Problem 1

(15 pt)

Answer the following questions, and provide an **explanation** for each question.

(a)

(5 pt) Can linear regression work when all X values are the same? When all Y values are the same?

(b)

(5 pt) Can linear regression be used when the X values are actually categories?

(c)

(5 pt) Will the regression line be the same if you exchange X and Y?

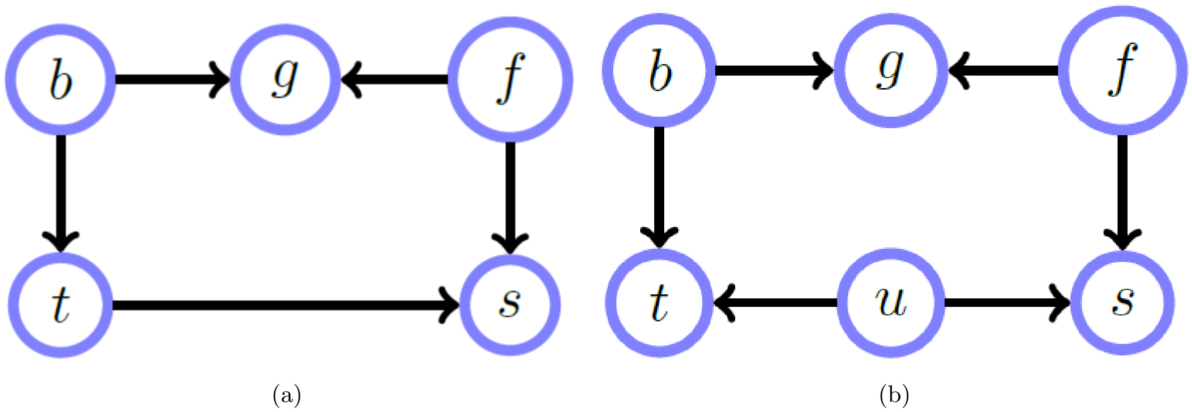

Figure 1: Some DGMs.

Problem 2

(15 pt)

Answer the following questions, and provide an **explanation** for each question.

(a)

(5 pt) Consider the graph in Figure 1(a). Are the variables t and f unconditionally independent, i.e. $t \perp f$?

(b)

(5 pt) Consider the graph in Figure 1(a). Are the variables t and f independent conditioned on g, i.e. $t \perp f | g$?

(c)

(5 pt) Consider the graph in Figure 1(b). Are the variables $\{b, f\}$ and u unconditionally independent, i.e. $\{b, f\} \perp u$?

Problem 3

(20 pt)

In the following Figure 2, **draw** the first principal component direction in the left figure, and the first Fisher's linear discriminant direction in the right figure. And **explain why**. Note: for PCA, ignore the fact that points are labeled (as round, diamond or square) since PCA does not use label information. For linear discriminant, consider both diamond and square points as the positive class, and round points as the negative class.

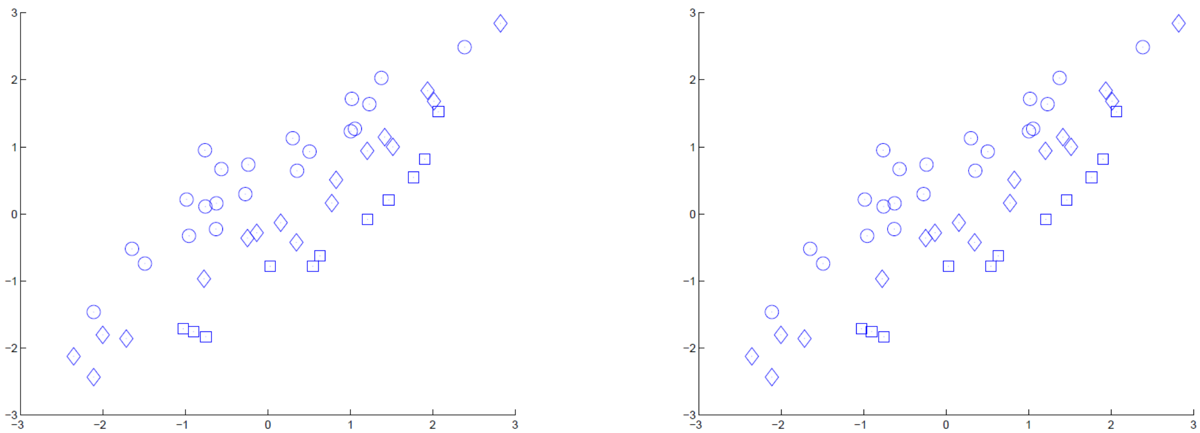

Figure 2: Draw the first principal component and linear discriminant component, respectively

Problem 4
(30 pt)

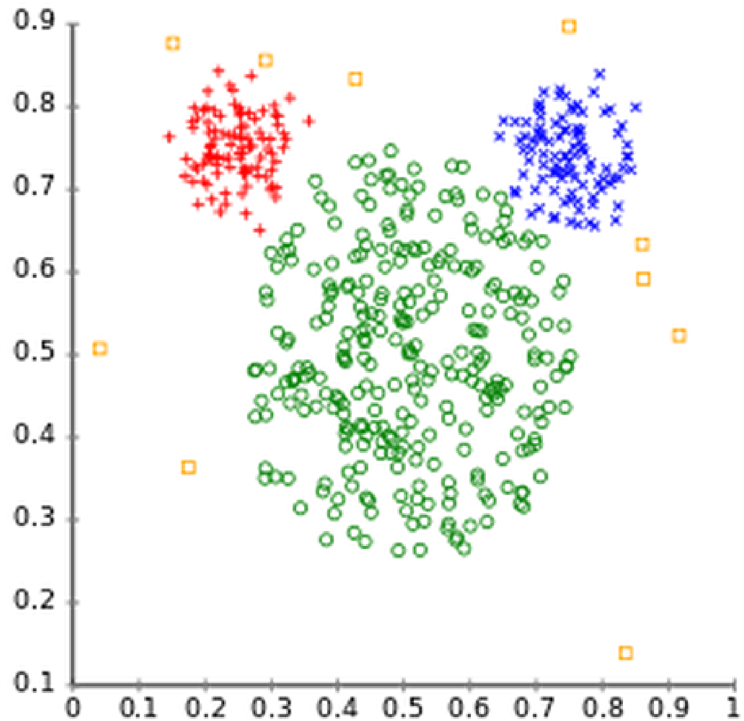

Figure 3: Randomly drawn data points

Figure 3 shows randomly drawn data points from three latent classes $(+, x, o)$. (Ignore the square points in the figure)
Assume we cluster this dataset by using k-means clustering with k of three and euclidean distance. **Draw** expected results with 3-means roughly over the figure, and **describe** the limitations of k-means for this dataset. **Describe** how GMM-EM methods can resolve this problem for this dataset.

Problem 5

(20 pt)

This problem is about the EM algorithm for mixtures of Bernoullis.
The expectation of the complete-data log likelihood with respect to the posterior distribution is given by

$$Q(\theta, \theta^t) = \sum_{i=1}^{N} \sum_{k=1}^{K} r_{ik} \left[\log \pi_k + \sum_{j=1}^{D} x_{ij} \log \mu_{kj} + (1 - x_{ij}) \log(1 - \mu_{kj}) \right]$$

The pdf of the beta distribution $Beta(\alpha, \beta)$ is given by $f(x; \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$.

(a)

(10 pt) Derive the M step for ML estimation of a mixture of Bernoullis.

$$\mu_{kj} = ?$$

(b)

(10 pt) Derive the M step for MAP estimation of a mixture of Bernoullis with a $Beta(\alpha, \beta)$ prior.

$$\mu_{kj} = ?$$