Assignment 5

Machine Learning, Summer term 2014, Ulrike von Luxburg

To be discussed in exercise groups on May 19-21

Exercise 1 (Primal hard margin SVM problem, 1+3 points) Given training data $(X_i, Y_i)_{i=1,\dots,n}$ with $X_i \in \mathbb{R}^d$ and $Y_i \in \{-1, +1\}$ the primal hard margin SVM problem is given as

$$\begin{aligned} &\min_{w \in \mathbb{R}^d, b \in \mathbb{R}} \quad \frac{1}{2}\|w\|^2 \\ &Y_i(w^T X_i - b) \geq 1, \quad i = 1, \dots, n \end{aligned} \tag{1}$$

(a) Recall the meaning of a hyperplane in canonical representation. Show that any solution of (1) gives rise to a hyperplane in canonical representation.

(b) Assume the data is linearly seperable, that is there exists a solution of (1). Show that this solution is unique.

Linear Programming (LP): A linear program is a special case of a convex optimization problem. We want to optimize a linear objective function, subject to linear constraints. For example, consider the following linear program:

$$\begin{aligned} \text{minimize} \quad & 4x_1 + 3x_2 - x_3 \\ \text{subject to} \quad & -x_1 + x_2 \leq 1 \\ & 4x_1 - 2x_2 + 3x_3 \leq -2 \\ & -2x_2 - 3x_3 + 4 \leq 0 \\ \text{and} \quad & x_i \leq 0\ ;\ i = 1, 2, 3 \end{aligned}$$

We can rewrite this linear program in a standard form

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax \leq b \\ \text{and} \quad & x \leq 0 \end{aligned} \tag{2}$$

where $x = (x_1, x_2, x_3)^T \in \mathbb{R}^3$, $A = \begin{bmatrix} -1 & 1 & 0 \\ 4 & -2 & 3 \\ 0 & -2 & -3 \end{bmatrix}$, $b = \begin{bmatrix} 1 \\ -2 \\ -4 \end{bmatrix}$ and $c = \begin{bmatrix} 4 \\ 3 \\ -1 \end{bmatrix}$.

Exercise 2 (LP in standard form, 2 points) Make a transformation of the variables such that you can write the following linear program in the standard form (2). Determine the corresponding matrix A and the vectors c and b.

$$\begin{aligned} \text{minimize} \quad & x_1 - 2x_2 + 4x_3 \\ \text{subject to} \quad & -x_1 + x_2 \geq 1 \\ & 3x_1 + 2x_3 \leq -1 \\ & -2x_1 - 5x_3 + 4 \leq 0 \\ & x_1 + x_2 + 8x_3 \leq 10 \\ \text{and} \quad & x_1, x_2 \leq 0 \\ & x_3 \geq 0 \end{aligned}$$

Exercise 3 (LP and its dual, 2+1 points) We want to derive the Lagrangian dual problem for the linear program (2). We assume $x, c \in \mathbb{R}^d$, $A \in \mathbb{R}^{n\times d}$, $b \in \mathbb{R}^n$. First form the Lagrangian

$$L(x, \lambda_1, \lambda_2) = c^T x + \lambda_1^T (Ax - b) + \lambda_2^T x$$

where $\lambda_1 \in \mathbb{R}^n$ and $\lambda_2 \in \mathbb{R}^d$ are vectors of Lagrange multipliers.

(a) For any pair $(\lambda_1, \lambda_2) \in \mathbb{R}^n \times \mathbb{R}^d$ determine

$$g(\lambda_1, \lambda_2) = \inf_{x\in\mathbb{R}^d} L(x, \lambda_1, \lambda_2).$$

g is called the Lagrange dual function. *Hint: In particular, this requires to determine the pairs* (λ_1, λ_2) *for which* $\inf_{x\in\mathbb{R}^d} L(x, \lambda_1, \lambda_2) = -\infty$.

The Lagrangian dual problem is given by

$$\begin{aligned} &\text{maximize} && g(\lambda_1, \lambda_2) \\ &\text{subject to} && \lambda_1, \lambda_2 \geq 0 \end{aligned}$$

(b) Show that in our case the dual problem can be written as a linear program. (You do not have to rewrite it in standard form (2)).

Optimization in MATLAB: For the following, we highly recommend to use the CVX optimization package (read `prepare05.pdf` for an introduction to this package - available on the course webpage). However, in principle you could also use the MATLAB functions `linprog` and `quadprog`.

Exercise 4 (Solving a linear program, 3 points) Solve in MATLAB the linear program

$$\begin{aligned} &\text{minimize} && c^T x \\ &\text{subject to} && Ax \leq b \end{aligned} \tag{3}$$

where $A = \begin{bmatrix} -1 & -1 \\ -0.5 & -1 \\ -2 & -1 \end{bmatrix}$, $b = \begin{bmatrix} -4 \\ -2 \\ -4 \end{bmatrix}$ and $c = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Then solve the program (3) with A and b replaced by

$$\widetilde{A} = \begin{bmatrix} -1 & -1 \\ -1 & -1 \\ -0.5 & -1 \\ -2 & -1 \end{bmatrix}, \qquad \widetilde{b} = \begin{bmatrix} -2 \\ -4 \\ -2 \\ -4 \end{bmatrix}.$$

Do you get the same solution? What would you expect? Try to solve the system by hand and explain.

Exercise 5 (SVM cancer detection, 4 points) In this exercise you should learn a (soft margin) SVM that classifies cancers as either benign (-1) or malignant (+1) depending on the characteristics of sample biopsies. Load the patients data from `cancer_data2014.mat` (available on the course webpage). For every patient, 9 attributes are measured:

1- Clump thickness 2- Uniformity of cell size 3- Uniformity of cell shape 4- Marginal Adhesion 5- Single epithelial cell size 6- Bare nuclei 7- Bland chomatin 8- Normal nucleoli 9- Mitoses.

For $C \in \{0.01, 0.1, 0.5, 1, 5, 10, 50\}$ train a SVM classifier on the training data and evaluate it on the test data. Plot the train and the test error (with respect to the 0-1-loss) as a function of C. What is the effect of choosing a large C on the training error? Does this effect coincide with what you are expecting?