Homework #4 - Machine Learning

Srikanth Pai, Asst. Professor, MSE

February 1, 2026

Problem Set

1. Let $C \subset \mathbb{R}^{n+m}$ be a convex set. Define the projection onto the first n coordinates by

$$\pi(C) = \{x \in \mathbb{R}^n : \exists y \in \mathbb{R}^m \text{ such that } (x, y) \in C\}.$$

   Prove that $\pi(C)$ is convex.

2. Let $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ be convex, and define

$$\phi(x) := \inf_{y \in \mathbb{R}^m} F(x, y).$$

   (a) Show that

$$\operatorname{epi}(\phi) = \pi\big(\operatorname{epi}(F)\big),$$

   where π denotes projection onto the (x, t) coordinates.

   (b) Using the previous problem, prove that ϕ is convex.

3. Consider the optimisation problem

$$\min_{\beta, \sigma \geq 0} \quad f(\beta) + \frac{\lambda}{2} \sum_{i=1}^{n} g(\beta_i, \sigma_i),$$

   where $f : \mathbb{R}^n \to \mathbb{R}$ is convex, $\lambda \geq 0$, and

$$g(x, y) = \begin{cases} \dfrac{x^2}{y} + y, & y > 0, \\ 0, & x = 0,\ y = 0, \\ +\infty, & \text{otherwise.} \end{cases}$$

   (a) Show that the objective function is convex.

   (b) For fixed $x \in \mathbb{R}$, compute

$$\inf_{y \geq 0} \left(\frac{x^2}{y} + y \right).$$

   (c) Using the previous parts, show that minimising over $\sigma \geq 0$ yields

$$\min_{\beta \in \mathbb{R}^n} \quad f(\beta) + \lambda \|\beta\|_1.$$

4. Let $v \in \mathbb{R}^n$ and consider the optimisation problem

$$\min_{x \in \mathbb{R}^n} \quad \frac{1}{2}\|x - v\|_2^2 \quad \text{subject to} \quad \|x\|_1 \leq 1.$$

(a) Write down the Lagrangian for this problem.

(b) Define the dual function

$$q(\lambda) := \inf_{x\in\mathbb{R}^n} L(x,\lambda), \qquad \lambda \geq 0,$$

and show that the minimiser in the definition of $q(\lambda)$ satisfies

$$x_i(\lambda) = \operatorname{sign}(v_i)\,(|v_i| - \lambda)_+.$$

(c) Using complementary slackness, show that if $\|v\|_1 \leq 1$, then $x^\star = v$ is optimal.

(d) Otherwise, show that the optimal multiplier $\lambda^\star$ satisfies

$$\sum_{i=1}^{n}(|v_i| - \lambda^\star)_+ = 1,$$

and hence characterise the projection $x^\star$.

5. Consider the quadratic optimisation problem

$$\min_{x\in\mathbb{R}^n} \quad \frac{1}{2}x^\top Qx + c^\top x \quad \text{subject to} \quad Ax = b,$$

where Q is symmetric positive definite and A has full row rank.

(a) Write down the Lagrangian and derive the KKT conditions.

(b) Show that the optimal primal–dual pair $(x^\star, \lambda^\star)$ satisfies

$$\begin{pmatrix} Q & A^\top \\ A & 0 \end{pmatrix}\begin{pmatrix} x^\star \\ \lambda^\star \end{pmatrix} = \begin{pmatrix} -c \\ b \end{pmatrix}.$$

(c) Explain why this system has a unique solution.

6. Let $X_1, \dots, X_n$ be i.i.d. random variables with distribution

$$X_i \sim \mathcal{N}(\mu, \sigma^2),$$

where μ is known and $\sigma^2 > 0$ is unknown.

(a) Write down the likelihood function $L(\sigma^2)$.

(b) Show that the likelihood can be written in the form

$$L(\sigma^2) \propto (\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right).$$

7. Suppose the prior density of σ^2 is of the form

$$p(\sigma^2) \propto (\sigma^2)^{-(\alpha+1)} \exp\left(-\frac{\beta}{\sigma^2}\right), \qquad \sigma^2 > 0,$$

for some $\alpha, \beta > 0$.

(a) Show that the posterior density of σ^2 has the same functional form.

(b) Identify the updated parameters of the posterior distribution.

(c) Conclude that this family of distributions is conjugate to the Gaussian likelihood.

8. Let $X_1, \dots, X_n$ be i.i.d. with

$$X_i \mid \lambda \sim \text{Poisson}(\lambda), \qquad \lambda > 0.$$

Assume the prior density

$$p(\lambda) \propto \lambda^{\alpha-1} e^{-\beta\lambda}, \qquad \lambda > 0.$$

Show that the posterior distribution of λ has the same functional form and determine its parameters.

9. Let $X_1, \dots, X_n$ be i.i.d. with

$$X_i \mid \mu \sim \mathcal{N}(\mu, \sigma^2),$$

where σ^2 is known.

Assume the prior

$$\mu \sim \mathcal{N}(\mu_0, \tau^2).$$

(a) Derive the posterior distribution of μ.

(b) Compute the maximum a posteriori (MAP) estimator of μ.

(c) Show that the MAP estimator minimises

$$\sum_{i=1}^{n} (x_i - \mu)^2 + \frac{\sigma^2}{\tau^2} (\mu - \mu_0)^2.$$

10. Let (x_i, y_i) for $i = 1, \dots, n$ be fixed data with

$$Y_i \mid \beta \sim \mathcal{N}(x_i^\top \beta, \sigma^2),$$

where σ^2 is known.

Assume the prior

$$\beta \sim \mathcal{N}(0, \tau^2 I_d).$$

Show that the MAP estimator of β minimises

$$\sum_{i=1}^{n} (y_i - x_i^\top \beta)^2 + \frac{\sigma^2}{\tau^2} \|\beta\|_2^2.$$