Quiz # 5 **Name:** ______________________

Mar 26, 2020 **Student ID:** ______________________

CMPT 727: 2020 Spring

Instructor: Maxwell Libbrecht

Instructions:

Take home exam. Electronic media limited to Wikipedia, published textbooks and calculator app. Any physical media is allowed.

Please answer the questions below. Show all your work.

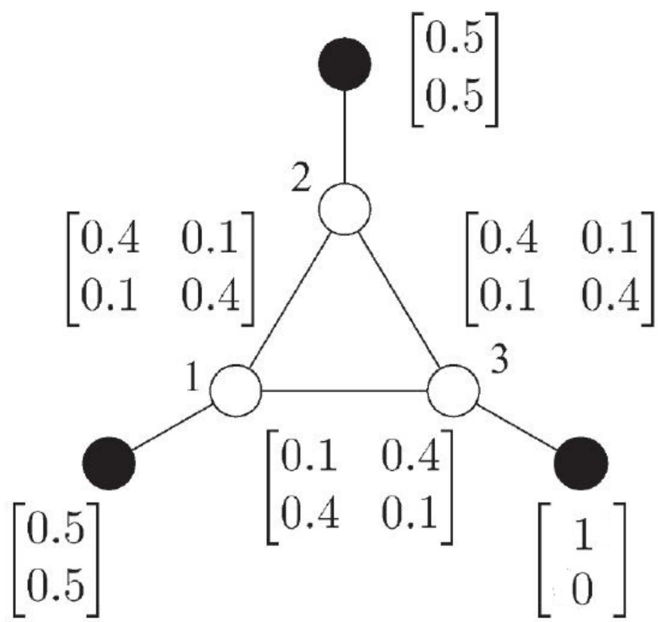
Problem 1. (6 pts) KL

Suppose we obverse 100 cars in a HOV lane and the number of passengers (exclude drivers) in a car ($p(X)$) are in the following distribution:

	0	1	2	3+
$p(x)$	0.15	0.2	0.45	0.2
$q_u(x)$	0.25	0.25	0.25	0.25
$q_b(x)$	0.08	0.32	0.42	0.18

We propose two possible models for this distribution: a uniform distribution $q_u(x)$ and a binomial distribution $q_b(x) = \text{Binom}(3, 0.57)$. Determine which model is more suitable according to the reverse KL divergence.

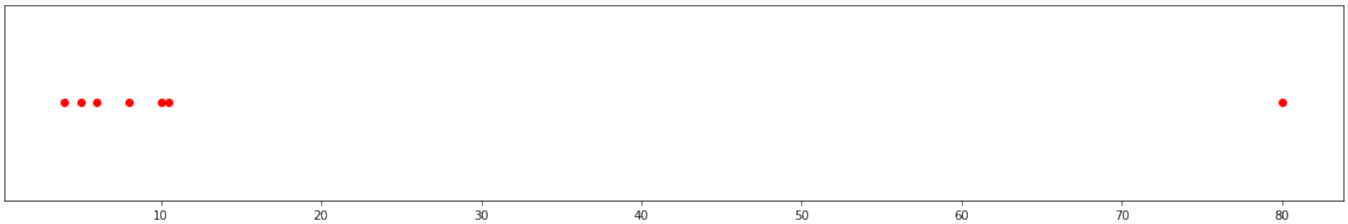
Problem 2. (8 pts) Loopy BP



Given the above pairwise UGM on binary nodes, calculate the posterior belief on Node 1 after one iteration of loopy belief propagation. You may leave your answer in fractional form.

Problem 3. (8 pts)

Consider the following 7 points in 1D space.



The data points are:$[4, 5, 6, 8, 10, 10.5, 80]$. We want to cluster into two clusters.

(a) Recall that k-medoids clustering is a method in which we choose k data points as cluster centers $c_j \in \{1 \dots N\}$, using the objective function

$$L(c_{1\dots k}) = \sum_{i=1}^{N} \text{min}_j \; d(x_i, x_{c_j}).$$

Find the cluster centers with the optimum loss when d is squared (Euclidean) distance.

(b) One way to do robust clustering (that is, to make it less sensitive to outliers) is to change the distance function. Consider the distance function $d(x_1, x_2) = \text{min}((x_1 - x_2)^2, \alpha)$, with $\alpha = 10$. Find the cluster centers with the optimum loss under this distance function.

Problem 4. (6 pts)

Five students ($S_1, ..., S_5$) have participated in six projects. The projects involve the following students respectively: $\{(S_1, S_4), (S_2, S_3, S_5), (S_4), (S_2, S_3), (S_1), (S_5, S_4)\}$. The grade on each project was either A or C. Each student is either studious or lazy. A project is likely to get a good grade if either student is studious. Draw a Bayes net, MRF and factor graph representing this scenario.