1 Linear Regression

1. T/F: Let r be the correlation coefficient between $\vec{x} = (x_1, \ldots, x_n)$ and $\vec{y} = (y_1, \ldots, y_n)$. Then $\arccos r$ is the angle between the vectors $\vec{x} - \bar{x}$ and $\vec{y} - \bar{y}$.

2. Recall that the CBS inequality gives us

$$|\vec{v} \cdot \vec{w}| \leq |\vec{v}| \cdot |\vec{w}|. \tag{1}$$

   (a) What does (1) tell us about the angle between $\vec{v}$ and $\vec{w}$?

   (b) Let $\vec{x} = (x_1, \ldots, x_n)$, and $\vec{y} = (y_1, \ldots, y_n)$.

   Let $\vec{v} = (x_1 - \bar{x}, \ldots, x_n - \bar{x})$, and $\vec{w} = (y_1 - \bar{y}, \ldots, y_n - \bar{y})$.

   What does (1) tell us about the correlation coefficient r between $\vec{x}$ and $\vec{y}$?

   (c) For what kind of vectors $\vec{v}$ and $\vec{w}$ is equality attained, when $\vec{v}$ and $\vec{w}$ are vectors with two components?

2 Needleman-Wunsch Algorithm

Move	X / _ (→)	_ / X (↓)	X / Y (↘)	X / X (↘)
Score	0	0	0	1

1. T/F:

   (a) A diagonal move is the only move that could increase the total score.

   (b) Any diagonal move can increase the total score.

2. Translate the following alignment into a path. What's its score?

   G C T A _
   G _ T A G

3. Translate the following path into an alignment. What's its score?

	G	T	C	G
T	→	↘		
A			↘	
G				↘
C				↓

4. (HW37 # 2 partial) Consider the following new scoring system in Table 1. Compare the two alignments. Which one is better?

5. (HW37 # 3 partial)

new gap	gap extension	match	mismatch
-5	-1	1	-1

Table 1: New scoring system

```
G  A  A  A  A  A  A  T     G  A  A  A  A  A  A  T
G  _  _  A  _  A  _  T     G  A  A  _  _  _  _  T
```

(a) Besides being square, what type of matrix is the similarity matrix in table 2?

(b) Write the similarity matrix for the alphabet below, and then compute its eigenvalues and eigenvectors.

   i. Alphabet "AC".

   ii. (Challenge) Alphabet "ACG".

	A	C	G	T
A	1	-1	-1	-1
C	-1	1	-1	-1
G	-1	-1	1	-1
T	-1	-1	-1	1

Table 2: Similarity matrix