Homework 2

1. Distance from Palo Alto to Beijing. The surface of the earth is reasonably approximated as a sphere with radius $R = 6367.5$km. A location on the earth’s surface is traditionally given by its latitude $\theta$ and its longitude $\lambda$, which correspond to angular distance from the equator and prime meridian, respectively. The 3-D coordinates of the location are given by

$$R(\sin \lambda \cos \theta, \cos \lambda \cos \theta, \sin \theta).$$

(In this coordinate system $(0, 0, 0)$ is the center of the earth, $R(0, 0, 1)$ is the North pole, and $R(0, 1, 0)$ is the point on the equator on the prime meridian, due south of the Royal Observatory outside London. And no, you don’t need to know any of this.)

The distance through the earth between two locations (3-vectors) $a$ and $b$ is $\|a - b\|$. The distance along the surface of the earth between points $a$ and $b$ is $R_\odot(a, b)$. Find these two distances between Palo Alto and Beijing, with latitudes and longitudes given below.

<table>
<thead>
<tr>
<th>City</th>
<th>Latitude $\theta$</th>
<th>Longitude $\lambda$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beijing</td>
<td>39.914°</td>
<td>116.392°</td>
</tr>
<tr>
<td>Palo Alto</td>
<td>37.429°</td>
<td>-122.138°</td>
</tr>
</tbody>
</table>

Hint. In Julia, the functions $\cos$ and $\sin$ compute $\cos$ and $\sin$ taking the argument in radians. Julia also has the functions $\cosd$ and $\sind$ which take the argument in degrees.

2. Difference of squared distances. Determine whether the difference of the squared distances to two fixed vectors $c$ and $d$, defined as

$$f(x) = \|x - c\|^2 - \|x - d\|^2,$$

is linear, affine, or neither. If it is linear, give its inner product representation, i.e., an $n$-vector $a$ for which $f(x) = a^T x$ for all $x$. If it is affine, give $a$ and $b$ for which $f(x) = a^T x + b$ holds for all $x$. If it is not, give specific $x$, $y$, $\alpha$, and $\beta$ for which superposition fails, i.e.,

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

(Provided $\alpha + \beta = 1$, this shows the function is neither linear nor affine.)

3. Reverse triangle inequality. Suppose $a$ and $b$ are vectors of the same size. The triangle inequality states that $\|a + b\| \leq \|a\| + \|b\|$. Show that we also have

$$\|a + b\| \geq \|a\| - \|b\|. $$
Swapping $a$ and $b$ here yields $\|a + b\| \geq \|b\| - \|a\|$, so we have

$$|\|a\| - \|b\|| \leq \|a + b\| \leq \|a\| + \|b\|.$$ 

**Hints.**

- First draw a picture to get the idea.
- Apply the triangle inequality to $(a + b) + (-b)$.

4. **Regression model sensitivity.** Consider the regression model $\hat{y} = x^T \beta + v$, where $\hat{y}$ is the prediction, $x$ is an feature vector, $\beta$ is a coefficient vector, and $v$ is the offset term. If $x$ and $\tilde{x}$ are feature vectors with corresponding predictions $\hat{y}$ and $\tilde{y}$, show that $|\hat{y} - \tilde{y}| \leq \|\beta\| \|x - \tilde{x}\|$. This means that when $\|\beta\|$ is small, the prediction is not very sensitive to a change in the feature vector.

5. **Linear independence of stacked vectors.**

   (a) Suppose $a_1, \ldots, a_k$ are linearly independent $n$-vectors, and $b_1, \ldots, b_k$ are any $m$-vectors. When are the stacked vectors

   \[
   c_1 = \begin{bmatrix} a_1 \\ b_1 \end{bmatrix}, \ldots, c_k = \begin{bmatrix} a_k \\ b_k \end{bmatrix}
   \]

   linearly independent? Your answer must be one of: Always, Never, or Sometimes. ‘Sometimes’ means that $c_1, \ldots, c_k$ can be linearly independent for some choices of $a_i$ and $b_i$, and linearly dependent for other choices. You must justify your answer.

   (b) Answer the same question, except that now we assume that $a_1, \ldots, a_k$ are linearly dependent.

6. **Building a recommendation engine using k-means.** A set of $N$ users of a music-streaming app listens to songs from a library of $n$ songs over some period (say, a month). We describe this using an $N \times n$ matrix $P$ defined as

   \[
   P_{ij} = \begin{cases} 
   1 & \text{user } i \text{ has played song } j \\
   0 & \text{user } i \text{ has not played song } j
   \end{cases}
   \]

   You can assume that if a user listens to a song, she likes it.

   Your job (say, during a summer internship) is to design an algorithm that recommends to each user 10 songs that she has not listened to, but might like. (You can assume that for each user, there are at least 10 songs that she has not listened to.)

   To do this, you start by running $k$-means on the columns of $P^T$. (It’s not relevant here, but a reasonable choice of $k$ might be 100 or so.) This gives the centroids $z_1, \ldots, z_k$, which are $n$-vectors.

   Now what do you do? You can explain in words; you do not need to give a formula to explain how you make the recommendations for each user.
7. **Matrix-vector multiplication.** For each of the following matrices, describe in words how \( x \) and \( y = Ax \) are related.

(a) \( A = \begin{bmatrix} 0 & 0 & I_k \\ 0 & I_k & 0 \\ I_k & 0 & 0 \end{bmatrix} \).

(b) \( A = \begin{bmatrix} E & 0 & 0 \\ 0 & E & 0 \\ 0 & 0 & E \end{bmatrix} \), where \( E \) is the \( k \times k \) matrix with all entries \( 1/k \).

8. **Plotting in Julia.** Create the 100-vectors \( x \) and \( y \) in Julia where

\[
x_i = \sin(i/8), \quad y_i = \begin{cases} 0 & i \leq 25 \\ (i - 25)/50 & 25 < i < 75 \\ 1 & i \geq 75 \end{cases}
\]

Using MMAPlot, plot \( x \). On a second figure, plot \( x, y \) and \( x + y \). Use different colors to distinguish the three lines, and add labels to the lines.

9. **Nearest neighbor and smallest angle.** Using Julia, find the nearest neighbor of \( a = (1, 3, 4) \) among the vectors

\( x_1 = (4, 3, 5), \quad x_2 = (0.4, 10, 50), \quad x_3 = (1, 4, 10), \quad x_4 = (30, 40, 50) \).

Report the minimum distance of \( a \) to \( x_1, \ldots, x_4 \). Also, find which of \( x_1, \ldots, x_4 \) makes the smallest angle with \( a \) and report that angle.

10. **Checking superposition in Julia.** Generate a random \( 20 \times 10 \) matrix \( A \), as well as 10-vectors \( x \) and \( y \), and scalars \( \alpha \) and \( \beta \). Evaluate the two 20-vectors \( A(\alpha x + \beta y) \) and \( \alpha(Ax) + \beta(Ay) \), and verify that they are very close. (If the numerical calculations were done exactly, they would be equal. Due to very small rounding errors made in the floating-point calculations, they will not be exactly equal.)

**Hint.** The Julia function `rand` can be used to generate random scalars, vectors, and matrices. `rand()` generates a random number, `rand(n)` generates a random \( n \)-vector, and `rand(n,m)` generates a random \( n \times m \) matrix.