Optimisation Basics 3 — Coding Practice Questions

CM52054: Foundational Machine Learning
Practice set with fully worked answers

Newton's Method

1) **Newton's method for linear regression.**
Generate a synthetic 1D dataset from the model

$$y = 3x + 2 + \text{noise},$$

For a squared-error function $g(w) = \|Xw - y\|^2$, implement Newton's method:

$$w_{t+1} = w_t - [\nabla^2 g(w_t)]^{-1} \nabla g(w_t),$$

where

$$\nabla g(w) = 2X^\top (Xw - y), \qquad \nabla^2 g(w) = 2X^\top X.$$

Start from any initial guess and show that one Newton step recovers the closed-form solution up to numerical precision.

2) **Visualising gradient descent vs Newton on an anisotropic quadratic.**
Consider the 2D quadratic function

$$f(\mathbf{x}) = \tfrac{1}{2}\mathbf{x}^\top A \mathbf{x}, \quad A = \begin{bmatrix} 5 & 0 \\ 0 & 0.5 \end{bmatrix}.$$

Implement both gradient descent and Newton's method starting from $\mathbf{x}_0 = (6, 6)$ and plot their trajectories on the same contour plot to show that gradient descent zig-zags while Newton's method moves directly to the minimum.

3) **Optional: Numerical gradient & Hessian checker.**
For the function

$$f(x, y) = x^3 + xy^2,$$

(a) implement analytical gradient and Hessian;
(b) implement numerical approximations of gradient and Hessian using central finite differences;
(c) evaluate both at a few random points and print the differences to verify that your analytical derivatives are correct.

Analytical derivatives:

$$\frac{\partial f}{\partial x} = 3x^2 + y^2, \quad \frac{\partial f}{\partial y} = 2xy,$$

$$\nabla^2 f = \begin{bmatrix} 6x & 2y \\ 2y & 2x \end{bmatrix}.$$