Assignment 6

Machine Learning: Algorithms and Theory

Prof. Ulrike von Luxburg / Diego Fioravanti / Moritz Haas
Tobias Frangen / Siavash Haghiri

Summer term 2018 — due to **June 5**

1. When you hand in your assignment you need to hand in the notebook too. Please do not write down a report with just results and/or figures. Ideally you should email it, it is easier to correct and more eco friendly, but we accept printed versions too. From now on not handing in the notebook will result in 0 points for the programming part.
2. Before the end of the course you need to present at least one of your solution in the tutorial. If you do not do that you cannot take the exam! If for any reason you cannot attend let us know, it is possible to change group or find another solution.
3. Join the class on ILIAS otherwise we cannot contact you if we need to.

Exercise 1 (Duality, 1+1+1+1+1 points)
The goal of this exercise is to manipulate the Lagrangian of a simple problem to understand the notion of duality. For additional material on Lagrangian you can read Chapter 5 on duality in `Convex Optimization` by Boyd & Vandenberghe (2004).
Consider the following optimization problem:

$$\underset{x \in \mathbb{R}}{\text{minimize}} f(x) = (x-1)^2$$
$$\text{subject to } g(x) = 3x - 1 \leq 0$$

(a) Is it convex? Justify your answer. Plot the function and the feasible set by hand (a sketch is enough). How can you see from the plot whether the optimization problem is convex?

(b) Write down the Lagrangian $L(x, \alpha)$, where x denotes the primal variables and α the Lagrange multipliers, and compute the saddle point.

(c) Plot $L(x, \alpha)$ on a grid from $-1 \leq x, \alpha \leq 1$ with step size 0.1. Use the functions `numpy.linspace`, `numpy.meshgrid` and `Axes3D.plot_surface` to do so. Here `https://matplotlib.org/mpl_toolkits/mplot3d/tutorial.html` you can find examples for how to plot in 3D. Plot the saddle point obtained in the previous question. Once plotted rotate the figure to inspect $L(x, \alpha)$ and verify that the analytic saddle point is correct.

(d) Write down the dual function and simplify it. Is it concave? Justify your answer.

(e) Write down the dual optimization problem and solve it by hand. Use the solution of the dual problem to derive the solution of the primal problem. Does strong duality hold?

Exercise 2 (SVM by hand, 1+1+1+1 points)
Consider a dataset containing 3 data points in $\mathbb{R}^2$: $x_1 = (0, 0)$, $x_2 = (1, 2)$, $x_3 = (-1, 2)$ and their corresponding labels $y_1 = -1$, $y_2 = 1$, $y_3 = 1$.

(a) Write down the primal problem of the hard margin SVM on this dataset. Compute the Lagrangian $L(w, b, \alpha)$.

(b) What is the dimension of w? of α? Compute the saddle point and mark with $(*)$ and $(**)$ the steps of your derivation where you see the saddle point condition given in the lecture on slide 350.

(c) Apply `sklearn.svm.SVC`, with the correct parameters, on this dataset to compute the SVM model. Check its attributes `coef_`, that gives w, `dual_coef_`, that gives the products $y_i \alpha_i$, and `intercept_`, that gives b. Are the values equal to your analytic solution? Why?

(d) Explain what a support vector is. Add 2 data points to the dataset such that the number of support vectors becomes 2 for the new SVM model. Give the new values of α.

Exercise 3 (Duality gap, 1+1+1 points)
Consider the following optimization problem:

$$\begin{aligned} \underset{x \in \mathbb{R}}{\text{minimize}} f(x) &= x^4 - 10x^2 + x \\ \text{subject to } g(x) &= x^2 - 2x - 3 \leq 0 \end{aligned}$$

(a) Write down the feasible set of this optimization problem and solve its primal. You do not have to solve the primal by hand and you can use a solver of your choice, for example WolframAlpha.

(b) Compute the Lagrangian and the dual problem. Derive a lower bound on the primal using the optimal solution of the dual. You can use the fact that the dual function is optimal when $\alpha = 0.5$. You do not have to solve the dual by hand and you can use a solver of your choice, for example WolframAlpha.

(c) Does strong duality hold in this case? Explain why and quantify the duality gap if any.