Exercise 10 - Dual Coordinate Ascent for Linear SVM

In this exercise we implement coordinate ascent method for solving the dual of the soft-margin SVM problem. For simplicity, we restrict to the case where the offset b is fixed at zero. The dual problem in this case is given by

$$\max_{\alpha\in\mathbb{R}^n} \; \Psi(\alpha) \tag{1}$$
$$\text{subject to: } 0 \le \alpha_i \le \frac{C}{n}, \quad \forall i = 1,\ldots,n$$

where $\Psi(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1,j=1}^{n} \alpha_i\alpha_j y_i y_j \langle x_i, x_j\rangle$ and $x_i \in \mathbb{R}^d$ is the i^{th} training example and $y_i \in \{-1,+1\}$ is its label.

a. (**Bonus 2 Points**) Derive the dual formulation given above for the soft-margin SVM problem when the offset $b = 0$.

b. Coordinate descent/ascent is an optimization method which can be applied when the objective is differentiable and the optimization variables are not coupled by the constraints. This method solves a sequence of smaller subproblems where the objective is optimized over a single coordinate while fixing the values for the other coordinates. The coordinates for the subproblems are typically chosen in a cyclic order.

   The subproblem for the dual (1) over the coordinate r is given by

$$\alpha_r^{\text{new}} = \operatorname*{arg\,max}_{\alpha_r\in\mathbb{R}} \quad \Psi\Big(\alpha_1^{\text{old}},\ldots,\alpha_{r-1}^{\text{old}},\alpha_r,\alpha_{r+1}^{\text{old}}\ldots,\alpha_n^{\text{old}}\Big)$$
$$\text{subject to: } 0 \le \alpha_r \le \frac{C}{n}$$

   where all the variables except for α_r are fixed at previous values. Note that this is a quadratic problem in one variable α_r with an interval constraint: $\alpha_r \in [0, \frac{C}{n}]$.

   1. (**3 Points**) First derive the unconstrained solution $\bar{\alpha}_r$ of the above subproblem (i.e., ignoring the interval constraint). Next show that the solution in the presence of constraints is simply the projection of $\bar{\alpha}_r$ onto the interval $[0, \frac{C}{n}]$. That is

$$\alpha_r^{\text{new}} = \max\left\{0, \min\left\{\bar{\alpha}_r, \frac{C}{n}\right\}\right\}$$

   2. (**1 Point**) Show that the KKT conditions given in the lecture can be rewritten as the following:

$$\forall i = 1,\ldots,n:$$
$$\alpha_i = 0 \Rightarrow y_i\langle w, x_i\rangle \ge 1, \quad 0 < \alpha_i < \frac{C}{n} \Rightarrow y_i\langle w, x_i\rangle = 1, \quad \alpha_i = \frac{C}{n} \Rightarrow y_i\langle w, x_i\rangle \le 1$$

c. (**3 Points**) Complete the given Matlab function `CoordinateDescentSVM.m` which takes as arguments the training data `Xtrain` $\in \mathbb{R}^{n\times d}$, the class labels `ytrain` $\in \{-1,+1\}^n$, the error parameter C as well as the test data `Xtest`, `ytest` and returns the dual solution α, the primal solution w and training and test errors `TrainErrs`, `TestErrs` computed at each step of the coordinate descent method. Note that test data is only used to compute the test errors at intermediate steps! You have to fill-in the following blocks in the code labelled as `Fill-in`:

1. Compute the unconstrained solution $\bar{\alpha}_r$ and the constrained solution α_r^{new} of the subproblems
2. Compute the dual objective, training and test errors in each iteration from the current iterate $\alpha^k \in \mathbb{R}^n$
3. Implement stopping criteria: Check if the KKT conditions given above are satisfied by the current iterate α^k upto the given tolerance `EPS`.
4. Compute the primal solution w from the dual solution α

d. (**3 Points**) Train the SVM classifiers for different choices of the error parameter $C = \{10, 100, 200, 500\}$ on the given USPS digit dataset `DIGITS01`. This is a dataset of handwritten digits containing the digits 0 and 1. Load the file `DIGITS01`. The variables `Xtrain` and `Xtest` contain respectively the training and test digit data (each digit is an image of 16×16, so we have 256 gray values) and the variables `ytrain` and `ytest` contains the class labels (-1 for the digit 0 and 1 for the digit 1) for the training and test examples. Plot the training and test errors that you obtained in each iteration of the coordinate descent method and save them as `DIGITS01TrainErrs_C.png` and `DIGITS01TestErrs_C.png` for each choice of C.

Hints:

a. Avoid `for` loops in your implementation. All the computations that you need to fill-in can be expressed as matrix-vector multiplication. The operator `.*` (see `MATLAB` help for `times`) that computes element-by-element multiplication of vectors might be helpful here.

b. Note that the dual variable α and the primal variable w are related by

$$w = \sum_{i=1}^{n} y_i \alpha_i x_i,$$

where x_i is the i^{th} row of X. Recall that the prediction at a point x is given by $\text{sign}(\langle w, x \rangle)$ (since the offset $b = 0$).

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.