"Machine Learning and Computational Statistics"

9^{th} Homework (Part A)

Exercise 1:

Consider the set-up of the example given in slide 10 of the 9^{th} lecture. Verify that the node impurity decrease achieved by the rule "$x_1 \leq 3$" is equal to 0.42.

Exercise 2:

Wolfe dual representation: A convex programming problem is equivalent to

$$max_{\boldsymbol{\lambda} \geq \mathbf{0}} L(\boldsymbol{\theta}, \boldsymbol{\lambda})$$

$$\text{subject to } \frac{\partial}{\partial \boldsymbol{\theta}} L(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \mathbf{0}$$

Consider the SVM problem as it is stated in slide 31 of the 9^{th} lecture. Prove that its equivalent dual representation is the one shown in slide 32.

Hints: (a) The parameters in SVM are $\boldsymbol{\theta}$ and $\theta_{\mathbf{0}}$. Using the Karush-Kuhn-Tacker (KKT) conditions (1) and (2), derive the equations given at the beginning of the 32^{th} slide.

(b) Replace your findings to the Lagrangian function given in the 31^{th} slide and perform operations.

(c) Use the Wolfe dual representation given above to state the dual form of the SVM problem.

Exercise 3:

Consider the two-class two-dim. problem where class ω_1 (+1) consists of the vectors $\boldsymbol{x}_1$= $[-1, 1]^T$, $\boldsymbol{x}_2$= $[-1, -1]^T$, while class ω_2 (−1) consists of the vectors $\boldsymbol{x}_3$= $[1, -1]^T$, $\boldsymbol{x}_4$= $[1, 1]^T$.

(a) **Draw** the points and make a conjecture about the line the (linear) SVM classifier will return.
(b) **Using** the dual representation of the SVM problem, from ex. 1(c) derive
   (i) the Lagrange multipliers and
   (ii) the line that separates the data from the two classes.
(c) **Discuss** on the results.

Hints: 1. Defining y_1=+1, y_2=+1, y_3=-1, y_4=-1, substitute to the function

$$(\sum_{i=1}^{N} \lambda_i - \frac{1}{2} \sum_{ij} \lambda_i \lambda_j y_i y_j \boldsymbol{x}_i^T \boldsymbol{x}_j) \equiv J_1^*(\lambda)$$

y_i's and $\boldsymbol{x}_i$'s and express $J^*{}_1(\lambda)$ only in terms of λ_i's (keep in mind that the quantities $\boldsymbol{x}_i^T \boldsymbol{x}_j$ are scalars).

2. Taking the derivative of $J^*{}_1(\lambda)$ with respect to each λ_i and setting to zero, derive a system of equations for λ_i's and find ALL its solutions.

3. Determine the $\boldsymbol{\theta}$ vector, using the equations given in slide 32 of Lecture 9.

4. Determine the θ_0 parameter.