Optimisation Basics 2 — Practice Questions

CM52054: Foundational Machine Learning
Practice set with fully worked answers

Part A — Concept & Short-Answer

1) **Convex sets and convex functions.**
(a) Define a convex set.
(b) Define a convex function.
(c) Why is convexity useful in optimisation?

Answer:

a) A set C is convex if for any $x, y \in C$ and $\beta \in [0, 1]$:
$$(1 - \beta)x + \beta y \in C.$$

b) A function f is convex if for all x, y in its domain and $\beta \in [0, 1]$:
$$f((1 - \beta)x + \beta y) \leq (1 - \beta)f(x) + \beta f(y).$$

c) Convexity guarantees that any local minimiser is also a global minimiser.

2) **Linear least-squares formulation.**
Write the minimisation problem for fitting the linear model $f(x) = w^\top x$.

Answer:
$$w^* = \arg\min_w \sum_{i=1}^{N} (w^\top x_i - y_i)^2 = \arg\min_w \|Xw - y\|^2.$$

3) **Normal equation (outline derivation)**
Show that minimising $\|Xw - y\|^2$ leads to:
$$w^* = (X^\top X)^{-1} X^\top y.$$

Answer:
$$\nabla_w \|Xw - y\|^2 = 2X^\top Xw - 2X^\top y = 0 \Rightarrow X^\top Xw = X^\top y \Rightarrow w^* = (X^\top X)^{-1} X^\top y.$$

4) **Gradient of the squared error.**
Compute $\nabla_w L(w)$ for $L(w) = \|Xw - y\|^2$.

Answer:
$$\nabla_w L(w) = 2X^\top Xw - 2X^\top y.$$

5) **Gradient descent update rule.**
Write the update rule and specialise it for this loss.

Answer:
Generic:
$$w_{t+1} = w_t - \alpha \nabla L(w_t).$$
Least-squares form:
$$w_{t+1} = w_t - \alpha (2X^\top X w_t - 2X^\top y).$$

Part B — Algorithms & Variants

6) **Compare GD variants.**
Give one advantage and one disadvantage for each:

Method	Advantage	Disadvantage
Batch GD		
Stochastic GD		
Mini-batch GD		

Answer:

- Batch GD — stable direction; slow on large datasets.
- SGD — fast, can escape minima; noisy and unstable.
- Mini-batch — efficient and stable; batch size requires tuning.

7) **Termination criterion.**
State one stopping rule.

Answer: A common rule:

$$\|\alpha \nabla f(w_t)\| < \varepsilon.$$

Part C — More Challenging Questions

8) **Convexity and uniqueness of the least-squares solution.**
Consider the least-squares objective

$$L(w) = \|Xw - y\|^2.$$

(a) Show explicitly that $L(w)$ is a convex function of w by writing it as a quadratic form, i.e. second-degree polynomial in the variable w.
(b) Compute the second-order gradient $\nabla_w^2 L(w)$.
(c) **Optional:** Give a condition on X under which the optimal w^* is unique.

Answer:

a) Expand:

$$L(w) = (Xw - y)^\top (Xw - y) = w^\top X^\top X w - 2y^\top X w + y^\top y.$$

This is a quadratic function in w with quadratic term $w^\top X^\top X w$. Since $X^\top X$ is positive semi-definite (for any z, $z^\top X^\top X z = \|Xz\|^2 \geq 0$), $L(w)$ is convex.

b) Differentiating twice w.r.t. w:

$$\nabla_w L(w) = 2X^\top X w - 2X^\top y,$$

so

$$\nabla_w^2 L(w) = 2X^\top X.$$

c) The optimal solution is unique if and only if the objective is *strictly* convex. This occurs when $X^\top X$ is positive definite, i.e. when the columns of X are linearly independent (full column rank). In that case, $X^\top X$ is invertible and there is a unique w^* solving $X^\top X w^* = X^\top y$.