156 Discussion Problems, Week 4

Recall: a function $k : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is called a *kernel* if there is some function $\phi : \mathbb{R}^n \to \mathbb{R}^m$ with $m \in \mathbb{N} \cup \{\infty\}$ such that $k(\mathbf{x}, \mathbf{x}') = \phi(\mathbf{x})^\top \phi(\mathbf{x}')$.

Problem 1: Let $k(\mathbf{x}, \mathbf{x}') = \phi(\mathbf{x})^\top \phi(\mathbf{x}')$, $k'(\mathbf{x}, \mathbf{x}') = \phi'(\mathbf{x})^\top \phi'(\mathbf{x}')$ be two kernels, and let $c > 0$. Show that each of the following is also a kernel:

(a) $k_1(\mathbf{x}, \mathbf{x}') = k(\mathbf{x}, \mathbf{x}') + k'(\mathbf{x}, \mathbf{x}')$.

(b) $k_2(\mathbf{x}, \mathbf{x}') = ck(\mathbf{x}, \mathbf{x}')$.

Problem 2: Show that $k(\mathbf{x}, \mathbf{x}') = -\mathbf{x}^\top \mathbf{x}'$ is not a kernel.

Problem 3: Let $k(\mathbf{x}, \mathbf{x}') = \phi(\mathbf{x})^\top \phi(\mathbf{x}')$ be a kernel. Show that each of the following is also a kernel.

(a) $k_1(\mathbf{x}, \mathbf{x}') = f(\mathbf{x}) f(\mathbf{x}') k(\mathbf{x}, \mathbf{x}')$, where f is any real-valued function.

(b) $k_2(\mathbf{x}, \mathbf{x}') = k(\mathbf{x}, \mathbf{x}') k'(\mathbf{x}, \mathbf{x}')$, where k' is any other kernel.

(c) $k_3(\mathbf{x}, \mathbf{x}') = q(k(\mathbf{x}, \mathbf{x}'))$, where q is any polynomial with nonnegative coefficients.

(d) $k_4(\mathbf{x}, \mathbf{x}') = \exp(k(\mathbf{x}, \mathbf{x}'))$.

Problem 4 (Suggested problem 3 from assignment 4). We reproduce the statement:

Consider the soft-margin SVM problem:

$$\begin{aligned} &\text{minimize } \frac{1}{2}\|\mathbf{w}\|_2^2 + C \sum_{n=1}^{N} \xi_n \\ &\text{s.t. } t_n(\mathbf{w}^\top \phi(x_n) + b) \geq 1 - \xi_n \quad \forall n \\ &\xi_n \geq 0 \quad \forall n \end{aligned}$$

Sketch a two-dimensional two-class toy example and answer the following geometrically:

(a) Where does a data point lie relative to where the decision hyperplane and the maximum-margin hyperplanes are when $\xi_n = 0$? Is this data point classified correctly?

(b) Where does a data point lie relative to where the decision hyperplane and the maximum-margin hyperplanes are when $0 < \xi_n \leq 1$? Is this data point classified correctly?

(c) Where does a data point lie relative to where the decision hyperplane and the maximum-margin hyperplanes are when $\xi_n >$? Is this data point classified correctly?