Secure Binary Embeddings for Privacy Preserving Nearest Neighbors

- contributions of paper
  - a scheme for privacy preserving nearest neighbor search based on a secure stable embedding using quantized random projections
  - show how to use this scheme by presenting protocols for clustering and authentication applications

- quantization process
  - \( x \in \mathbb{R}^K, \ y_m = \langle x, a_m \rangle + w_m, \ q_m = Q\left(\frac{y_m}{\Delta_m}\right) \rightarrow q = Q(\Delta^{-1}(Ax + w)) \)
    - \( x \): \( K \)-dimensional signal
    - \( m = 1, \ldots, M \): measurement index
    - \( y_m \): unquantized measurements
    - \( a_m \): measurement vectors \( \rightarrow A \): random matrix i.i.d., \( \mathcal{N}(\mu = 0, \sigma^2) \)
    - \( w_m \): additive dither (noise), uniformly distributed in \([0, \Delta]\)
    - \( \Delta_m \): quantization precision parameters \( \rightarrow \Delta \): diagonal matrix
    - \( Q(\cdot) \): quantizer

- universal quantization
  - scalar quantizer with non-contiguous quantization regions
  - determine an upper bound for the probability that there exist two signals \( x \) and \( x' \) with distance greater than \( d \) that quantize to the same quantization vector given the number of measurements \( M \)

\[ P(q = q') = P(x, x' \text{ consistent} | d) = \frac{1}{2} + \sum_{i=0}^{\infty} \frac{e^{-\left(\frac{\pi(2i+1)\pi d}{\sqrt{2}\Delta}\right)^2}}{(\pi(i+1/2))^2} \] (Figure 3)

\[ P(q = q') = P(x, x' \text{ consistent} | d) \leq \frac{1}{2} + \frac{1}{2} e^{-\left(\frac{\pi d}{\sqrt{2}\Delta}\right)^2} \]

\[ \geq 1 - \sqrt{\frac{2}{\pi}} \cdot \frac{\sigma d}{\Delta} \]
\[ \geq 1 - \sqrt{\frac{2}{\pi}} \cdot \frac{\sigma d}{\Delta} \]
\[ \geq 1 + \frac{4}{\pi^2} e^{-\left(\frac{\pi d}{\sqrt{2}\Delta}\right)^2} \]

\[ d = \|x - x'\|_2, \ q = Q\left(\frac{\langle x, a \rangle + w}{\Delta}\right), \ q' = Q\left(\frac{\langle x', a \rangle + w}{\Delta}\right) \]
• secure binary embeddings
  – the quantization process used as an embedding has similar properties to Locality Sensitive Hashing (LSH)
  – information-theoretic security:
    \[ I(q_i; q'_i|d) = \sum_{q_i, q'_i \in \{0, 1\}} P(q_i, q'_i|d) \log \frac{P(q_i, q'_i|d)}{P(q_i|d)P(q'_i|d)} \leq 10e^{-\left(\frac{\pi \sigma d}{\sqrt{2}}\right)^2} \] (Figure 4)
    \[ I(q; q'|d) \leq 10Me^{-\left(\frac{\pi \sigma d}{\sqrt{2}}\right)^2} \]
  
  \( d < \Delta \): the distance between the quantized measurements provides information about the distance between the signals
  \( d > \Delta \): we cannot recover any information just by observing the quantized measurements

  – stable embedding:
    provide a relationship between the distance of the signals and the distance of their quantized measurements
    binary space: \( \{0, 1\}^M \rightarrow \) Hamming distance: \( d_H(q, q') = \frac{1}{M} \sum_m (q_m \oplus q'_m) \)
    with probability \( 1 - 2e^{2\log L - 2d^2M}, 1 - P_{cl(d)} - t \leq d_H(q, q') \leq 1 - P_{cl(d)} + t \)
    \( L \): number of points to be embedded securely
    \( P_{cl(d)} \): shorthand for \( P(x, x' \text{ consistent}|d) \)
    \( t \): control variable

• applications
  – privacy preserving clustering with a star topology
  – authentication using symmetric keys
  – privacy preserving clustering with two parties
Figure 1: Scalar quantization.

Figure 2: Universal quantization.
Figure 3: Probability of equal bits, $P(q_i = q'_i)$.

Figure 4: Cross-bit information, $I(q_i, q'_i | d)$. 