Machine Learning	Prof. Dr. Bernt Schiele
	Dr. Paul Swoboda & Dr. Gerard Pons-Moll
Exercise Sheet 4 - 12.11.2018	due: 19.11.2018, 14:15

Exercise 8 - Projected Gradient Descent for the Lasso

Let $y \in \mathbb{R}^n$ be the n outputs and $\Phi \in \mathbb{R}^{n \times D}$ the design matrix of a regression problem (D basis functions $\phi_1, \ldots, \phi_D$, then $\Phi_{ij} = \phi_j(x_i)$, $i = 1, \ldots, n$, $j = 1, \ldots, D$).

The Lasso problem

$$\min_{w \in \mathbb{R}^D} \frac{1}{n} \|Y - \Phi w\|_2^2 + \lambda \|w\|_1 \tag{1}$$

can be rewritten into the following smooth, constrained optimization problem

$$\begin{aligned} \min_{w^+, w^- \in \mathbb{R}^D} \quad & \frac{1}{n} \left\|Y - \Phi w^+ + \Phi w^-\right\|_2^2 + \lambda \sum_{i=1}^{D} w_i^+ + \lambda \sum_{i=1}^{D} w_i^- \\ \text{subject to:} \quad & w_i^+ \geq 0, \ i = 1, \ldots, D, \\ & w_i^- \geq 0, \ i = 1, \ldots, D. \end{aligned} \tag{2}$$

a. **(3 points)** The projection $P_C : \mathbb{R}^d \to \mathbb{R}$ onto a convex set C is defined for $x \in \mathbb{R}^d$ as

$$P_C(x) := \arg\min_{y \in C} \frac{1}{2} \|x - y\|_2^2 .$$

1. Show that the projection onto a convex set is uniquely defined (Under which condition on the objective is the global minimum unique ?)
2. Derive an analytical expression for the projection onto the convex set

$$C = \{x \in \mathbb{R}^d \,|\, x_i \geq 0\}, \quad \text{(positive orthant in } \mathbb{R}^d).$$

b. **(5 points)** Instead of the interior point method introduced in the lecture we use projected gradient descent which is a more simple method for constrained convex optimization. Let C be a convex, closed set and ϕ the differentiable, convex objective function, then the convex optimization problem $\min_{x \in C} \phi(x)$ can be solved via projected gradient descent which is defined as

$$x_{t+1} = P_C(x_t - \alpha_t \nabla \phi(x_t)),$$

where $\alpha_t > 0$ is the stepsize.

1. Complete the Matlab Function `Lasso` which has as arguments Y, Φ, λ and returns the weight vector w of the Lasso problem in Equation (1). Use projected gradient descent for the optimization problem (2).
2. Run your Lasso implementation with linear design (no offset) for the training data from the last exercise sheet (prediction of crime rate of U.S. cities) with $\lambda = 10^{-3}$. You should get 31 non-zero coefficients in the optimal weight vector.

   i. What is the influence of a feature with positive resp. negative component of the weight vector ?

   ii. Check the corresponding information on the features (see FeatureInformation file) to see if the features chosen by Lasso (corresponding to non-zero components in the weight vector) make actually sense for this problem

Both parts should be answered on paper or in a text file. The last question should be answered briefly (just check the 10 largest (in absolute value) components of the optimal weight vector).

Hints:

- A sum $f + g$ of convex functions f, g is strictly convex if f or g is strictly convex.
- For the computation of the projection, note that each component can be minimized independently of the other ones. Why ?

Exercise 9 - Derivation of a dual problem

Let $(x_i, y_i)_{i=1}^n$ be a training sample for a binary classification task, that is $x_i \in \mathbb{R}^d$ and $y_i \in \{-1, 1\}$. The so-called hard-margin Support Vector Machine (SVM) without offset corresponds to the optimization problem

$$\begin{aligned} \min_{w \in \mathbb{R}^d} \quad & \frac{1}{2} \|w\|_2^2 \\ \text{subject to}: \quad & y_i \langle w, x_i \rangle \geq 1, \qquad i = 1, \ldots, n \end{aligned}$$

a. **(3 Points)** Derive the dual problem.

b. **(1 Point)** Which problem, dual or primal, would you solve depending on n (number of training samples) versus d (number of features) ?

Hints:

- Note that inequality constraints have the form $g(x) \leq 0$

Submission instructions

- We accept both handwritten and electronic submissions. So you can choose what is more convenient for you. In any case, you should specify full names and immatriculation IDs of all team members. Obviously, programming tasks you can submit only electronically.
- Handwritten submissions should be submitted in the lecture hall of Monday's lecture (before the lecture starts).
- Electronic submissions should be emailed to the corresponding tutor:

  a. Apratim Bhattacharyya (Wednesday 8-10): abhattac@mpi-inf.mpg.de
  b. Maksym Andriushchenko (Thursday 8-10): s8mmandr@stud.uni-saarland.de
  c. Max Losch (Friday 16-18): mlosch@mpi-inf.mpg.de

  If not all 3 students belong to the same tutorial group, then you should email your submission to **only** one tutor (e.g. to the tutor of the first author of your homework), so please do not put other tutors in copy of the email.

  The email subject must have the following form: "[ML18/19 Exercise] Sheet X", where X is the number of the current exercise sheet. Then please specify in the email full names and immatriculation IDs of all team members. Then please attach all your files as a single zip archive, which consists of your immatriculation IDs, e.g. "2561234_2561235_2561236.zip".
- Reminder: you should submit in groups of 3. Otherwise, we will later on merge the groups smaller than 3 students.