6. Constrained Optimization and SVM

Dienstag, 27. November 2018 13:00

1 Duality of finding maxima of a set

Problem 1: Given a set of variables $x_1, ..., x_N \in \mathbb{R}$, define an equation that finds the largest value in the set via minimization. Then, use the Lagrange dual function to derive a second, equivalent maximization problem.

Problem 2: Given a set of variables $x_1, ..., x_N \in \mathbb{R}$, define an equation that calculates the sum of the k largest values via maximization. Then, use the Lagrange dual function to derive a second, equivalent minimization problem.

$$\text{maximize} \quad \sum_{i=1}^{N} w_i x_i$$

$$\text{subject to} \quad \sum_{i=1}^{N} w_i = k$$

$$w_i - 1 \leq 0 \qquad i = 1, \dots, N$$

$$-w_i \leq 0$$

$\boxed{\theta = \vec{w}}$, $\vec{\alpha} = (b, \vec{s}, \vec{t})$

1) $L(\vec{w}, b, \vec{s}, \vec{t}) = -\sum_{i=1}^{N} w_i x_i + b\left(\sum_{i=0}^{N} w_i - k\right) + \sum_{i=1}^{N} s_i (w_i - 1) - \sum_{i=1}^{N} t_i w_i$

2) $\vec{\nabla}_{\vec{w}} L = 0$

$\partial_{w_i} L(\vec{w}, b, s, t) = -x_i + b + s_i - t_i = 0$

$g(b, \vec{s}, \vec{t}) = L(\vec{w}^*(b, \vec{s}, \vec{t}), b, \vec{s}, \vec{t}) = \left[\sum_{i=1}^{N} w_i^* \underbrace{(-x_i + b + s_i - t_i)}_{0}\right] - kb - \sum_{i=1}^{N} s_i =$

$= -kb - \sum_{i=1}^{N} s_i$

3) ~~maximize~~ minimize $\quad +kb + \sum_{i=1}^{N} s_i$

subject to $\quad b \geq x_i - s_i$

$s_i \geq 0 \qquad i = 1, \dots, N$

$t_i \geq 0$

$-x_i + b + s_i = t_i \geq 0$

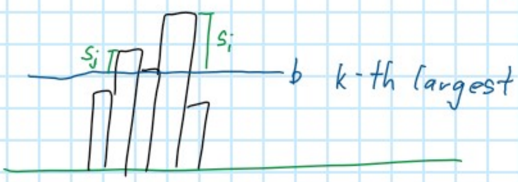
$b \geq x_i - s_i$

$s_i \geq x_i - b$

b:

$f_i(x) = 0$

$\Leftrightarrow f_i(x) \geq 0 \wedge f_i(x) \leq 0$



2 Constrained Optimization Toy Problem

Suppose we have 40 pieces of raw material. Toy A can be made of one piece material with 3 EUR machining fee. A larger toy B can be made from two pieces of material with 5 EUR machining fee.

Because distribution costs decrease with larger quantities, we can sell x pieces of toy A for $20 - x$ EUR each, and y pieces of toy B for $40 - y$ EUR each. From our experience, toy B is more popular than toy A; therefore, we will produce not more of toy A than of toy B.

To get the maximum profit, we want to calculate the amount of toy A and toy B that we should produce.

Problem 3: Write down the constrained optimization problem and the associated Lagrangian.

$$\min \; f(x,y) = -[x(20-x) + y(40-y) - 3x - 5y] =$$
$$= x^2 - 17x + y^2 - 35y$$
$$\text{s.t.} \quad f_1(x,y) = x + 2y - 40 \leq 0$$
$$f_2(x,y) = x - y \leq 0$$
$$\text{Lag.:} \quad L(x,y,\alpha_1,\alpha_2) = x^2 - 17x + y^2 - 35y + \alpha_1(x+2y-40) + \alpha_2(x-y)$$
$$\alpha_1 \geq 0 \quad \alpha_2 \geq 0$$

Problem 4: Write down the Karush–Kuhn–Tucker (KKT) conditions for the above optimization problem.

Primal feas.: $x + 2y - 40 \leq 0$

$\underline{x - y \leq 0}$

Dual feas.: $\alpha_1 \geq 0$

$\alpha_2 \geq 0$

Complementary slackness: $\alpha_1(x + 2y - 40) = 0$

$\alpha_2(x - y) = 0$

x, y min

$$0 \quad \frac{\partial L}{\partial x} = 2x - 17 + \alpha_1 + \alpha_2 = 0$$
$$0 \quad \frac{\partial L}{\partial y} = 2y - 35 + 2\alpha_1 - \alpha_2 = 0$$

Problem 5: Obtain the solution to the constrained optimization problem by solving the KKT conditions. Do not worry about non-integer production quantities.

Problem 5: Obtain the solution to the constrained optimization problem by solving the KKT conditions. Do not worry about non-integer production quantities.

Guesses:
$$\left[\begin{array}{l} \alpha_2 = 0 \\ \rightarrow x + 2y - 40 = 0 \end{array}\right] \begin{array}{l} \leftarrow \text{not constrained} \\ \leftarrow \text{constrained } (\alpha_1 > 0) \end{array}$$

$$x = 40 - 2y$$

$$2(40 - 2y) - 17 + \alpha_1 = 0$$
$$2y - 35 + 2\alpha_1 = 0$$

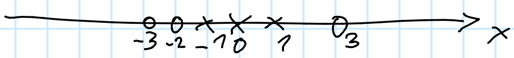
$$y = 16{,}1 \Rightarrow x = 7{,}8$$

$$\alpha_1 = 1{,}4 > 0$$

3 Concrete SVM Example

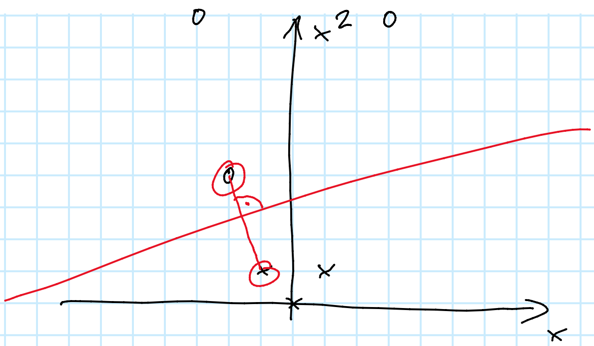
You are given a data set with data from a single feature x in $\mathbb{R}$ and corresponding labels $y \in \{+1, -1\}$. Data points for $+1$ are at $-3, -2, 3$ and data points for -1 are at $-1, 0, 1$.

Problem 6: Can this data set in its current feature space be separated using a linear separator? Why/why not?



Now, we define a simple feature map $\boxed{\phi(x) = (x, x^2)}$ that transforms points in $\mathbb{R}$ to points in $\mathbb{R}^2$.

Problem 7: After applying ϕ to the data, can it now be separated using a linear separator? Why/why not? (Plotting the data may help you with your answer.)



Problem 8: *Construct* a maximum-margin separating hyperplane (i.e. you do not need to solve a quadratic program). Clearly mark the support vectors. Also draw the resulting decision boundary in the feature space $\phi(x) = (x, x^2)$. Is it possible to add another point to the training set in such a way, that the hyperplane *does not* change? Why/why not?

Problem 9: For this specific training set write down the SVM optimization problem, the Lagrangian, the Lagrange dual function and the dual problem.

Problem 10: Write down the KKT conditions for this training set explicitly and verify that the maximum-margin hyperplane you constructed satisfies them.

Complementary slackness:

$$\alpha_1 = \alpha_3 = \alpha_5 = \alpha_6 = 0$$

$\frac{\partial L}{\partial b} = 0$: $\quad \alpha_1 + \alpha_2 + \alpha_3 = \alpha_4 + \alpha_5 + \alpha_6$

$$\alpha_2 = \alpha_4$$

$\frac{\partial L}{\partial w_i} = 0$: $\quad w_1 = -3\alpha_1 - 2\alpha_2 + 3\alpha_3 + \alpha_4 - \alpha_6$

$$w_2 = 9\alpha_1 + 4\alpha_2 + 9\alpha_3 - \alpha_4 - \alpha_6$$

$$w_1 = -\alpha_2$$

$$w_2 = 3\alpha_2$$

$$m = \frac{2}{\|\vec{w}\|_2} = \|\phi(x_4) - \phi(x_2)\| = \sqrt{1^2 + 3^2} = \sqrt{10}$$

$$\|\vec{w}\| = \sqrt{w_1^2 + w_2^2} = \sqrt{\alpha_2^2 + 9\alpha_2^2} = \sqrt{10}\,\alpha_2 = \frac{2}{m} = \frac{2}{\sqrt{10}}$$

$$\alpha_2 = \alpha_4 = \frac{2}{10} = 0{,}2$$