HW Assignment 4, Theory

Problems marked with a (∗) are mandatory for ITCS 8156 students. Bonus problems are optional, solving them will result in extra points.

1 Positive Definite Matrices (20 points)

A matrix $A \in R^{N \times N}$ is called positive semidefinite if $\mathbf{x}^T A \mathbf{x} \geq 0$, for any $\mathbf{x} \in R^{N \times 1}$. Using this definition, prove the following two properties:

1. (10 points) Show that a diagonal matrix $\mathbf{A}$ whose elements satisfy $A_{ii} \geq 0$ is positive semidefinite.
2. (10 points) Show that the sum of two positive semidefinite matrices is itself positive semidefinite.

2 Kernel Techniques (50 points)

In this section, you are asked to use the definition of kernels to prove that a given function of two samples x and y is a valid kernel.

1. (Normalized Kernel, 30 points) Let $K : X \times X \rightarrow R$ be a kernel function defined over a sample space X.
   (a) (20 points) Prove that the function below (a *normalized kernel*) is a valid kernel.
   $$K'(x, y) = \frac{K(x, y)}{\sqrt{K(x, x)K(y, y)}}$$
   (b) (10 points) Why it would not make sense to normalize a Gaussian kernel?
   (c) (10 bonus points) Which types of input data would benefit from normalizing the kernel function? Explain why, and provide real world examples.
2. (Product Kernel, 20 points)) Show that if $K_1(\mathbf{x}, \mathbf{y})$ and $K_2(\mathbf{x}, \mathbf{y})$ are valid kernel functions, then $K(\mathbf{x}, \mathbf{y}) = K_1(\mathbf{x}, \mathbf{y})K_2(\mathbf{x}, \mathbf{y})$ is also a valid kernel.
3. (Rotation Kernel, 20 bonus points) Using the definition of valid kernels, show that if A is a symmetric and positive semidefinite matrix, then $k(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T A \mathbf{y}$ is a valid kernel.
   Hint: Inderjit Dhillon's Linear Algebra Background describes some useful properties of symmetric positive semidefinite matrices.

3 Support Vector Machines (∗), (40 points)

In this section, you are asked to prove basic properties for the separable case (1) and non-separable case (2) of the SVM optimization objective.

1. (20 points) Consider the constrained optimization SVM problem for the separable case shown on slide 7. Show that, if the 1 on the right-hand side of the inequality constraint is replaced by some arbitrary constant $\gamma > 0$, the resulting maximum margin hyperplane is unchanged.

   Hint: Let $(\mathbf{w}, b)$ *be the solution for the original SVM problem where 1 is used, and* $(\mathbf{w}', b')$ *the solution for when* γ *is used. Show that the hyperplane* $\{\mathbf{x} \mid \mathbf{w}'^T\mathbf{x} + b' = 0\}$ *is the same as the hyperplane* $\{\mathbf{x} \mid \mathbf{w}^T\mathbf{x} + b = 0\}$.

2. (20 points) Prove that the sum of slacks $\sum \xi_n$ from the objective function of the SVM formulation with soft margin is an upper bound on the number of misclassified training examples.

4 Large Margin Perceptron $(*)$, (60 points)

Let $\mathbf{u}$ be a current current vector of parameters and $\mathbf{x}$ and $\mathbf{y}$ two training examples such that $\mathbf{u}^T(\mathbf{x}-\mathbf{y}) < 1$. Use the technique of Lagrange multipliers to find a new vector of parameters $\mathbf{w}$ as the solution to the convex optimization problem below:

minimize:

$$J(\mathbf{w}) = \frac{1}{2}\|\mathbf{w} - \mathbf{u}\|^2$$

subject to:

$$\mathbf{w}^T(\mathbf{x} - \mathbf{y}) \geq 1$$

5 Submission

Submit your responses on Canvas as one file named `theory.pdf`. It is recommended to use an editor such as Latex or Word or Jupyter-Notebook that allows editing and proper formatting of equations. Alternatively, if you choose to write your solutions on paper, submit an electronic scan / photo of it on Canvas. Make sure that your writing is legible and the scan has good quality.