Übung zur Vorlesung

Maschinelles Lernen 2

Sommersemester 2012

Abteilung Maschinelles Lernen
Institut für Softwaretechnik und theoretische Informatik
Fakultät IV, Technische Universität Berlin
Prof. Dr. Klaus-Robert Müller
Email: klaus-robert.mueller@tu-berlin.de

Exercise Sheet 11

Due **July 1**, 9am local time on ISIS

20. **Convex optimization (25 + 25 + 25 + 25 P)**

(a) Let $X \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$. Write

$$\mathcal{C}(X,b) := \{w \in \mathbb{R}^n \;:\; X \cdot w \geq b\},$$

where the inequality is, as usual, component-wise. *Describe* the set $\mathcal{C}(X,b)$ in terms of X and b, in words, and by giving a graphic interpretation. *Prove* that $\mathcal{C}(X,b)$ is always a convex subset of $\mathbb{R}^n$.

(b) Let $X \in \mathbb{R}^{m \times n}$ such that $\mathcal{C}(X,b)$ is non-empty for some $b \in \mathbb{R}^m, b > 0$. Consider the optimization problem

$$\begin{aligned} \max_{\boldsymbol{w} \in \mathbb{R}^n, \lambda \in \mathbb{R}} \quad & \lambda \\ \text{subject to} \quad & \|\boldsymbol{w}\|_2^2 \leq 1, \quad \text{and} \\ & \mathbb{1} \cdot \lambda \leq X \cdot w. \end{aligned} \tag{1}$$

where $\mathbb{1} \in \mathbb{R}^m$ is the vector of ones, and $\|.\|_2$ is the usual Euclidean norm on $\mathbb{R}^n$.

Prove that problem (1) is a convex optimization problem. *Describe* the optimal solution (w_{max}, λ_{max}) in terms of some suitable $\mathcal{C}(.,.)$, by providing a graphical interpretation.

(*Hint:* Problem (1) is - due to a better graphical interpretation - formulated as a maximization problem, while convex problems are usually formulated as minimization problems. However, each minimization problem can be transformed into a maximization problem and vice versa by a change of sign.

(c) Consider the optimization problem

$$\begin{aligned} \min_{\boldsymbol{\alpha} \in \mathbb{R}^m} \quad & \left\|X^\top \cdot \boldsymbol{\alpha}\right\|_2 \\ \text{subject to} \quad & \mathbb{1}^\top \alpha = 1, \quad \text{and} \\ & \alpha \geq 0. \end{aligned} \tag{2}$$

Show that Problem (2) is the Lagrange dual of Problem (1).

(*Hint:* Initially, there is a separate Lagrange/slack variable for the boundary condition $\|\boldsymbol{w}\|^2 \leq 1$. It carries through the computation but can be removed by an explicit maximization.)

(d) *Describe* how a solution for the primal problem (1) can be obtained from a solution of the dual Problem (2). *Describe* the optimal solution (α_{min}) in terms of some suitable $\mathcal{C}(.,.)$, by providing a graphical interpretation, and relate your description of the optimum in (b) to it. *Show* that the duality gap is zero, and explain what this means in your graphical interpretation.

Please ask questions in the ISIS discussion forums for Machine Learning 2:
`https://www.isis.tu-berlin.de/course/view.php?id=8005`