Similarity Search Tutorial

Lin Guosheng
School of Computer Science and Engineering
Nanyang Technological University

Q1 LSH

- **Q1**: given the query sample and the samples in the database below, find the top 2 nearest neighbours in the database. Use LSH-random projection with the hash functions given below.

Query sample:

A	[0, -1]

Database samples (5):

B	[-2, 0]
C	[1, 2]
D	[2, 1]
E	[1, -1]
F	[-1, 2]

w_1 =[-1 1] ^{T};
w_2 =[-1 0] ^{T};
w_3 =[0 1] ^{T};
w_4 =[1 -1] ^{T};
w_5 =[1 0] ^{T};
w_6 =[-1 -1] ^{T};

Note: the question is modified from the LSH example in the lecture class by adding two additional hashing functions (w_5 and w_6)

Q2 PQ

- **Q2. Product Quantization (PQ)**

Two input vectors are given below:
A: [1, 0, 2, 1, -1, 0]
B: [-1, 3, 1, 4, 3, 2]

Define 3 Subspaces. Each subspace has 3 centroids:
Subspace 1: C1: [-2, 1], C2: [3, 2], C3: [2, -1]
Subspace 2: C1: [1, 0], C2: [2, 3], C3: [-1, 3]
Subspace 3: C1: [-2, 3], C2: [1, -1], C3: [1, 4]

a): Use PQ to compute the compressed vectors of A and B

b): Construct distance lookup tables and calculate the approximate (symmetric case) squared-L2 distance of A and B using the compressed vectors.