Machine Learning B (2025)
Home Assignment 2

Bar Segal xsb740

06/05/2025

Contents

1 Convex geometry (25 points)

Question 1. Carathéodory's theorem for conv$S$

Let $S = \{x_1, \dots, x_n\} \subset \mathbb{R}^d$ and $x \in \mathrm{conv}S$; so $x = \sum_{i=1}^n \lambda_i x_i$ with $\lambda_i \geq 0$ and $\sum_i \lambda_i = 1$. Choose such a representation with the *fewest* non-zero coefficients, and let $I = \{i \mid \lambda_i > 0\}$.

If $|I| \leq d+1$ we are done. Otherwise $|I| \geq d+2$, and the points $\{x_i\}_{i \in I} \subset \mathbb{R}^d$ are affinely dependent: there exist scalars α_i (not all zero) satisfying

$$\sum_{i \in I} \alpha_i x_i = 0, \qquad \sum_{i \in I} \alpha_i = 0.$$

Split I into $I^+ := \{i : \alpha_i > 0\}$ and $I^- := \{i : \alpha_i < 0\}$ (both non-empty) and set

$$\varepsilon := \min_{i \in I^+} \frac{\lambda_i}{\alpha_i} > 0, \qquad \lambda_i' := \lambda_i - \varepsilon \alpha_i.$$

Then $\lambda_i' \geq 0$ for every i and at least one λ_i' (the minimiser in I^+) equals 0, while

$$\sum_i \lambda_i' = \sum_i \lambda_i - \varepsilon \sum_i \alpha_i = 1.$$

Hence $x = \sum_i \lambda_i x_i = \sum_i \lambda_i' x_i$ is expressed with one fewer support point. Iterating this finite deletion process leaves at most $d+1$ points in the support. $\square$

Question 2. The bound $d+1$ is necessary

A convex combination of only d points lies in their affine span, whose dimension is at most $d-1$. Such combinations therefore cannot fill the interior of a d-simplex, the convex hull of $d+1$ affinely independent points.

Example in $\mathbb{R}^2$. Let $S = \{(0,0), (1,0), (0,1)\}$. The convex hull is the filled triangle. Combining any two of the vertices produces a point on an edge, never the interior point $(\frac{1}{3}, \frac{1}{3})$. Thus $d = 2$ still requires $d+1 = 3$ points.

2 Convex functions (25 points)

Question 3.1 Epigraph of a convex function is convex

Let $f : \mathbb{R}^d \to \mathbb{R}$ be convex and write

$$\mathrm{epi}(f) := \{(x,t) \in \mathbb{R}^d \times \mathbb{R} \mid f(x) \leq t\}.$$

Pick (x_1,t_1), $(x_2,t_2) \in \mathrm{epi}(f)$ and let $\lambda \in [0,1]$. Define $(x_\lambda, t_\lambda) = \lambda(x_1,t_1) + (1-\lambda)(x_2,t_2)$. Then, by convexity of f,

$$f(x_\lambda) \le \lambda f(x_1) + (1-\lambda) f(x_2) \le \lambda t_1 + (1-\lambda) t_2 = t_\lambda,$$

so $(x_\lambda, t_\lambda) \in \mathrm{epi}(f)$. Hence $\mathrm{epi}(f)$ is convex.

Question 3.2 Convex epigraph $\Longrightarrow$ convex function

Assume $\mathrm{epi}(f)$ is convex. For any $x_1, x_2 \in \mathbb{R}^d$ and $\lambda \in [0,1]$ the points $(x_1, f(x_1))$ and $(x_2, f(x_2))$ lie in $\mathrm{epi}(f)$; hence their convex combination

$$\big(\lambda x_1 + (1-\lambda)x_2,\ \lambda f(x_1) + (1-\lambda) f(x_2)\big) \in \mathrm{epi}(f).$$

Unpacking the definition of $\mathrm{epi}(f)$ gives the usual convexity inequality for f. Therefore f is convex.

Question 4. Convexity of the perspective $g(x,t) = t\,f(x/t)$

Let f be convex and set $g(x,t) = t\,f(x/t)$ on $\mathbb{R}^d \times \mathbb{R}_{++}$. Proving $\mathrm{epi}(g)$ is convex suffices.

Take $(x_i, t_i, u_i) \in \mathrm{epi}(g)$ for $i = 1,2$ and $\lambda \in [0,1]$; put $(\bar{x}, \bar{t}, \bar{u}) = \lambda(x_1,t_1,u_1) + (1-\lambda)(x_2,t_2,u_2)$ (with $\bar{t} > 0$).

Write $y_i = x_i/t_i,\ v_i = u_i/t_i$; then $(y_i, v_i) \in \mathrm{epi}(f)$. Because $\mathrm{epi}(f)$ is convex (parts 3.1–3.2),

$$\Big(\frac{\bar{x}}{\bar{t}},\ \frac{\bar{u}}{\bar{t}}\Big) \in \mathrm{epi}(f) \implies \frac{\bar{u}}{\bar{t}} \ge f\big(\bar{x}/\bar{t}\big).$$

Multiplying by $\bar{t}$ yields $\bar{u} \ge \bar{t}\,f(\bar{x}/\bar{t})$; hence $(\bar{x}, \bar{t}, \bar{u}) \in \mathrm{epi}(g)$. Thus *epig* is convex, and by 3.2 the perspective g itself is convex.

3 Lagrange duality (25 points)

question 5. Dual of a simple linear programme

Primal (standard form).

$$\begin{aligned} \min_{w \in \mathbb{R}^d} \quad & c^\top w \\ \text{s.t.} \quad & Aw = b,\ \ w \ge 0, \end{aligned} \qquad A \in \mathbb{R}^{p\times d},\ b \in \mathbb{R}^p,\ c \in \mathbb{R}^d.$$

Lagrangian. Introduce multipliers $\nu \in \mathbb{R}^p$ for the equality $Aw = b$ and $\lambda \in \mathbb{R}^d_{\ge 0}$ for the inequality $-w \le 0$:

$$\mathcal{L}(w,\nu,\lambda) = c^\top w + \nu^\top(Aw - b) - \lambda^\top w = (c + A^\top \nu - \lambda)^\top w - b^\top \nu.$$

Dual function. Taking $\inf_w \mathcal{L}$ gives

$$\phi(\nu, \lambda) = \begin{cases} -\, b^\top \nu, & c + A^\top \nu - \lambda = 0, \\ -\infty, & \text{otherwise.} \end{cases}$$

Because $\lambda \geq 0$, the stationarity condition can be written $A^\top \nu + c \geq 0$.

Dual problem. Eliminating λ yields the single-variable dual

$$\boxed{\begin{array}{ll} \max\limits_{\nu \in \mathbb{R}^p} & -\, b^\top \nu \\ \text{s.t.} & A^\top \nu + c \geq 0. \end{array}}$$

question 6. KKT point $\Longrightarrow$ strong duality

Consider the convex programme

$$\begin{array}{lll} \min\limits_{x \in \mathbb{R}^n} & f_0(x) & \\ \text{s.t.} & f_i(x) \leq 0 \ (i = 1{:}m), & h_j(x) = 0 \ (j = 1{:}p), \end{array}$$

with each f_0, f_i convex and differentiable and each h_j affine.

Lagrangian and dual.

$$\mathcal{L}(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{p} \nu_j h_j(x), \quad \lambda \geq 0, \ \phi(\lambda, \nu) = \inf_x \mathcal{L}(x, \lambda, \nu).$$

Suppose a KKT point exists. Let (x^*, λ^*, ν^*) satisfy

1. Primal feasibility: $f_i(x^*) \leq 0, \ h_j(x^*) = 0$.
2. Dual feasibility: $\lambda^* \geq 0$.
3. Complementary slackness: $\lambda_i^* f_i(x^*) = 0$.
4. Stationarity: $\nabla_x \mathcal{L}(x^*, \lambda^*, \nu^*) = 0$.

Items 4 and the definition of ϕ give $\phi(\lambda^*, \nu^*) = \mathcal{L}(x^*, \lambda^*, \nu^*)$. Items 1–3 collapse the right-hand side to $f_0(x^*)$. Hence

$$\phi(\lambda^*, \nu^*) = f_0(x^*).$$

Conclude strong duality. Weak duality always yields $\max_{\lambda\geq 0,\nu}\phi(\lambda,\nu)\leq \min_x f_0(x)$. Because the KKT point attains equality, both inequalities are tight:

$$\min_x f_0(x) = f_0(x^*) = \phi(\lambda^*,\nu^*) = \max_{\lambda\geq 0,\nu}\phi(\lambda,\nu).$$

Thus the optimal values of the primal and dual coincide; strong duality holds.

4 SVM (25 points)

question 7. Sparse–SVM equals a hard–margin SVM (when λ_i may be signed)

Ignoring the sign constraint $\lambda_i \geq 0$, the Sparse–SVM reads

$$\min_{\lambda\in\mathbb{R}^n,\, b\in\mathbb{R}} \frac{1}{2}\|\lambda\|_2^2 \quad \text{s.t.} \quad y_i\Big(\sum_{j=1}^n \lambda_j y_j x_j^\top x_i + b\Big) \geq 1,\ i = 1{:}n. \tag{$\star$}$$

Define the (data–dependent) feature map

$$\varphi : \mathbb{R}^m \to \mathbb{R}^n, \qquad \varphi(x) := \big(y_1 x_1^\top x, \dots, y_n x_n^\top x\big)^\top,$$

and set $w := \lambda \in \mathbb{R}^n$. Then for every sample x_i, $y_i\big(w^\top\varphi(x_i) + b\big) \geq 1$, while the objective is $\frac{1}{2}\|w\|_2^2$. Hence ($\star$) is nothing but the standard hard–margin SVM applied to the transformed data $\{(\varphi(x_i), y_i)\}_{i=1}^n$. So, *dropping the non-negativity of λ converts the Sparse–SVM into an ordinary linear SVM in feature space.*

Question 8. Dual of the Sparse–SVM (with $\lambda_i \geq 0$ kept)

Primal formulation.

$$\begin{aligned} \min_{\lambda\in\mathbb{R}^n_{\geq 0},\, b\in\mathbb{R}} \quad & \frac{1}{2}\|\lambda\|_2^2 \\ \text{s.t.} \quad & g_i(\lambda, b) := 1 - y_i\Big(\sum_{j=1}^n \lambda_j y_j x_j^\top x_i + b\Big) \leq 0 \quad (i = 1{:}n). \end{aligned} \tag{P}$$

Lagrangian. With multipliers $\alpha_i \geq 0$ for $g_i \leq 0$ and $\mu_i \geq 0$ for $-\lambda_i \leq 0$:

$$\mathcal{L}(\lambda, b, \alpha, \mu) = \tfrac{1}{2}\|\lambda\|^2 + \sum_{i=1}^n \alpha_i - \lambda^\top(H\alpha + \mu) - b\sum_{i=1}^n \alpha_i y_i,$$

where $H_{ij} = y_i y_j x_i^\top x_j$ and $H\alpha := h$.

Dual variables.

- **Bias:** stationarity in b: $\sum_{i=1}^{n} \alpha_i y_i = 0$.
- **Weights:** minimise $\frac{1}{2}\|\lambda\|^2 - \lambda^\top(h+\mu)$ over $\lambda \geq 0$. The solution is the projection of $h+\mu$ onto the non-negative orthant:

  $$\lambda^\star = \big[h+\mu\big]_+.$$

  Taking $\mu^\star = \max\{0, -h\}$ gives $\lambda^\star = \max\{0, h\}$ and complementary slackness $\lambda^\star \odot \mu^\star = 0$.

Dual objective. Insert $\lambda^\star$ into $\mathcal{L}$:

$$\phi(\alpha) = \sum_{i=1}^{n} \alpha_i \; - \; \frac{1}{2} \big\|\max\{0, H\alpha\}\big\|_2^2.$$

Dual problem.

$$\boxed{\begin{array}{lll} \max\limits_{\alpha \in \mathbb{R}^n} & \displaystyle\sum_{i=1}^{n} \alpha_i - \frac{1}{2}\big\|\max\{0, H\alpha\}\big\|_2^2 & \\ \text{s.t.} & \displaystyle\sum_{i=1}^{n} \alpha_i y_i = 0, & \alpha_i \geq 0 \; (i = 1{:}n). \end{array}}$$

Because $H \succeq 0$, the objective is concave, so this is a convex maximisation problem. Keeping the constraints $\lambda_i \geq 0$ manifests through the $\max\{0, \cdot\}$ term; dropping them would replace that norm by $\|H\alpha\|_2^2$ and enlarge the dual feasible set.

Thank you.