1 Linear Algebra Practice

Problem 1. (VMLS 1.10 Total score.) The record for each student in a class is given as a 10vector r, where $r_1, \ldots, r_8$ are the grades for the 8 homework assignments, each on a $0-10$ scale, r_9 is the midterm exam grade on a $0-120$ scale, and r_{10} is the final exam score on a $0-160$ scale. The student's total course score s, on a $0-100$ scale, is based 25% on the homework, 35% on the midterm exam, and 40% on the final exam. Express s in the form $s = w^T r$. (That is, determine the 10-vector w.) You can give the coefficients of w to 4 digits after the decimal point.

Problem 2. (VMLS 2.1 Linear or not?.) Determine whether each of the following scalar-valued functions of n vectors is linear. If it is a linear function, give its inner product representation, i.e., an n-vector a for which $f(x) = a^T x$ for all x. If it is not linear, give specific x, y, α, and β for which superposition fails, i.e.,

$$f(\alpha x + \beta y) \neq \alpha f(x) + \beta f(y).$$

(a) The spread of values of the vector, defined as $f(x) = \max_k x_k - \min_k x_k$.

(b) The difference of the last element and the first, $f(x) = x_n - x_1$.

(c) The median of an n-vector, where we will assume $n = 2k + 1$ is odd. The median of the vector x is defined as the $(k+1)$ st largest number among the entries of x. For example, the median of $(-7.1, 3.2, -1.5)$ is -1.5.

(d) The average of the entries with odd indices, minus the average of the entries with even indices. You can assume that $n = 2k$ is even.

(e) Vector extrapolation, defined as $x_n + (x_n - x_{n-1})$, for $n \geq 2$. (This is a simple prediction of what x_{n+1} would be, based on a straight

Problem 3. (VMLS 2.9 Taylor approximation.) Consider the function $f : \mathbf{R}^2 \rightarrow \mathbf{R}$ given by $f(x_1, x_2) = x_1 x_2$. Find the Taylor approximation $\hat{f}$ at the point $z = (1,1)$. Compare $f(x)$ and $\hat{f}(x)$ for the following values of x :

$$x = (1,1), \quad x = (1.05, 0.95), \quad x = (0.85, 1.25), \quad x = (-1, 2).$$

Make a brief comment about the accuracy of the Taylor approximation in each case.

Problem 4. (VMLS 3.5 Other common norms.) Any real-valued function f that satisfies the four properties given on page 46 (nonnegative homogeneity, triangle inequality, nonnegativity, and definiteness) is called a vector norm, and is usually written as $f(x) = \|x\|_{\mathrm{mn}}$, where the subscript is some kind of identifier or mnemonic to identify it. The most commonly used norm is the one we use in this book, the Euclidean norm, which is sometimes written with the subscript 2, as $\|x\|_2$. Two other common vector norms for n-vectors are the 1-norm $\|x\|_1$ and the ∞-norm $\|x\|_\infty$, defined as

$$\|x\|_1 = |x_1| + \cdots + |x_n|, \quad \|x\|_\infty = \max\{|x_1|, \ldots, |x_n|\}.$$

These norms are the sum and the maximum of the absolute values of the entries in the vector, respectively. The 1-norm and the ∞-norm arise in some recent and advanced applications, but we will not encounter them in this book. Verify that the 1-norm and the ∞-norm satisfy the four norm properties

Problem 5. (VMLS 3.27 A new measure of deviation.) The standard deviation is a measure of how much the entries of a vector differ from their mean value. Another measure of how much the entries of an n-vector x differ from each other, called the mean square difference, is defined as

$$\operatorname{MSD}(x)=\frac{1}{n^2}\sum_{i,j=1}^{n}\left(x_i-x_j\right)^2$$

(The sum means that you should add up the n^2 terms, as the indices i and j each range from 1 to n.) Show that $\operatorname{MSD}(x)=2\operatorname{std}(x)^2$. Hint. First observe that $\operatorname{MSD}(\tilde{x})=\operatorname{MSD}(x)$, where $\tilde{x}=x-\operatorname{avg}(x)1$ is the de-meaned vector. Expand the sum and recall that $\sum_{i=1}^{n}\tilde{x}_i=0$.

Problem 6. (VMLS 4.2 k-means with specialized vectors.) Suppose that the vectors $x_1,\ldots,x_N$ are clustered using k-means, with group representatives $z_1,\ldots,z_k$.

(a) Suppose the original vectors x_i are nonnegative, i.e., their entries are nonnegative. Explain why the representatives z_j are also nonnegative.

(b) Suppose the original vectors x_i represent proportions, i.e., their entries are nonnegative and sum to one. (This is the case when x_i are word count histograms, for example.) Explain why the representatives z_j also represent proportions, i.e., their entries are nonnegative and sum to one.

(c) Suppose the original vectors x_i are Boolean, i.e., their entries are either 0 or 1 . Give an interpretation of $\left(z_j\right)_i$, the i th entry of the j group representative.

Problem 7. (VMLS 5.2 Linear dependence.) An intern at a quantitative hedge fund examines the daily returns of around 400 stocks over one year (which has 250 trading days). She tells her supervisor that she has discovered that the returns of one of the stocks, Google (GOOG), can be expressed as a linear combination of the others, which include many stocks that are unrelated to Google (say, in a different type of business or sector). Her supervisor then says: "It is overwhelmingly unlikely that a linear combination of the returns of unrelated companies can reproduce the daily return of GOOG. So you've made a mistake in your calculations." Is the supervisor right? Did the intern make a mistake? Give a very brief explanation.

Problem 8. (VMLS 5.6 Running GS.) Consider the list of nn-vectors

$$a_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad a_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \dots, \quad a_n = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

(The vector a_i has its first i entries equal to one, and the remaining entries zero.) Describe what happens when you run the Gram-Schmidt algorithm on this list of vectors, i.e., say what $q_1, \dots, q_n$ are. Is $a_1, \dots, a_n$ a basis?

2 Implementation

Problem 9. (Implementing Lloyd's Algorithm.)
Recall LLoyd's (k-means) algorithm to find clusters in a dataset. In this exercise, we will implement the algorithm on a simple synthetic dataset.

(a) Using the python `numpy.random` package, sample 100 points in $\mathbb{R}^2$ from a normal distribution centered at $(0, 0)$ with standard deviation 0.3.

(b) Repeat the above with centers at $(1, 1)$ and $(-1, 1)$ and same standard deviation as above. Collect all 300 points together into a list called `X`.

(c) Given a list of vectors `data` and a list of k centroids `centroids`, write a function `group_assignment()` that assigns each vector to the centroid closest to it.

(d) Using the provided functions `update_centroid()` and `clustering_objective()`, implement the k-means algorithm which terminates when distance the chosen centroids at successive iterations is less than 10^{-6}.