Machine Learning B

2024-2025

Home Assignment 2

Nirupam Gupta
Department of Computer Science
University of Copenhagen

The deadline for this assignment is **6 May 2025, 17:00**. You must submit your *individual* solution electronically via the Absalon home page.

A solution consists of:

- A PDF file with detailed answers to the questions, which may include graphs and tables if needed. Do *not* include your full source code in the PDF file, only selected lines if you are asked to do so.
- A .zip file with all your solution source code with comments about the major steps involved in each question (see below). Source code must be submitted in the original file format, not as PDF. The programming language of the course is Python.
- IMPORTANT: Do NOT zip the PDF file, since zipped files cannot be opened in *SpeedGrader*. Zipped PDF submissions will not be graded.
- Your PDF report should be self-sufficient. I.e., it should be possible to grade it without opening the .zip file. We do not guarantee opening the .zip file when grading.
- Your code should be structured such that there is one main file (or one main file per question) that we can run to reproduce all the results presented in your report. This main file can, if you like, call other files with functions, classes, etc.
- Handwritten solutions will not be accepted. Please use the provided latex template to write your report.

1 Convex geometry (25 points)

A set C is called a *cone* if for all $x \in C$ and $\theta \geq 0$, $\theta x \in C$. A point of the form $\sum_{i=1}^{n} \theta_i x_i$ with $\theta_1, \ldots, \theta_n \geq 0$ is called the *conic combination* of the set of points $S = (x_1, \ldots, x_n)$. We define *conic hull* of S to be

$$\mathsf{Cone}(S) = \left(\sum_{i=1}^{n} \theta_i x_i \,\middle|\, x_i \in S,\ \theta_i \geq 0,\ \forall i \right)$$

For an arbitrary positive integer $d \geq 1$ and an arbitrary set of n points $S = (x_1, \ldots, x_n)$ in $\mathbb{R}^d$, we have the following lemma.

Lemma 1. Let $n \geq d$. Any point $x \in \mathsf{Cone}(S)$ can be written as a conic combination of at most d points in S. That is, for all $x \in \mathsf{Cone}(S)$ there exists a set $\mathcal{I} \subseteq \{1, \ldots, n\}$ with $|S'| = d$ such that $x = \sum_{i \in \mathcal{I}} \theta_i \, x_i$ where $\theta_i \geq 0$, $\forall i \in \mathcal{I}$.

<u>Please Respond to the following two questions.</u>

Question 1 (15 pts). Consider an arbitrary positive integer $d \geq 1$ and an arbitrary set of n points $S = (x_1, \ldots, x_n)$ in $\mathbb{R}^d$ with $n > d$. Recall from the lectures that $\mathsf{Co}(S)$ denotes the *convex hull* of set S. Prove the following property:

> **(P1)** Any $x \in \mathsf{Co}(S)$ can be written as a convex combination of at most $d+1$ points in S.

Question 2 (10 pts). Does **(P1)** hold if we replace $d+1$ points with just d points? Give formal justification for your answer.

2 Convex functions (25 points)

Please respond to the following two questions, based on the lecture on April 30.

Question 3 (10 pts, equally distributed to both parts). Let $f : \mathbb{R}^d \to \mathbb{R}$ being a convex function.

(1) Prove that the *epigraph* of f, defined to be

$$\mathsf{Epi}(f) = \big((x, t) \in \mathbb{R}^d \times \mathbb{R} \,\big|\, f(x) \leq t \big),$$

is a convex set.

(2) Prove the converse, i.e., for a function $f : \mathbb{R}^d \to \mathbb{R}$, if its *epigraph* is convex then f is a convex function.

Question 4 (15 pts). Let $f : \mathbb{R}^d \to \mathbb{R}$ being a convex function. For $x \in \mathbb{R}^d$ and $t \in \mathbb{R}_{++}$ (where $\mathbb{R}_{++}$ denotes the set of strictly positive real-values). Prove that $g(x,t) = tf\left(\frac{x}{t}\right)$ is also convex. [**Hint:** use Question 3.]

[**Optional question**] Use this property to prove that relative entropy, i.e., KL-divergence between two probability distributions is a convex function.

3 Lagrange duality (25 points)

Please respond to the following two questions, based on the lecture on April 30.

Question 5 (10 pts). Derive the Lagrange dual problem of the following primal problem:

$$\begin{array}{ll} \underset{w \in \mathbb{R}^d}{\text{Minimize}} & c^\intercal w \\ \text{Subject to} & Aw = b \quad ; \ A \in \mathbb{R}^{p \times d} \ , \ b \in \mathbb{R}^p \\ & w \succeq 0 \end{array}$$

Question 6 (15 pts). Suppose the primal optimization problem is convex and has differentiable objective and constraint functions. Show that if there exists a solution to the KKT conditions, then *strong duality* holds for this optimization problem.

4 SVM (25 points)

Consider dataset $S = \{(x_1, y_1), \ldots, (x_n, y_n)\}$ such that $x_i \in \mathbb{R}^m$ and $y_i \in \{-1, +1\}$ for all i. Suppose the points are linearly separable. Consider the following alternative to the SVM problem formulation shown in the lecture.

Let $w = \sum_{i=1}^n \lambda_i y_i x_i$ where $\lambda_i \in \mathbb{R}$ for all i. Instead of maximizing the margin of separation, we aim to "sparsify" the above representation of w by minimizing the number of non-zero λ_i's. This can be approximately achieved by minimizing $\sum_{i=1}^n \lambda_i^2$. The resulting optimization problem is given by:

$$\begin{array}{ll} \underset{\lambda_1, \ldots, \lambda_n, b \,\in\, \mathbb{R}}{\text{Minimize}} & \frac{1}{2}\sum_{i=1}^n \lambda_i^2 \\ \text{Subject to} & 1 - y_i(\sum_{j=1}^n \lambda_j y_j x_j^\intercal x_i + b) \le 0 \quad , \ i = 1, \ldots, n \\ & \lambda_i \ge 0, \ i = 1, \ldots, n \end{array} \qquad \text{(Sparse SVM)}$$

Please respond to the following two questions.

Question 7 (10 points). Ignoring the additional constraint: $\lambda_i \geq 0$ for all i, show that (Sparse SVM) coincides with an instance of the original (Linear SVM) optimization problem that we studied in the lecture.

Question 8 (15 points). Derive the Lagrange dual problem of (Sparse SVM).