AI 512 Part 1: Assignment -1 (10 Marks)

Q1 *Nearest point to a line.* Let a and b be different n-vectors. The line passing through a and b is given by the set of vectors of the form $(1-\theta)a + \theta b$, where θ is a scalar that determines the particular point on the line.

Let x be any n-vector. Find a formula for the point p on the line that is closest to x. The point p is called the *projection* of x onto the line. Show that $(p - x) \perp (a - b)$, and draw a simple picture illustrating this in 2-D. *Hint.* Work with the square of the distance between a point on the line and x, *i.e.*, $\|(1-\theta)a + \theta b - x\|^2$. Expand this, and minimize over θ. **2 Marks**

Q2 *Distance versus angle nearest neighbor.* Suppose $z_1, \ldots, z_m$ is a collection of n-vectors, and x is another n-vector. The vector z_j is the (distance) nearest neighbor of x (among the given vectors) if

$$\|x - z_j\| \leq \|x - z_i\|, \quad i = 1, \ldots, m,$$

i.e., x has smallest distance to z_j. We say that z_j is the *angle nearest neighbor* of x if

$$\angle(x, z_j) \leq \angle(x, z_i), \quad i = 1, \ldots, m,$$

i.e., x has smallest angle to z_j.

(a) Give a simple specific numerical example where the (distance) nearest neighbor is not the same as the angle nearest neighbor. **1 Mark**

(b) Now suppose that the vectors $z_1, \ldots, z_m$ are normalized, which means that $\|z_i\| = 1$, $i = 1, \ldots, m$. Show that in this case the distance nearest neighbor and the angle nearest neighbor are always the same. *Hint.* You can use the fact that arccos is a decreasing function, *i.e.*, for any u and v with $-1 \leq u < v \leq 1$, we have $\arccos(u) > \arccos(v)$. **2 Marks**

Q3

Iterative method for least squares problem. Suppose that A has linearly independent columns, so $\hat{x} = A^\dagger b$ minimizes $\|Ax - b\|^2$. In this exercise we explore an iterative method, due to the mathematician Lewis Richardson, that can be used to compute $\hat{x}$. We define $x^{(1)} = 0$ and for $k = 1, 2, \ldots,$

$$x^{(k+1)} = x^{(k)} - \mu A^T (Ax^{(k)} - b),$$

where μ is a positive parameter, and the superscripts denote the iteration number. This defines a sequence of vectors that converge to $\hat{x}$ provided μ is not too large; the choice $\mu = 1/\|A\|^2$, for example, always works. The iteration is terminated when $A^T(Ax^{(k)} - b)$ is small enough, which means the least squares optimality conditions are almost satisfied. To implement the method we only need to multiply vectors by A and by A^T. If we have efficient methods for carrying out these two matrix-vector multiplications, this iterative method can be faster than **QR Algorithm** (although it does not give the exact solution). Iterative methods are often used for very large scale least squares problems.

(a) Show that if $x^{(k+1)} = x^{(k)}$, we have $x^{(k)} = \hat{x}$. **2 Marks**

(b) Generate a random 20×10 matrix A and 20-vector b, and compute $\hat{x} = A^\dagger b$. Run the Richardson algorithm with $\mu = 1/\|A\|^2$ for 500 iterations, and plot $\|x^{(k)} - \hat{x}\|$ to verify that $x^{(k)}$ appears to be converging to $\hat{x}$. **3 Marks**

(Code in python, use the pseudo inverse formula to compute $\hat{x}$ .Include the python code separately along with the solution report)