NUMERICAL OPTIMIZATION

Sheet 11: All about Support Vector Machines

EXERCISE ONE Recall the problem *SVM*, or *Support Vector Machines.* In it, you are given a list of pairs x_i, y_i, where $x \in \mathbb{R}^n$ is a point and $y_i \in \{-1, 1\}$. The task is to find the best hyperplane separating the 1-points from the -1-points, and we measure its quality by maximizing the margin between all points and the hyperplane.

Your first task is to recall (or re-create) its constrained optimization formulation.

Hints:

1. The goal is to have a variable $b \in \mathbb{R}$ and a vector of variables $w \in \mathbb{R}^n$.
2. The objective function should be $\min ||w||_2^2$.
3. There should be one constraint in the convex program for each data point $\{x_i, y_i\}$.
4. Recall that a hyperplane is a set of points $\{x | w^T x + b = 0\}$.
5. The following fact will be helpful – for two parallel hyperplanes $\{x_1 | w^T x_1 + b = 1\}$ and $\{x_2 | w^T x_2 + b = -1\}$, their distance is equal to $\frac{2}{||w||_2}$.
6. Why is it enough that we minimize $||w||_2^2$?

EXERCISE TWO With the formulation recovered, it is time to start dualization of this SVM formulation. We start slow:

1. Write the Lagrangian of this problem.
2. Recall the general form of any dual for a constrained optimization problem.
3. We can actually already tell if strong duality holds for SVMs. Or, more precisely, that it holds if a very simple condition is met. Can you use some criterion to check if strong duality holds?

EXERCISE THREE We continue with dualization. Write the function $g(\lambda)$ and the infimum section in it. This infimum is going over all choices of $w \in \mathbb{R}^n$ and $b \in \mathbb{R}$.

As with many other dualization exercises, we can read from a function $g(\lambda)$ a condition that if it is not met, then $g(\lambda) = -\infty$. Can you detect this condition and also which variable we can “send to infinity” unless this condition is met?

EXERCISE FOUR With this condition, we need to still compute the infimum section of $g(\lambda)$. However, since the dual function is convex, we can use very simple machinery from mathematical analysis. (We have also done this before.) Find the point where the infimum section of $g(\lambda)$ takes the minimal value and write the value of $g(\lambda)$ at that point.

Then, finally, write the dual formulation of *Support Vector Machines.*