# CS4442B & CS9542B: Artificial Intelligence II – Assignment #2

Due: 23:55, March 3rd (Sunday), 2024

## Instructions

Your submission should include: 1) a .pdf file that containing all the answers to the questions. For the questions requiring to plot figures, you should also include all the figures in the .pdf fie. 2) source code (e.g., .py files, python notebook, .m files...)

## 1 Generative Models, Naive Bayes Classifier [20 points]

Based on the Table on page 13 of Lecture 6, compute the following

(a) [5 points] $p(\text{Water} = \text{warm}|\text{Play} = \text{yes})$, $p(\text{Water} = \text{warm}|\text{Play} = \text{no})$

(b) [5 points] $p(\text{Play} = \text{yes}|\text{Water} = \text{warm})$, $p(\text{Play} = \text{no}|\text{Water} = \text{warm})$

(c) [5 points] $p(\text{Play} = \text{yes}|\text{Forecast} = \text{same})$, $p(\text{Play} = \text{yes}|\text{Forecast} = \text{change})$

(d) [5 points] $p(\text{Water} = \text{warm}|\text{Play} = \text{yes})$, $p(\text{Water} = \text{warm}|\text{Play} = \text{no})$ with Laplace smoothing

## 2 Kernels [30 points]

In this problem, we consider constructing new kernels by combining existing kernels. Recall that for some function $k(x, z)$ to be a kernel, we need to be able to write it as a dot product of vectors in some high-dimensional feature space defined by $\phi$:

$$k(x, z) = \phi(x)^\top \phi(z)$$

Mercer's theorem gives a necessary and sufficient condition for a function $k$ to be a kernel function: its corresponding kernel matrix $K$ has to be symmetric and positive semidefinite.

Suppose that $k_1(x, z)$ and $k_2(x, z)$ are two valid kernels. For each of the cases below, **STATE** whether $k$ is also a valid kernel. If it is, **prove** it. If it is not, **give a counterexample**. You can use either Mercer's theorem, or the definition of a kernel as needed to prove it (If you use any properties on page 10 of Lecture 8, we need to prove them first).

(a) [10 points] $k(x, z) = a_1 k_1(x, z) - a_2 k_2(x, z)$, where $a_1, a_2 > 0$ are real numbers

(b) [10 points] If $k(x, z) = e^{\frac{x^\top z}{\sigma^2}}$ is a valid kernel, prove that the Gaussian kernel $k(x, z) = e^{-\frac{||x-z||_2^2}{2\sigma^2}}$ is also a valid kernel.

(c) [10 points] Suppose that you are given instances which are strings of characters. Given a parameter $L$, you are considering the following parameterized kernel $k_L$ for comparing the instances: and two strings $u$ and $v$, you will count the number of subsequences up to length $L$ that the two strings have in common. For example, if $u$ = "car" and $v$ = "ark", and $L = 1$, then $k_L(u, v) = 2$ (since 'a' and 'r' are common to the two strings). For the same strings, if $L = 2$, then $k_L(u, v) = 3$, since 'a', 'r' and "ar" are in common. Let $L = 2$, is $k_L$ a valid kernel function?

# 3 PCA and Eigenface [50 points]

For this exercise, you will use principal component analysis (PCA) to analyze face images in any programming language of your choice (e.g., Python/Matlab/R). The data set `faces.dat`; each row represents an image (400 images), and each column represents a pixel ($64 \times 64 = 4096$ pixels).

(a) [5 points] **Display** the 200th image.

(b) [5 points] **Remove** the mean of the images (4 points), and then **display** the 200th image (1 point).

(c) [10 points] **Perform PCA** on the mean-centered data matrix. You can either implement PCA by yourself using eigenvalue decomposition over the sample covariance matrix, or use a existing machine learning toolbox. (6 points)

**Sort** the eigenvalues in a descending order and **plot** them. **Report** the largest eigenvalue in scientific notation and keep four decimal places. (4 points)

(d) [5 points] You will find the last (i.e., 400th) eigenvalue is 0. **Explain** why.

(e) [5 points] Based on the eigenvalues, **determine** the dimensionality of the data you want to keep (i.e., how many principal components you want to keep), which accounts for most of the variance. (2 points)

**Explain** your reason. (3 points)

(f) [10 points] **Display** the top-5 leading eigenvectors (corresponding to the top-5 largest eigenvalues) in 5 figures.

(g) [10 points] **Display, respectively, the reconstructed/recovered** 200**th images** using 10, 100, 200, and 399 principal components. (Hint: Lecture 9 (page 19), we have learned that $\hat{x} = vv^\top x$ if we project $x$ into 1-dimensional space using the 1st principal component. Reconstructed $\hat{x}$ using top-$K$ principal components is a straightforward extension: $\hat{x} = \sum_{k=1}^{K} v_k v_k^\top x$)

Hint: in order to display the images and principal components (i.e., 3a, 3b, 3f, 3g), you should reshape the 4096-dimensional vector into a $64 \times 64$ matrix.