Problem 1. Kernels

Let $k_1$ and $k_2$ be two PSD kernels on a space $\mathcal{X}$.

1. Show that the kernel $k$ defined by $k(u,v) = k_1(u,v)k_2(u,v)$ for any $u,v \in \mathcal{X}$ is PSD. 
   [Hint: consider the Hadamard product between eigenvalue decompositions of the Gram matrices associated to $k_1$ and $k_2$].

2. Let $g: \mathcal{C} \rightarrow \mathbb{R}$ be a given function. Show that the kernel $k$ defined by $k(u,v) = g(u)g(v)$ is PSD.

3. Let $Q$ be a polynomial with nonnegative coefficients. Show that the kernel $k$ defined by $k(u,v) = Q(k_1(u,v))$ for any $u,v \in \mathcal{X}$ is PSD.

4. Show that the kernel $k$ defined by $k(u,v) = \exp(k_1(u,v))$ for any $u,v \in \mathcal{X}$ is PSD.
   [Hint: use series expansion].

5. Let $\mathcal{X} = \mathbb{R}^d$ and $\| \cdot \|$ denote the Euclidean norm on $\mathbb{R}^d$. Show that the kernel $k$ defined by $k(u,v) = \exp(-\|u - v\|^2)$ is PSD.

Problem 2. Convexity and Projections

1. Give an algorithm that computes projections on the set

$$C = \{ x \in \mathbb{R}^d : \max_{1 \leq i \leq d} |x_i| \leq 1 \}$$

and prove a rate of convergence.

2. Give an algorithm that computes projections on the set

$$\Delta = \{ x \in \mathbb{R}^d : \sum_{i=1}^{d} x_i = 1, x_i \geq 0 \}$$

and prove a rate of convergence.
3. Recall that the Euclidean norm on \( n \times n \) real matrices is also known as the Frobenius norm and is defined by \( \|M\|^2 = \text{Trace}(M^\top M) \). Let \( S_n \) be the set of \( n \times n \) symmetric matrices with real entries. Let \( S_n^+ \) denote the set of \( n \times n \) symmetric positive definite matrices with real entries, that is \( M \in S_n \) if and only if \( x^\top M x \geq 0, \ \forall \ x \in \mathbb{R}^n \).

(a) Show that \( S_n^+ \) is convex and closed.

(b) Give an explicit formula for the projection (with respect to the Frobenius norm) of a matrix \( M \in S_n \) onto \( S_n^+ \).

4. Let \( C \subset \mathbb{R}^d \) be a closed convex set and for any \( x \in \mathbb{R}^d \) denote by \( \pi(x) \) its projection onto \( C \). Show that for any \( x, y \in \mathbb{R}^d \), it holds

\[ \|\pi(x) - \pi(y)\| \leq \|x - y\| \]

where \( \| \cdot \| \) denotes the Euclidean norm.

Show that for any \( y \in C \),

\[ \|\pi(x) - y\| \leq \|x - y\| , \]

Problem 3. Convex conjugate

For any function \( f : D \subset \mathbb{R}^d \to \mathbb{R} \), define its convex conjugate \( f^* \) by

\[ f^*(y) = \sup_{x \in C} \left( y^\top x - f(x) \right) . \]  \hspace{1cm} (1)

The domain of the function \( f^* \) is taken to be the set \( D = \{ y \in \mathbb{R}^d : f^*(y) < \infty \} \).

1. Find \( f^* \) and \( D \) if

(a) \( f(x) = 1/x, \ C = (0, \infty) \),

(b) \( f(x) = \frac{1}{2} |x|^2, \ C = \mathbb{R}^d \),

(c) \( f(x) = \log \sum_{j=1}^{d} \exp(x_j), \ x = (x_1, \ldots, x_d), \ C = \mathbb{R}^d \).

Let \( f \) be strictly convex and differentiable and that \( C = \mathbb{R}^d \).

2. Show that \( f(x) \geq f^{**}(x) \) for all \( x \in C \).

3. Show that the supremum in (1) is attained at \( x^* \) such that \( \nabla f(x^*) = y \).

4. Recall that \( D_f(\cdot, \cdot) \) denotes the Bregman divergence associated to \( f \). Show that

\[ D_f(x, y) = D_{f^*}(\nabla f(y), \nabla f(x)) \]
Problem 4. Around gradient descent

In what follows, we want to solve the constrained problem:

$$\min_{x \in C} f(x).$$

where $f$ is a $L$-Lipschitz convex function and $C \subset \mathbb{R}^d$ is a compact convex set with diameter at most $R$ (in Euclidean norm). Denote by $x^*$ a minimum of $f$ on $C$.

1. Assume that we replace the updates in the projected gradient descent algorithm by

$$y_{s+1} = x_s - \eta \frac{g_s}{\|g_s\|}, \quad g_s \in \partial f(x_s).$$

$$x_{s+1} = \pi_C(y_{s+1}),$$

where $\pi_C(\cdot)$ is the projection onto $C$.

What guarantees can you prove for this algorithm under the same assumptions?

2. Consider the following updates:

$$y_s \in \arg\min_{y \in C} \nabla f(x_s)^{\top} y$$

$$x_{s+1} = (1 - \gamma_s)x_s + \gamma sy_s,$$

where $\gamma_s = 2/(s + 1)$.

In what follows, we assume that $f$ is differentiable and $\beta$-smooth:

$$f(y) - f(x) \leq \nabla f(x)^{\top} (y - x) + \frac{\beta}{2} |y - x|^2.$$

(a) Show that

$$f(x_{s+1}) - f(x_s) \leq \gamma_s (f(x^*) - f(x_s)) + \frac{\beta}{2} \gamma_s^2 R^2$$

(b) Conclude that for any $k \geq 2$,

$$f(x_k) - f(x^*) \leq \frac{2\beta R^2}{k + 1}.$$
18.657 Mathematics of Machine Learning
Fall 2015

For information about citing these materials or our Terms of Use, visit: http://ocw.mit.edu/terms.