CSCI 5521: Intro to Machine Learning)*
Sample Midterm Exam

Name: **Student ID:**

Question 1 (30 points)
Consider the class of K-interval classifier I_K in $\mathbb{R}$ which is specified by K intervals $[a_1, b_1]$, $[a_2, b_2]$,...,$[a_K, b_K]$ and labels any example positive iff it lies inside any of the K intervals.

1. What is the VC dimension of I_1 denoted by VC(I_1)? Prove your answer. You need to show why the classifiers can shatter VC(I_1) data points but not VC(I_1)+1 data points.

*Instructor: Rui Kuang (kuang@umn.edu).

2. What is the VC dimension of I_2 denoted by VC(I_2)? Prove your answer. You need to show why the classifiers can shatter VC(I_2) data points but not VC(I_2)+1 data points.

Question 2 (30 points)

The *Poisson distribution* has been widely used to model count data, i.e. the observation $x \in \{0, 1, 2...\}$ is a count. The Poisson distribution has the following probability density function

$$f(x|\lambda) = \frac{\lambda^x \exp\Big(-\lambda\Big)}{x!}, \lambda > 0,$$

where the parameter λ is referred to as event rate. Suppose $x^1, x^2, \ldots, x^N$ are N independent and identically distributed (i.i.d) samples of a Poisson distribution, find the maximum likelihood estimation of the parameter λ.

1. Write the log-likelihood function $\mathcal{L}(\lambda|x^1, x^2, \ldots, x^N)$.

2. Find the maximum likelihood estimation of λ by setting the derivative of $\mathcal{L}$ with respect to λ to zero.

Question 2 (40 points)

Suppose $\mathbf{x}^t \backsim \mathcal{N}(\mu, \Sigma), t = 1, \ldots, N$ are i.i.d samples of a 6-D multivariate Gaussian distribution. Assume the covariance matrix of the Gaussian distribution to be a block diagonal matrix in the following form

$$\Sigma = \begin{pmatrix} \hat{\Sigma} & 0 & 0 \\ 0 & \hat{\Sigma} & 0 \\ 0 & 0 & \hat{\Sigma} \end{pmatrix},$$

where $\hat{\Sigma}$ is a 2-by-2 sub-covariance matrix.

1. Observe the structure of the covariance matrix Σ and describe what assumption is made about the data. Based on the observation, write the maximum likelihood function and derive the estimation of the parameters μ and $\hat{\Sigma}$. (Note: it is incorrect to estimate a full Σ and take the diagonal blocks as the solution.)

2. Compare and discuss the model complexity of the following three assumed covariance matrices Σ.

   - **Model 1**: Σ is a full matrix.
   - **Model 2**: Σ is a diagonal matrix.
   - **Model 3**: Σ is a block diagonal matrix as in this question.