Exercises in Convex Optimization
Lecture 6

Michel Baes, Patrick Cheridito

October 25, 2018

Design of a cylindrical can The goal is to design a cylindrical can with height h and radius r such that the volume is at least V and the total surface area is minimal.

a) The problem can be posed as follows:

$$\begin{array}{ll} \min & 2\pi(r^2 + rh) \\ \text{subject to} & \pi r^2 h \geq V \\ & r > 0 \\ & h > 0 \end{array}$$

Why is this problem not convex? How can this problem be transformed into a convex optimization problem? (Many different possibilities for treating the problem exist. One of them starts by taking *logarithms* of the variables.)

b) Solve the transformed convex problem, in order to give the optimal h and r in terms of V.

`Lecture_6/CylindricalCan`

Kernel Methods Let us assume, similarly to Lecture 4, that we have a possibly large set of distinct points $\{x_1, \ldots, x_N\}$ from a set $\mathcal{X}$ (which was in the course $\mathbb{R}^d$). These points are separated in two classes, so that x_i has class $p_i = +1$ or $p_i = -1$; Assume that N_+ of the given points are in class $+1$ and N_- are in class -1. The *linear classification problem*, that we studied in the lecture, consists in finding a hyperplane $H = \{x : w^T x + w_0 = 0\}$ that separates the points in the class $\{-1\}$ from the points in the class $\{+1\}$ optimally (which meant in the course "with the widest separation margin"). Clearly, such a separating hyperplane does not always exist. Nevertheless, we can extend this linear separation problem to allow classification for points that are not linearly separated.

We need a kind of *similarity measure*[1] k from $\mathcal{X} \times \mathcal{X}$ to $\mathbb{R}$. This "measure" must be symmetric and positive semidefinite:

1. $k(x, y) = k(y, x)$ for $x, y \in \mathcal{X}$;

[1] Here, "measure" is is not understood in the measure theoretic sense.

2. for every finite set $x_1, \dots, x_n \in \mathcal{X}$, the matrix $[k(x_i, x_j)]_{ij}$ is positive semidefinite.

This measure will, in our case be defined from a *feature function* $\Phi : \mathcal{X} \to \mathcal{Z}$, where $\mathcal{Z}$ is a vector space endowed with a scalar product $\langle \cdot, \cdot \rangle$, and $k(x, y) = \langle \Phi(x), \Phi(y) \rangle$. (Sometimes, the space $\mathcal{Z}$ can have an enormous dimension, even infinite in many applications.)

Now, instead of trying to separate the x_i's in $\mathcal{X}$ with a hyperplane, the idea is to try to separate the $\Phi(x_i)$'s in $\mathcal{Z}$ with a hyperplane.

The trivial classification kernel Suppose that $\mathcal{X} = \mathbb{R}^d$, $\mathcal{Z} = \mathbb{R}^N$ and that $\Phi : \mathbb{R}^d \to \mathbb{R}^N$ is any function for which $\Phi(x_j) = p_j \cdot [p_1; \dots; p_N]$. Then $k(x_i, x_j) = N$ iff $p_i = p_j$, and $-N$ otherwise. Can you determine a hyperplane that separates perfectly the points $\Phi(x_i)$? Prove that the N-by-N matrix $[k(x_i, x_j)]_{ij}$ is positive semidefinite. Is it positive definite?

The circular separator Suppose that $\mathcal{X} = \mathbb{R}^2$ and define $\Phi(u, v) := (u, v, u^2 + v^2)$ for $u, v \in \mathbb{R}$, so that $\mathcal{Z} = \mathbb{R}^3$. Suppose that every point $p_i = -1$ iff $||x_i||_2 \leq r$ and $p_i = +1$ otherwise (Note that such a set cannot be separated by any hyperplane (i.e. straight line) in $\mathbb{R}^2$). Formulate the linear separation problem of the $\Phi(x_i)$'s in $\mathbb{R}^3$ and solve it (Define $r_i := u_i^2 + v_i^2$, with $x_i = (u_i, v_i)$ for all i).

It turns out that, given a similarity measure k with properties as above, it is *always* possible to find $\mathcal{Z}$, a scalar product on $\mathcal{Z}$, and a function $\Phi : \mathcal{X} \to \mathcal{Z}$. We shall verify this result when $\mathcal{X}$ is a finite set[2].

Suppose $\mathcal{X} = \{x_1, \dots, x_N\}$. Define $\Phi : \mathcal{X} \mapsto \mathbb{R}^N$ so that $\Phi(x) = k(x, \cdot)$, i.e. $\Phi(x)_i = k(x, x_i)$. for all $x \in \mathcal{X}$ and $1 \leq i \leq N$.

1. We define $\mathcal{Z}$ as the linear span of $\{\Phi(x_1), \dots, \Phi(x_N)\}$ and endow it with the bilinear form
$$\langle f, g \rangle = \sum_{i,j=1}^{N} \alpha_i \beta_j k(x_i, x_j)$$
for $f = \sum_{i=1}^{N} \alpha_i \Phi(x_i)$, $g = \sum_{j=1}^{N} \beta_j \Phi(x_j) \in \mathcal{Z}$.
Prove that, even if f or g can have many expansions as a sum of $\Phi(x_i)$'s, the value of $\langle f, g \rangle$ does not depend on this expansion.
2. Prove that $\langle \cdot, \cdot \rangle$ is a bilinear, symmetric and that $\langle f, f \rangle \geq 0$ for all $f \in \mathcal{Z}$.
3. Prove that $k(x, y) = \langle \Phi(x), \Phi(y) \rangle$ for every $x, y \in \mathcal{X}$.
4. Prove that if k is positive definite (in the sense that the matrix $[k(x_i, x_j)]_{ij}$ is positive definite), then $\langle \cdot, \cdot \rangle$ is a scalar product.

`Lecture_6/KernelMethods`

[2] This important result is actually true for *any* set, as proved by *Mercer's Theorem*. It gives rise to vast generalization of this method, beyond classification problems (e.g. interpolation problems, ...).