Homework 1

BAIM Mohamed Jalal
Machine Learning

September 22, 2024

1 Part 1

1.1 Closed Form LSE

Given:

- A target vector $y \in \mathbb{R}^n$, where n is the number of observations.
- A design matrix $X \in \mathbb{R}^{n \times p}$.
- A parameter vector $a \in \mathbb{R}^p$ that we wish to estimate.

We define the predicted response $h(x)$ as:

$$h(x) = Xa.$$

Our objective is to minimize the objective function $J(a)$:

$$J(a) = \sum_{i=1}^{n} (y_i - h(x_i))^2 = \|y - Xa\|^2$$

$$J(a) = (y - Xa)^\top (y - Xa)$$

$$\begin{aligned} J(a) &= (y - Xa)^\top (y - Xa) \\ &= y^\top y - y^\top Xa - a^\top X^\top y + a^\top X^\top Xa. \end{aligned}$$

Since both $y^\top Xa$ and $a^\top X^\top y$ are scalars they are equivalent so:

$$J(a) = y^\top y - 2a^\top X^\top y + a^\top X^\top Xa.$$

Let's find the parameter vector a that minimizes $J(a)$:

$$\frac{\partial J(a)}{\partial a} = -2X^\top y + 2X^\top Xa = 0.$$

$$X^\top Xa = X^\top y.$$

Finally :

$$a = (X^\top X)^{-1} X^\top y.$$

1.2 Steepest Descent Method

We are given a function $f : \mathbb{R}^n \to \mathbb{R}$, and our goal is to find x^* such that $f(x^*)$ is minimized.

Starting from an initial point x_1, we iteratively update the solution by moving in the direction of the gradient. The update rule is given by:

$$x_{n+1} = x_n + \alpha_n d_n,$$

where $d_n = -\nabla f(x_n)$ is the negative of the gradient at the current point.

Applying Steepest Descent to LSE:
We aim to minimize the following objective function for the LSE function:

$$J(a) = \|y - Xa\|^2.$$

$$J(a) = \left(y^T y - 2y^T Xa + a^T X^T Xa\right).$$

Taking the gradient with respect to a:

$$\nabla_a J(a) = -2X^T y + 2X^T Xa = -2X^T (y - Xa)$$

We can now apply the steepest descent update rule:

$$a_{k+1} = a_k - 2\alpha_k \left(X^T (Xa_k - y)\right).$$

1.3 Newton's Method

We aim to minimize the least squares objective function:

$$J(a) = \|y - Xa\|^2.$$

The gradient of $J(a)$ is:

$$\nabla_a J(a) = -X^T y + X^T Xa = X^T (Xa - y)$$

and the Hessian matrix is:

$$H(a) = X^T X$$

The update rule for Newton's method is:

$$a_{k+1} = a_k - H(a_k)^{-1} \nabla_a f(a_k),$$

which simplifies to:

$$a_{k+1} = a_k + (X^T X)^{-1} X^T (y - Xa_k).$$