CSCI 567, Summer 2022
Prof. Victor Adamchik

Theory Assignment 2

Due: 11:59 pm, June 6, 2022

Instructions

Submission: Assignment submission will be via courses.uscden.net. By the submission date, there will be a folder named 'Theory Assignment 1' set up in which you can submit your files. Please be sure to follow all directions outlined here.

You can submit multiple times, but only *the last submission* counts. That means if you finish some problems and want to submit something first and update later when you finish, that's fine. In fact you are encouraged to do this: that way, if you forget to finish the homework on time or something happens (remember Murphy's Law), you still get credit for whatever you have turned in.

Problem sets must be typewritten or neatly handwritten when submitted. In both cases, your submission must be a single PDF. It is strongly recommended that you typeset with LaTeX. There are many free integrated LaTeX editors that are convenient to use (e.g Overleaf, ShareLaTeX). Choose the one(s) you like the most. This tutorial Getting to Grips with LaTeX is a good start if you do not know how to use LaTeX yet.

Please also follow the rules below:

- The file should be named as `firstname_lastname_USCID.pdf` e.g., `Don_Quijote_de_la_Mancha_8675309045.pdf`).
- Do not have any spaces in your file name when uploading it.
- Please include your name and USCID in the header of your report as well.

Collaboration: You may discuss with your classmates. However, you need to write your own solutions and submit separately. Also in your report, you need to list with whom you have discussed for each problem. Please consult the syllabus for what is and is not acceptable collaboration. Review the rules on academic conduct in the syllabus: a single instance of plagiarism can adversely affect you significantly more than you could stand to gain.

Notes on notation:

- Unless stated otherwise, scalars are denoted by small letter in normal font, vectors are denoted by small letters in bold font and matrices are denoted by capital letters in bold font.
- $\|.\|$ means L2-norm unless specified otherwise i.e. $\|.\| = \|.\|_2$

Problem 1 Kernel function (32 points)

Recall from lecture that there are two definitions of a kernel function, $k(\boldsymbol{x}, \boldsymbol{x}')$.

1. First k is called a kernel function if there exists a basis function $\phi : \mathbb{R}^D \to \mathbb{R}^M$ such that $k(\boldsymbol{x}, \boldsymbol{x}') = \phi(\boldsymbol{x})^T \phi(\boldsymbol{x}')$.
2. Second, we have Mercer's Theorem which states that k is a kernel function if and only if, for any set of $x_1, x_2, \cdots, x_n \in \mathbb{R}^D$, the resulting Gram matrix is PSD.

Throughout this problem, you can use either of these definitions to check or prove that some function is a valid kernel.

1.1 Consider the function $k(\boldsymbol{x}, \boldsymbol{x}') = \boldsymbol{x}^T \boldsymbol{x}' + (\boldsymbol{x}^T \boldsymbol{x}')^2$ over $\boldsymbol{x} \in \mathbb{R}^2$. Is this a valid kernel function? Show why or why not. **(10 points)**

1.2 Consider the function $k(\boldsymbol{x}, \boldsymbol{x}') = (f(\boldsymbol{x}) + f(\boldsymbol{x}'))^2$ for any function $f : \mathbb{R}^D \to \mathbb{R}$. Is this a valid kernel function? Show why or why not. **(12 points)**

1.3 Now, assume $k_1(\boldsymbol{x}, \boldsymbol{x}')$ and $k_2(\boldsymbol{x}, \boldsymbol{x}')$ are kernel functions. Prove by the Mercer Theorem (from lecture 5) that a linear combination $k(\boldsymbol{x}, \boldsymbol{x}') = \alpha k_1(\boldsymbol{x}, \boldsymbol{x}') + \beta k_2(\boldsymbol{x}, \boldsymbol{x}')$ for some $\alpha, \beta \geq 0$ is also a kernel function. **(10 points)**

Problem 2 Support Vector Machines (32 points)

Consider the dataset consisting of points (x, y), where x is a real value, and $y \in \{-1,\ 1\}$ is the class label. Let's start with three points $(x_1, y_1) = (-1,\ -1)$, $(x_2, y_2) = (1, -1)$, $(x_3, y_3) = (0,\ 1)$.

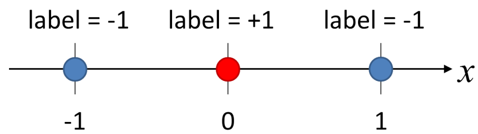

Figure 1: Three data points considered in Problem 2

2.1 Can three points shown in Figure 1, in their current one-dimensional feature space, be perfectly separated with a linear separator? Why or why not? **(4 points)**

2.2 Now we define a simple feature mapping $\phi(x) = [x, x^2]^T$ to transform the three points from one- to two-dimensional feature space. Plot the transformed points in the new two-dimensional feature space. Is there a linear decision boundary that can separate the points in this new feature space? Why or why not? **(4 points)**

2.3 Given the feature mapping $\phi(x) = [x, x^2]^T$, write down the 3×3 kernel (or Gram) matrix $\mathbf{K}$ for the three data points. Show that this Gram matrix is positive semi-definite. Write the Kernel function K(x,y)(defined as $K(x, y) = \phi(x)^T \phi(y)$). **(8 points)**

2.4 Write down the dual formulation of this problem (plugging in the numerical values evaluated using the kernel function). **(8 points)**

2.5 Solve the dual form analytically. Then obtain primal solution $\mathbf{w}^*, b^*$ using dual solution. **(8 points)**

Problem 3 Constrained Optimization (36 points)

Machine learning problems, especially clustering problems, sometimes involve optimization over a **simplex**. In this exercise, you will solve two optimization problems over the simplex. Recall a $K-1$ dimensional simplex Δ is defined as:

$$\Delta = \{\boldsymbol{q} \in \mathbb{R}^K | q_k \geq 0, \forall k \text{ and } \sum_{k=1}^{K} q_k = 1\},$$

which means that a $K-1$ dimensional simplex has K non-negative entries, and the sum of all K entries is 1. This property coincides with the property of the probability distribution of a discrete random variable of K possible outcomes. Thus, the simplex is usually seen in clustering problems.

3.1 Given $a_1, ..., a_K \in \mathbb{R}_{\neq 0}$ (the set of non-zero real numbers), solve the following optimization over the simplex. (find the optimal value $\boldsymbol{q}^*$ of $\boldsymbol{q}$) **(18 points)**

$$\arg\max_{\boldsymbol{q} \in \Delta} \sum_{k=1}^{K} a_k^2 \ln q_k$$

(a) Write down the Lagrangian of this problem. (Hint: use the constraints on q_k given by the simplex Δ) **(4 points)**

(b) Apply KKT conditions on the Lagrangian you derived above to find $\boldsymbol{q}^*$. (Hint: the solution can be written in the form of $q_k^* = ...$) **(12 points)**

(c) The solution you acquired will not have a simple form if a_k is allowed to be 0. Explain why. (Hint: point out the relevant variable. One sentence explanation is sufficient) **(2 points)**

3.2 Next given $c_1, ..., c_K \in \mathbb{R}$, solve the following optimization problem following the same steps in part 1.1. $\boldsymbol{q}$ is under the same constraints as in part 1.1: **(18 points)**

$$\arg\max_{\boldsymbol{q}\in\Delta} \sum_{k=1}^{K} (q_k c_k - q_k \ln q_k)$$