ECE367 Problem Set 1

Problem 1.1:

a) The function $\ell_1 : \mathbb{R}^n \to \mathbb{R}$ is defined as: $\|x\|_1 = \sum_{i=1}^{n} |x_i|$. To show this is a valid norm, we will show that it satisfies the 3 conditions on a function for it to be a norm of some vector space.

i) $\|u\| \geq 0 \quad \forall u \in V$, and $\|u\| = 0$ iff $u = 0$

$\|u\|_1 = \sum |v_i|$. Since each term $|v_i|$ in the summation is positive, the entire sum must be positive for any $v \in V$.

$\|u\| = 0 \Rightarrow u = 0$: If $\sum |v_i| = 0$, then $|v_i| = 0 \;\; \forall i \in [n]$, since the absolute value of any real number is always positive or 0. $|v_i| = 0$ implies $v_i = 0$, so by transitivity, $\sum |v_i| = 0$ only if $v = 0$.

$u = 0 \Rightarrow \|u\| = 0$: If $u = 0, \Rightarrow v_i = 0 \;\; \forall i \in [n] \Rightarrow |v_i| = 0 \;\; \forall i \in [n] \Rightarrow \sum_{i=1}^{n} |v_i| = 0$

ii) $\|u + v\| \leq \|u\| + \|v\| \quad \forall u, v \in V$

$\|u + v\| = \sum_{i=1}^{n} |u_i + v_i|$

By triangle inequality on real numbers, $|u_i + v_i| \leq |u_i| + |v_i|$. So:

$\|u + v\|_1 = \sum_i |u_i + v_i| \leq \sum (|u_i| + |v_i|) = \sum |u_i| + \sum |v_i| = \|u\|_1 + \|v\|_1$

iii) $\|\alpha u\| = |\alpha| \|u\|$

$\|\alpha u\|_1 = \sum_i |\alpha u_i| = \sum |\alpha||u_i| = |\alpha| \sum |u_i| = |\alpha| \|u\|_1$

b) The function $\ell_\infty : \mathbb{R}^n \to \mathbb{R}$ is defined as: $\|x\|_\infty = \max_{k \in [n]} |x_k|$. To show this is a valid norm, we will show that it satisfies the 3 conditions on a function for it to be a norm of some vector space.

i) $\|v\| \geq 0 \quad \forall v \in V$, and $\|v\| = 0$ iff $v = 0$

$\|u\|_\infty = \max_{k \in [n]} |x_k| \geq 0$ since the absolute value of any real number is positive or zero

$\|u\|_\infty = 0 \Rightarrow u = 0$: $\max_{k \in [n]} |v_k| = 0$.

Since absolute value of real number is always greater or equal to 0, if the maximum value of $|v_k|$ is 0, then $|v_i| = 0 \;\; \forall i \in [n]$. It follows that $v_i = 0 \;\; \forall i \in [n]$, so that $v = 0$

$u = 0 \Rightarrow \|u\|_\infty = 0$: If $v = 0$, $|v_k| = |0| = 0 \;\; \forall k \in [n]$. So, $\max_{k \in [n]} |v_k| = 0$.

ii) $\|u + v\| \leq \|u\| + \|v\|$:

$\|u + v\|_\infty = \max_{k \in [n]} |u_i + v_i|$

WLOG, assume $\operatorname{argmax}_{k \in [n]} |u_i + v_i| = j$. Then: $\max_{k \in [n]} |u_i + v_i| = |u_j + v_j| \leq |u_j| + |v_j| \leq \max_{k \in [n]} |u_k| + \max_{k \in [n]} |v_k| = \|u\|_\infty + \|u\|_\infty$

iii) $\|\alpha u\|_\infty = |\alpha| \|u\|_\infty$

$\|\alpha u\|_\infty = \max_{k \in [n]} |\alpha u_k| = \max_{k \in [n]} |\alpha||u_k| = |\alpha| \max_{k \in [n]} |u_k| = |\alpha| \|u\|_\infty$

2) Show: $\frac{1}{\sqrt{n}}\|x\|_2 \le \|x_\infty\| \le \|x\|_2 \le \|x\|_1 \le \sqrt{n}\|x\|_2 \le n\|x_\infty\|$

Cauchy-Schwarz Inequality:

$$\frac{|\langle x,y\rangle|}{\|x\|\|y\|} \le 1$$

$$\frac{|x^T y|}{\|x\|\|y\|} \le 1$$

We will show the inequalities from left to right:

① $\frac{1}{\sqrt{n}}\|x\|_2 \le \|x\|_\infty$:

$$\|x\|_2^2 = \sum_{i=1}^{n} |x_i|^2 \le \sum_{i=1}^{n} \left(\max_{k\in[n]} |x_k|\right)^2 = n\|x\|_\infty^2 \Rightarrow \frac{1}{\sqrt{n}}\|x\|_2 \le \|x\|_\infty$$

② $\|x\|_\infty \le \|x\|_2$:

Use cauchy-schwarz with y being the vector with all 0s except 1 at index where x has its maximum element

Then: $\frac{|\langle x,y\rangle|}{\|x\|\|y\|} \le 1 \Rightarrow \frac{\max_{k\in[n]} |x_k|}{\|x\|\|y\|} \le 1 \Rightarrow \frac{\|x\|_\infty}{\|x\|_2} \le 1$ (since $\|y\|=1$) $\Rightarrow \|x\|_\infty \le \|x\|_2$

③ $\|x\|_2 \le \|x\|_1$:

$$\|x\|_2^2 = \sum_{i=1}^{n} |x_i|^2 = x_1^2 + x_2^2 + \dots + x_n^2$$

$$\|x\|_1^2 = \left(\sum_{i=1}^{n} |x_i|\right)^2 = (|x_1|+|x_2|+\dots+|x_n|)^2 = x_1^2 + \dots + x_n^2 + 2\sum_{i=1}^{n}\sum_{j=1, j\ne i}^{n} |x_i||x_j|$$

$$\|x\|_1^2 - \|x\|_2^2 = 2\sum_{i=1}^{n}\sum_{j=1, j\ne i}^{n} |x_i||x_j| \ge 0, \text{ so } \|x\|_2^2 \le \|x\|_1^2 \Rightarrow \|x\|_2 \le \|x\|_1$$

④ $\|x\|_1 \le \sqrt{n}\|x\|_2$

Use Cauchy-Schwarz with:

$$y_i = \begin{cases} 1, & x_i \ge 0 \\ -1, & x_i < 0 \end{cases}$$

Then: $|\langle x,y\rangle| = \|x\|_1$

$$\frac{|\langle x,y\rangle|}{\|x\|_2\|y\|_2} = \frac{\|x\|_1}{\|x\|_2\sqrt{n}} \le 1 \Rightarrow \sqrt{n}\|x\|_2 \ge \|x\|_1$$

⑤ This follows directly from ①

3a) Suppose $a_1, \dots a_k$ lin. ind. Can we conclude $c_1, \dots c_k$ are lin. ind?

Assume for sake of contradiction $c_1, c_2, \dots c_k$ are linearly dependent. Then $\exists\, \alpha_1, \dots \alpha_k \in \mathbb{R}$, not all 0, such that:

$$\alpha_1 c_1 + \alpha_2 c_2 + \dots + \alpha_k c_k = 0$$

$$\alpha_1 \begin{bmatrix} a_1 \\ b_1 \end{bmatrix} + \alpha_2 \begin{bmatrix} a_2 \\ b_2 \end{bmatrix} + \dots + \alpha_k \begin{bmatrix} a_k \\ b_k \end{bmatrix} = 0$$

This implies: $\alpha_1 a_1 + \alpha_2 a_2 + \dots + \alpha_k a_k = 0$

But $a_1, a_2, \dots a_k$ are linearly independent. Contradiction, so $c_1, c_2, \dots c_k$ are independent.

3b) If $c_1, c_2, \dots c_k$ linearly dependent, then $\exists\, \alpha_1, \alpha_2, \dots \alpha_k$ not all 0 such that $\alpha_1 c_1 + \dots + \alpha_k c_k = 0$. So we require constants $\alpha_i, i \in [k]$ such that $\sum \alpha_i a_i = 0$ and $\sum \alpha_i b_i = 0$. While the first is guaranteed by our assumption that $a_1 \dots a_k$ are dependent, the second one isn't. In general then, the statement is false. Particularly, if $c_1, \dots c_k$ are independent, then $c_1, c_2, \dots c_k$ cannot be dependent.

4)

$\langle x+y, x-y\rangle = \langle x,x\rangle - \cancel{\langle x,y\rangle} + \cancel{\langle y,x\rangle} + \langle y,y\rangle = \|x\|^2 - \|y\|^2 = 1-1=0$

Since $x+y$ and $x-y$ are orthogonal, they are linearly independent. Further,

$\text{span}\{x,y\} = \text{span}\{x+y, x-y\}$.

So, $\{x+y, x-y\}$ forms an orthogonal basis for $\mathbb{R}^n$

5) A function $f(x,y): V\times V \to \mathbb{R}$ is an inner product if it satisfies:

i) $\langle x,y\rangle = \langle y,x\rangle$

ii) $\langle \alpha x, y\rangle = \alpha\langle x,y\rangle$

iii) $\langle x+y, z\rangle = \langle x,z\rangle + \langle y,z\rangle$

iv) $\langle x,x\rangle \geq 0$ and $\langle x,x\rangle = 0$ iff $x=0$

We will go through all conditions:

i) $f(x,y) = \sum_k \alpha_k x_k y_k = \sum_k \alpha_k y_k x_k = f(y,x) \Rightarrow$ No conditions on α_k

ii) $f(kx,y) = \sum_k \alpha_k (k x_k) y_k = k\sum_k \alpha_k x_k y_k = k f(x,y) \Rightarrow$ No condition on α_k

iii) $\langle x+y, z\rangle = \sum_k \alpha_k (x_k+y_k) z_k = \sum \alpha_k x_k z_k + \sum \alpha_k y_k z_k = f(x,z)+f(y,z) \to$ No conditions on α_k

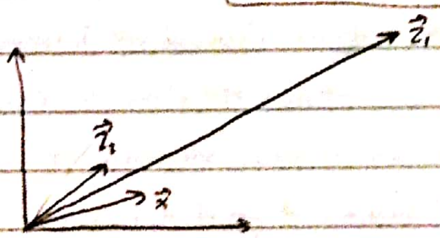
iv) $\langle x,x\rangle \geq 0 \Rightarrow \sum \alpha_k x_k x_k = \sum \alpha_k x_k^2$.

$\sum \alpha_k x_k^2 \geq 0$ iff $\alpha_k \geq 0 \ \forall\ k\in[n]$

$\langle x,x\rangle = 0$ iff $x=0 \Rightarrow \alpha_k \neq 0 \ \forall\ k\in[n]$.

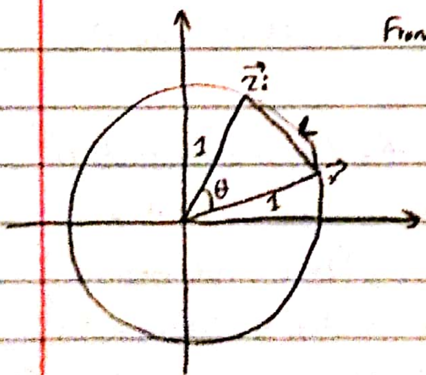
So, we require $\boxed{\alpha_k > 0 \ \forall\ k\in[n]}$

6a)



Here, closest distance to $\vec{x}$ is $\vec{z}_2$, but closest angle to $\vec{x}$ is $\vec{z}_1$.

b) If $\|z_i\| = 1 \ \forall\ i\in[n]$, then the tips of the vectors lie on a circle with radius 1.



From cosine law,

$L = 1^2+1^2 - 2(1)(1)\cos\theta = 2 - 2\cos\theta = 2(1-\cos\theta)$ where L is distance and θ is angle.

$1-\cos\theta = \frac{L}{2} \Rightarrow \cos\theta = 1-\frac{L}{2} \Rightarrow \theta = \arccos\left(1-\frac{L}{2}\right)$. $0\leq L \leq 2$

This is a monotonically increasing function. As $L\uparrow$, $\theta\uparrow$. So, the vector with smallest distance also has smallest angle.

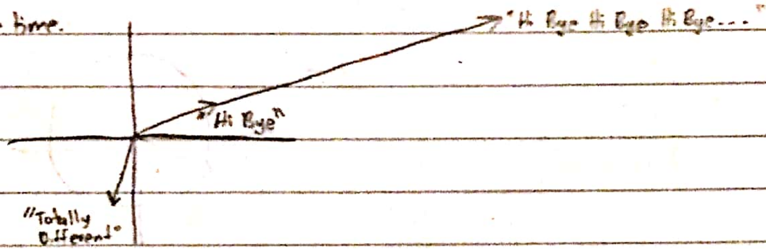
Application Problems

1.7a)

Document 7 and Document 8 have minimum euclidean distance of 20.27 between them

Document 9 and Document 10 have minimum angle of 31.52° between them

These results are not the same. This is expected in general. Distance and angle measure 2 different things. Distance measures difference in words. If two documents are about the same thing but one is much longer than the other, distance will be larger than 2 completely different articles each very short. Angle is a measure of how similar two documents are by comparing the proportions of each word in the entire document. For example, "Hi Bye" and "Hi Hi Bye Bye" are collinear because in each case Hi and Bye both appear 50% of the time.



b) Document 9 and 10 have minimal distance of 0.064299 between them

Document 9 and 10 have minimum angle of 30.52° between them

The answers now match. This makes sense. This normalization is used to replace term count in each document by relative term count (a fraction) representing what fraction of total words that word is. By using a measure of importance of each word in a document rather than literal word counts, we get a computation for distance that is smaller if two documents match in terms of the importance of each word, which is a better way to summarize an article into a vector.

Of course, angles remain the same as we are just scaling each vector in the normalization process. Angle and distance are now both trying to measure the same things.

c) Document 9 and 10 have minimum euclidean distance of 0.082128 between them

d) The previous normalization step scaled vectors to represent the importance of each word in each document. So, given a document, its components reflected how important those words were in the document.

The new step, $\times\sqrt{\log\left(\frac{|D|}{f_{doc}(t)}\right)}$ changes components to include more information, specifically, the importance of a word in a particular document compared to other documents

$$w(t,d) = \frac{f_{term}(t,d)}{\sum_{i}^{n} f_{term}(t,d)} \sqrt{\log\left(\frac{|D|}{f_{doc}(t)}\right)}$$

← account for how important document d is for term t.

↑ account for how important each t is in document d

So, if one word appears twice and another once in same document, ~~but~~ that word that appears once only appears in this document, it will be scaled up more than the word that appears twice, to reflect its importance in that particular doc

Scanned with CamScanner

1.7a

Find the two documents with minimum distance

```
load wordVecV.mat
[W, D] = size(V);
minlen = sum((V(:,1)-V(:,2)).^2);
minlen_v1 = 1;
minlen_v2 = 2;

fprintf("Unnormalized:\n")
```

```
Unnormalized:
```

```

for i = 1:D-1
   for j = i+1:D
      pair_len = sum((V(:,i)-V(:,j)).^2);
      if pair_len < minlen
          minlen = pair_len;
          minlen_v1 = i;
          minlen_v2 = j;
      end
   end
end
fprintf('Document %d and Document %d have minimum euclidean distance of %.02f between them\n',
```

```
Document 7 and Document 8 have minimum euclidean distance of 24.72 between them
```

```

% Find the two documents with minimal angle
minang = acos(dot(V(:,1), V(:,2))/(norm(V(:,1))*norm(V(:,2))))*180/pi;
minang_v1 = 1;
minang_v2 = 2;
for i = 1:D-1
   for j = i+1:D
      pair_ang = acos(dot(V(:,i), V(:,j))/(norm(V(:,i))*norm(V(:,j))))*180/pi;
      if pair_ang < minang
          minang = pair_ang;
          minang_v1 = i;
          minang_v2 = j;
      end
   end
end
fprintf('Document %d and Document %d have minimum angle of %.02f degrees between them\n', minar
```

```
Document 9 and Document 10 have minimum angle of 30.52 degrees between them
```

1.7b

```
Vnorm = V;
for i = 1:D
    Vnorm(:,i) = V(:,i)/sum(V(:,i));
end
sum(Vnorm); % All columns should sum to 1 now
```

```
% Now do the same thing as before, use this as new f(t,d) matrix

fprintf("\nNormalized:\n")
```

```
Normalized:
```

```
% Length
minlennorm = sum((Vnorm(:,1)-Vnorm(:,2)).^2);
minlennorm_v1 = 1;
minlennorm_v2 = 2;
for i = 1:D-1
   for j = i+1:D
      pair_len = sum((Vnorm(:,i)-Vnorm(:,j)).^2);
      if pair_len < minlennorm
         minlennorm = pair_len;
         minlennorm_v1 = i;
         minlennorm_v2 = j;
      end
   end
end
fprintf('Document %d and Document %d have minimum euclidean distance of %f between them\n', mir
```

```
Document 9 and Document 10 have minimum euclidean distance of 0.064299 between them
```

```
% Angle
% Find the two documents with minimal angle
minangnorm = acos(dot(Vnorm(:,1), Vnorm(:,2))/(norm(Vnorm(:,1))*norm(Vnorm(:,2))))*180/pi;
minangnorm_v1 = 1;
minangnorm_v2 = 2;
for i = 1:D-1
   for j = i+1:D
      pair_ang = acos(dot(Vnorm(:,i), Vnorm(:,j))/(norm(Vnorm(:,i))*norm(Vnorm(:,j))))*180/pi;
      if pair_ang < minangnorm
         minangnorm = pair_ang;
         minangnorm_v1 = i;
         minangnorm_v2 = j;
      end
   end
end
fprintf('Document %d and Document %d have minimum angle of %.02f degrees between them\n', minar
```

```
Document 9 and Document 10 have minimum angle of 30.52 degrees between them
```

1.7c

```
% V_tfidf = Vnorm;
for i = 1:W
   V_tfidf(i,:) = Vnorm(i,:) * sqrt(log(D/nnz(Vnorm(i,:))));
end

fprintf("\nTFIDF:\n")
```

TFIDF:

```
% Length
minlentfidf = sum((V_tfidf(:,1)-V_tfidf(:,2)).^2);
minlentfidf_v1 = 1;
minlentfidf_v2 = 2;
for i = 1:D-1
   for j = i+1:D
      pair_len = sum((V_tfidf(:,i)-V_tfidf(:,j)).^2);
      if pair_len < minlentfidf
          minlentfidf = pair_len;
          minlentfidf_v1 = i;
          minlentfidf_v2 = j;
      end
   end
end
fprintf('Document %d and Document %d have minimum euclidean distance of %f between them\n', mir
```

```
Document 9 and Document 10 have minimum euclidean distance of 0.082128 between them
```