Machine Learning 1 Homework Week 5

Monday, September 28, 2015
Deadline: Friday, October 9, 2015, 23:59

1 Lagrange Multipliers: Warm-up

In class we have seen the formulation of the Support Vector Machines (SVMs), which has been successfully used for many classification tasks. We saw that the parameters of the SVMs are learned using the Lagrange Multipliers. The Lagrange Multiplier is a method of optimizing a function subject to a set of constraints, *e.g.* in SVMs we have the following optimization problems

$$\min_{\mathbf{w}} \frac{1}{2}\|\mathbf{w}\|^2 \tag{1}$$

$$s.t. \; \forall i \in \{1, 2, \ldots, N\} : y_i(\mathbf{w} \cdot \mathbf{x_i} - b) \geq 1 \tag{2}$$

In this exercise, we will deal the same optimization problems using Lagrange Multipliers. Suppose we would like to maximize the function

$$f(\mathbf{x}) = 1 - x_1^2 - 2x_2^2 \tag{3}$$

which has two input dimensions x_1 and x_2 (they can be considered as the parameters that we would like to learn). This function is plotted in Figure 1(a). We can see the function is concave and there is no local minimum. The optimization of the function is subject to a constraint function. Therefore the optimal solution that is found has to satisfies the constraints.

For example, if we set the constraints $x_1 + x_2 = 1$, then the optimization problem is to find the maximal value of $f(\mathbf{x})$ where $\mathbf{x}$ is also on the constraint plane. Figure 1(b) shows the constraint function (black) separates $f(\mathbf{x})$ into two parts. The 3D view of $f(\mathbf{x})$ is slighted changed in Figure 1(b) for better visualization.

Answer the following questions:

1. Find the maximum of $1 - x_1^2 - 2x_2^2$, subject to the constraint that $x_1 + x_2 = 1$.

2. Find the maximum of $1 - x_1^2 - x_2^2$ subject to the constraint $x_1 + x_2 - 1 \geqslant 0$

3. Find the maximum of $1 - x_1^2 - x_2^2$ subject to the constraint $-x_1 - x_2 + 1 \geqslant 0$

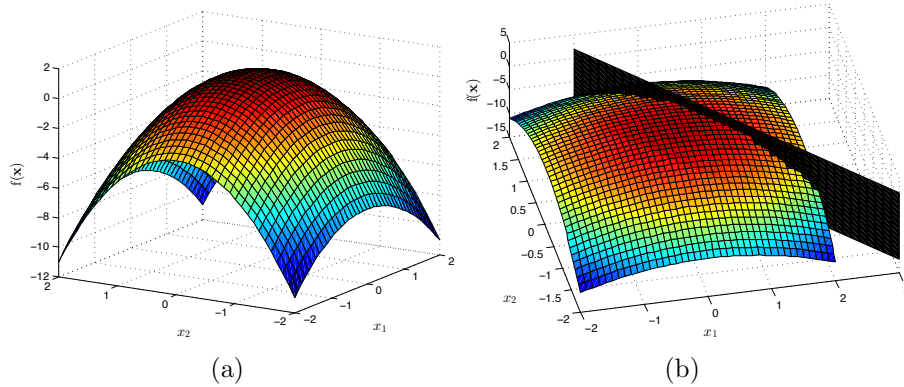(a)   (b)

Figure 1: (a) $f(\mathbf{x}) = 1 - x_1^2 - 2x_2^2$ Plot of the example function. Plot (b) also illustrates the constraint surface.

4. Find the maximum of $x_1 + 2x_2 - 2x_3$, subject to the constraint that $x_1^2 + x_2^2 + x_3^2 = 1$.

5. A company manufactures a chemical product out of two ingredients, known as ingredient X and ingredient Y. The number of doses produced, D, is given by $6x^{2/3}y^{1/2}$, where x and y are the number of grams of ingredients X and Y respectively. Suppose ingredient X consts 4 euro per gram, and ingredient Y costs 3 euro per gram. Find out the maximum number of doses that can be made if no more than 7000 euro can be spent on the ingredients.

2   Kernel Outlier Detection

Consider the picture in Figure 2. The dots represent data-items. Our task is to derive an algorithm that will detect the outliers (in this example there are 2 of them). To that end, we draw a circle rooted at location $\boldsymbol{a}$ and with radius R. All data-cases that fall outside the circle are detected as outliers.

We will now write down the primal program that will find such a circle:

$$\min_{\boldsymbol{a},R,\boldsymbol{\xi}} R^2 + C \sum_{i=1}^{N} \xi_i$$

$$s.t. \ \forall i : \|\boldsymbol{x_i} - \boldsymbol{a}\|^2 \leq R^2 + \xi_i, \ \xi_i \geq 0$$

In words: we want to minimize the radius of the circle subject to the constraint that most data-cases should lay inside it. Outliers are allowed to stay outside but they pay a price proportional their distance from the circle boundary and C.
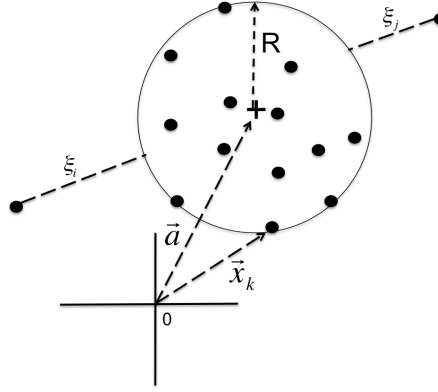
Answer the following questions:

Figure 2: Kernel Outlier Detection

1. Introduce Lagrange multipliers for the constraints and write down the primal Lagrangian. Use the following notation: $\{\alpha_i\}$ are the Lagrange multipliers for the first constraint and $\{\mu_i\}$ for the second constraint.

2. Write down all KKT conditions. (Hint: take the derivative w.r.t. R^2 instead of R).

3. Identify the complementary slackness conditions. Use these conditions to derive what data-cases (e.g. in Figure 2) will have $\alpha_i > 0$ (support vectors) and which ones will have $\mu_i > 0$.

4. Derive the dual Lagrangian and specify the dual optimization problem. Kernelize the problem, i.e. write the dual program only in terms of kernel entries and Lagrange multipliers.

5. The dual program will return optimal values for $\{\alpha_i\}$. In terms of these, compute the optimal values for the other dual variables $\{\mu_i\}$.

   Then, solve the primal variables $\{\boldsymbol{a}, R, \boldsymbol{\xi}\}$ (in that order) in terms of the dual variables $\{\mu_i, \alpha_i\}$. Note that you do not need to know the dual optimization program to solve this question. You only need the KKT conditions.

6. Assume we have solved the dual program. We now want to apply it to new test cases. Describe a test in the dual space (i.e. in terms if kernels and Lagrange multipliers) that could serve to detect outliers. (Students who got stuck along the way may describe the test in primal space).

7. What kind of solution do you expect if we use $C = 0$. And what solution if we use $C = \infty$?

8. Describe geometrically what kind of solutions we may expect if we use a RBF kernel (Gaussian) with very small bandwidth (sigma = small), i.e. describe how these solutions can be different geometrically (in x-space) from the case with a linear kernel.

9. Now assume that you are given labels (e.g. y=1 for outlier and y=-1 for "inlier"). Change the primal problem to include these labels and turn it into a classification problem similar to the SVM. (You do not have to derive the dual program).