Locality-Sensitive Hashing

- implementations
  - likelike: http://code.google.com/p/likelike/
    * specific for item recommendation in E-commerce sites
    * features must be positive integers
  - LSHKIT: http://lshkit.sourceforge.net/index.html
    * requires filling C++ class templates with configuration values

- phoneme proximity
  - use E2LSH package
    * ready to use
    * allows to specify values for $R$ and $P_1$
      (if $\|p - q\| \leq R$ then $\Pr[H(h(q) = h(p))] \geq P_1$)
  - LSH parameters
    * $R$ range: 10 → 20, step: 1
    * $P_1$ range: 0.80 → 0.98, step: 0.02
    * TIMIT corpus speakers: $abc0$, $ask0$, $cth0$
  - analysis of the results
    * figures 1-3 present the average of correct classification over all the phones
    * figures 4-6 present the average detection of positive examples, $tp/p$, over all the phones
    * results are independent from the value of $P_1$
    * best classification results for $R = 11$ and $R = 12$
    * best positive example detection for $R = 15$ and $R = 16$
Figure 1: Phone classification, TIMIT corpus speaker \textit{abc0}.

(a) With \textit{h#} phone.  
(b) Without \textit{h#} phone.

Figure 2: Phone classification, TIMIT corpus speaker \textit{aks0}.

(a) With \textit{h#} phone.  
(b) Without \textit{h#} phone.
Figure 3: Phone classification, TIMIT corpus speaker cth0.

Figure 4: Positive examples detection, TIMIT corpus speaker abc0.
Figure 5: Positive examples detection, TIMIT corpus speaker *aks0*.

Figure 6: Positive examples detection, TIMIT corpus speaker *cth0*.