Number:

Signature:

Bishop's University

CS 509 - Pattern Recognition

Final exam

Fall 2018

The exam is worth 100 points total and has five problems. Be sure to read the whole exam before attempting any of it. The exam is open notes. For that purpose, only handwritten notes, and class notes are allowed. You have 180 minutes to complete the exam. Use the provided white space to respond to each question. Please, write legibly.

Problem 1: "Recall" Questions (10 points)

1. Describe Bayes Decision Rule. **(2 points)**

2. What quantity is PCA maximizing during dimension reduction? **(2 points)**

3. What does ISOMAP focus on preserving during dimension reduction? **(2 points)**

Problem 2: Non-Parametric Methods (20 points)

You are given a dataset $D = \{0,1,1,2,2,2,3,4,4\}$. Using techniques from non-parametric density estimation, answer the following questions:

1. Draw a histogram of D with a bin-width of 1 and bins centered at $\{0,1,2,3,4\}$. (**6 points**)

2. Write the formula for the kernel density estimate given an arbitrary kernel K. (**2 points**)

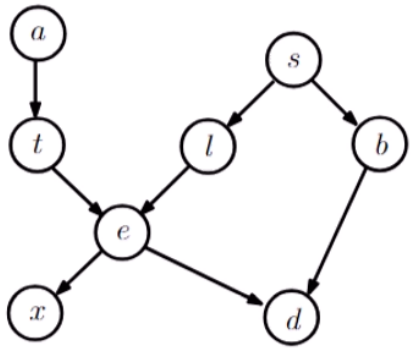
3. Select a triangle kernel as your window function:

$$K(u) = (1 - |u|)\delta(|u| \leq 1)$$

Where u is a function of the distance of sample x_i to the value in question x divided by the bandwidth: $u = \frac{x - x_i}{h}$. Compute the kernel density estimates for the values $x = \{2,4\}$ with bandwidths of 2. (**6 points**)

4. What are the limitations of neural networks that were overcome by convolutional neural networks (CNN)? Justify your answer. **(4 points)**

Problem 3: Graphical models (20 points)



The chest clinic network above concerns the diagnosis of lung disease (tuberculosis, lung cancer, or both, or neither). In this model, a travel abroad is assumed to increase the probability of lung cancer. We have the following binary variables.

x	positive X-ray
d	Dyspnea (shortness of breath)
e	Either Tuberculosis or Lung Cancer
t	Tuberculosis
l	Lung cancer

b	Bronchitis
a	A travel abroad
s	Smoker

1. Write down the joint probability implied by the graph. (**4 points**)

2. Are the following independence statements implied by the graph?
   a. $Lung\ cancer \perp\!\!\!\perp bronchitis | smoking$ (**8 points**)

2. Are the following independence statements implied by the graph?
   a. $Lung\ cancer \perp\!\!\!\perp bronchitis | smoking$ (**8 points**)

   b. $A\ travel\ abroad \perp\!\!\!\perp smoking | lung\ cancer, shortness\ of\ breath$ (**8 points**)

Problem 4: MLE and MAP (30 points)

In this problem, we will find the maximum likelihood estimator (MLE) and maximum *a posteriori* (MAP) estimator for the mean of a *univariate normal* distribution. For that purpose, Let $X = \{x_1, \dots, x_n\}$ be an identically, independent distributed (iid) sample drawn from a normal distribution with *known* variance σ^2 and unknown μ defined as follows:

$$f(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

1. What is the joint probability distribution $P(X|\theta)$ of the sample? **(3 points)**

2. What is the maximum likelihood estimation (MLE) of the parameter μ? **(5 points)**

3. Let μ has a prior *Gamma* distribution given by

$$p(\mu) = \frac{\mu^{\alpha-1} e^{\frac{-\mu}{\beta}}}{\Gamma(\alpha)\beta^{\alpha}}$$

Where α and β are uknown.

a. Please write down the objective function of maximum a posteriori (MAP) estimation of the parameter μ. (**4 points**)

b. Find the posterior and the Bayes Estimator. **(10 points)**

c. Estimate μ_{MAP} that maximizes the objective function in a) using MAP estimation. **(10 points)**

Problem 5: EM algorithm for a Gaussian mixture (20 points)

Consider a special case of a Gaussian mixture model in which the covariance matrices Σ_k of the components are all constrained to be diagonals, such as $\Sigma_k = \sigma_k^2 I$, where σ_k^2 is the variance and I is a $d \times d$ identity matrix. Derive the EM equations for maximizing the likelihood function under such a model.

The Gaussian distribution is defined for $x \in \mathbb{R}^d$ as follows:

$$p(x|\mu_k, \Sigma_k) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_k|^{\frac{1}{2}}} exp\left[-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k)\right]$$

Hints:

1. For a symmetric matrix $A \in \mathbb{R}^{n\times n}$ and a vector $x \in \mathbb{R}^n$,

$$\frac{\partial}{\partial x}[x^T A x] = 2Ax$$

2. For $\Sigma_k = \sigma_k^2 I$ we have:

$$p(x|\mu_k, \sigma_k^2 I) = \frac{1}{(2\pi)^{\frac{d}{2}}\sigma_k^d} exp\left[-\frac{1}{2}\frac{(x-\mu_k)^T(x-\mu_k)}{\sigma_k^2}\right]$$