a table, the mutual information between x_i, x_j for all $1 \leq i, j \leq 5$ based on given dataset.

(e) (4 points) Construct the Chow-Liu tree based on given dataset. Write the corresponding probability factorization.