Probability and Computation: Problem Sheet 6 Solutions

Question 1. *Let $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ be any vector. Now consider a random vector $y = (y_1, y_2, \dots, y_n) \in \mathbb{R}^n$, where each y_i, $1 \le i \le n$ is in $\{-1, +1\}$ uniformly at random.*

1. *Compute $\mathbf{E}\left[\, x^T y \,\right]$ and $\mathbf{E}\left[\, (x^T y)^2 \,\right]$.*
2. *We are now trying to generalise these results. Which properties of the distribution the y_i's are sampled from are needed to recover the results from Part 1?*
3. *What is the challenge in proving concentration for $(x^T y)^2$, for example, why cannot we apply a Chernoff Bound?*

Question 2. *(easy) What is the time complexity for applying the Random Projection Method to a set of P input vectors?*

Solution: We first need to construct the random $M \times N$-matrix, which takes $O(MN)$ time. Then we need to multiply the random matrix to each of the P input vectors, each of which takes $O(MN)$ time, amounting to a total time of $O(MNP)$.

Question 3. *Demonstrate that it is essential for the Random Projection Method to choose the linear function f randomly. Specifically, prove that for any linear function $f : \mathbb{R}^N \to \mathbb{R}^M$ there are two vectors $x_1, x_2 \in \mathbb{R}^N$, $x_1 \neq x_2$, such that $f(x_1) = f(x_2)$.*
Hint: *Your proof should at some point exploit that $N > M$.*

Solution: As in the description of the Random Projection Method, the linear function $f : \mathbb{R}^N \to \mathbb{R}^M$ applied to any vector $w \in \mathbb{R}^N$ is the same as the matrix-vector product

$$f(w) = R \cdot w,$$

where R is a $M \times N$ matrix. Or written more explicitly,

$$f \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ \vdots \\ w_N \end{pmatrix} = \begin{pmatrix} \vdots & \vdots & & \vdots \\ r_1 & r_2 & \cdots & r_N \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \end{pmatrix} \cdot \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ \vdots \\ w_N \end{pmatrix},$$

where $\{r_1, r_2, \dots, r_N\}$ is a set of M-dimensional vectors. Since $N > M$ and the due to the fact that N is the largest number of linearly independent vectors, it follows that the set of vectors $\{r_1, r_2, \dots, r_N\}$ is not linearly independent. Thus there are coefficients $\alpha_1, \alpha_2, \dots, \alpha_N$, not all being equal to zero, such that

$$\sum_{j=1}^{N} \alpha_j \cdot r_j = \vec{0}.$$

Thus for every entry $1 \le i \le M$,

$$\sum_{i=1}^{N} \alpha_j \cdot r_{i,j} = \sum_{i=1}^{N} r_{i,j} \cdot \alpha_j = 0.$$

This implies that

$$f\begin{pmatrix}\alpha_1\\ \alpha_2\\ \vdots\\ \vdots\\ \alpha_N\end{pmatrix} = \begin{pmatrix}0\\0\\ \vdots\\0\end{pmatrix}.$$

Based on this finding, it is straightforward to pick two vectors $x_1 \in \mathbb{R}^N$ and $x_2 \in \mathbb{R}^N$ such that $x_1 \neq x_2$ but $f(x_1) = f(x_2)$. Let $x_1 \in \mathbb{R}^n$ be arbitrary and $x_2 = x_1 + \alpha$. Then, thanks to the linearity of f,

$$f(x_2) = f(x_1 + \alpha) = f(x_1) + f(\alpha) = f(x_1).$$

Question 4. *Consider the deterministic weighted majority algorithm. [LECTURE 15]*

1. *Prove or disprove that the algorithm always makes a number of mistakes M^T within T rounds that satisfies: $M^T \geq \min_{i\in[n]} m_i^T$.*
2. *Show that for any deterministic algorithm, there exists an input such that $M^T \geq 2 \cdot \min_{i\in[n]} m_i^T$.*