Lecture 5 - Extensions

Recall:

Any LST function satisfies the triangle inequality. Proof via

\[ d_{x|x'} \leq d_x + d_{x|x'} \]

**Lemma:** Jacquet mean is a kernel (and so are other similarities)

**Proof:**

\[ d_{x|x'} = \langle e_x, e_{x'} \rangle \geq 0 \]

**Lemma:** Any \( \text{sin}(x_1, x_1') = \mathbb{E} \left[ \delta_x(x_1), \delta_{x'}(x_1') \right] \)

can be an embedded as binary mapping

**Proof:** Construct new hash function

\[ \delta_x(x) \in \{0, 1\} \]

\[ \mathbb{E} \left[ \delta_x(x), \delta_y(x) \right] = \frac{1}{2} + \frac{1}{2} \text{sin} \left( \frac{x}{2} \right) \]

**Lemma:** Not all kernels can be embedded as a Sin kernel

**Proof:**

\[ \langle (1,0), (-1,0) \rangle = -1 \neq 0 \]

**Extension:**

\[ \text{sin}(x_1, x_1') = \mathbb{E} \left[ \delta_x(x_1), \delta_{x'}(x_1') \right] \]

**Lemma:** \( \text{sin}(x_1, x_1') \) embeds all kernel \( \Phi \)

and to append map \( a, b \) to binary

\[ \rightarrow \text{Sign is easy} \rightarrow \text{Sign} (a) \rightarrow \Phi (i) \]

This is the absolute upper bound per coordinate

The only chance to match

**Condition**

1) Trace class \( \leq \lambda; \leq 0 \)

2) \( \lambda, \text{sign} \ll \ll \Phi(x) \ll \ll 0 \)

or rather \[ \leq \lambda; \ll \ll \Phi(x) \ll \ll \leq 0 \]

but all hand satisfy this since we have

\[ k(x, x') = \text{sign}(x) \Phi_i(x) \]

So, no summation must be \( \infty \)

**Proposition of the Embedding basis**

We can use this to approximate kernels, too

\[ k(x_1, x_1') \ll \ll \frac{1}{\mu} \ll \ll \text{sign}(x) \ll \ll \Phi_i(x) \ll \ll \ll \Phi_i(x') \]

\[ \rightarrow \text{we would only 2 bits per dimension} \]
Properties of the Min-Wise Hash

**Min-wise hash (recall)**

\[
\text{Sim}(X, X') = \frac{\mathbb{E}}{|V|} \left[ \min \left( \pi(X) \right) \equiv \min \left( \pi(X') \right) \right]
\]

in practice take \( \min \) rather than \( k \)-times \( \min \)

**Definition:**

- Min-wise indep. permutation family
- \( P_f(\min(\pi(X)) = \pi(x)) = \frac{1}{|X|} \)
  \hspace{1cm} \text{for } x \in X

**Surprising theorem:** (Broder & Mitzenmacher, 2001)

For any mapping \( f : P(X) \rightarrow \Omega \) there exists a permutation \( \pi_f \) such that

\[
f(X) = f(\pi_f(\min(\pi_f(X)))
\]

and (obviously) there is a distribution over \( \pi_f \)

such that

\[
\mathbb{E} \left[ \min(\pi_f(X)) = \min(\pi_f(X')) \right]
\]

**Auxiliary Results**

**Theorem to prove:**

Assume that there is a mapping \( f \), \( P(f) \) s.t.

\[
P_f(\{ f(A) = f(B) \}) = \frac{|A \cap B|}{|A \cup B|}
\]

Then there exist \( \pi_f \) such that

\[
f(X) = f\left( \pi_f^{-1}(\min(\pi_f(X))) \right)
\]

**Lemma 1:** \( P_f( f(X) = f(x1) ) = \frac{1}{|X|} \)

(by definition)

**Lemma 2:** \( P_f( f(x1) = f(x) ) = 0 \)

(by definition)

**Lemma 3:** \( P_f( f(X) \in \{ f(x1) \mid \forall i \} ) = 1 \)

(by Lemma 1)

**Lemma 4:** For \( X \leq Y \) in law

\[
P_f( f(X) = f(Y) ) = \frac{|X|}{|Y|}
\]

And if \( f(Y) \in \{ f(x1) \mid \forall i \} \)

then \( f(Y) = f(x1) \)

**Proof:** Part 1 is trivial

**Advance Preview:*

- \( f \) is invertible (do show)
- Extract \( x_i \): use \( f^{-1}(x_1, \ldots, x_i) \)
Fail bonds

Corollary: Accuracy reduces variance

\[ x_i \text{ with } E\{x_i\} = \mu \]

\[ \text{Var}\{x_i\} = E\{x_i^2\} - (E\{x_i\})^2 = \sigma^2 \]

Define \[ X_i = \frac{1}{n} \sum_{j=1}^{n} x_{ij} \]

\[ E\{X_i\} = \mu \quad \text{Var}\{X_i\} = \frac{1}{n} \sigma^2 \]

(since variance add up inprod)

Corollary:

\[ \Pr \left\{ \left| X - \mu \right| > \frac{\sigma}{\sqrt{n}} \right\} \leq \frac{1}{\sqrt{n}} \]

→ Terrible scaling behavior in \( S \) not good in \( Y \)

→ Want logarithmic in \( S \), but cannot use Bernstein / Chernoff since the moments are not bounded ...

Chernoff bound: for \( X_i \in \mathbb{R} \), we have

\[ \Pr \left\{ \sum_{i=1}^{S} X_i > \sum_{i=1}^{S} E[X_i] + \varepsilon \right\} \leq \exp\left( -2 \frac{\varepsilon^2}{C^2} \right) \]

when \( C^2 = \sum_{i=1}^{S} (b_i - a_i)^2 \)

Key idea: Controlling variance & sample size separately

\[ X_i = \frac{1}{n} \sum_{j=1}^{n} x_{ij} \]

\[ X = \text{median} \left( \{X_i\}_{i=1}^{n} \right) - \mu \]

\[ \Pr \left\{ \left| (\hat{X} - \mu) > \frac{\sigma}{\sqrt{n}} \right| \right\} < \frac{1}{\sqrt{n}} \]

\[ \Pr \left\{ \left| \text{med} \{X_i - \mu\} \right| > 8.6 \right\} \leq \frac{1}{n} \frac{8.6^2}{2} \]

Lots of things just go bad before this happen

this is a R.V. with \( \sum_{i=1}^{S} \leq \frac{1}{\sqrt{n}} \)

\[ \implies \quad \varepsilon = \delta \left( \frac{1}{2} - \frac{1}{\sqrt{n}} \right) \]

\[ C = \delta \]

\[ \implies \Pr \left\{ \text{failure} \right\} \leq \exp\left( -2 \delta \left( \frac{1}{2} - \frac{1}{\sqrt{n}} \right)^2 \right) \]

\[ \implies \sigma + \frac{1}{\sqrt{n}} = \frac{1}{4} \implies \delta = \frac{1}{\sqrt{n}} \]

\[ \implies \text{ prob is } \exp\left( -2 \frac{\delta^2}{C^2} \right) \]

McDiarmid ineq. (\ref{95})

\[ \left| f(x) - f(i) \right| \leq C_i \]

\[ \implies \\Pr \left\{ \sum_{i=1}^{S} f(x_i) > \sum_{i=1}^{S} f(x) \right\} \leq \exp\left( -2 \frac{\varepsilon^2}{C^2} \right) \]

when \( C^2 = \sum_{i=1}^{S} C_i^2 \)

Self-bounding ineq of McDiarmid \( \leq \text{rad} (X; \delta) \)

\[ \left| g(x) - g(i) \right| \leq a \cdot g(x) + b \]

where \( g(i) = \min_{X_i} g(X) \leq \left( \text{rad} + \text{bias} \right) \)

then \( \Pr \left\{ g(x) - \sum_{i=1}^{S} g(x_i) \geq \varepsilon \right\} \leq \exp\left( -2 \frac{\varepsilon^2}{C^2} \right) \]

\[ \leq \frac{1}{2} - a \cdot g(x) \]
Proof of main theorem:

Recall: Gauss-Markov inequality

\[ \Pr \{ X > \mu C \} \leq \frac{1}{C} \text{ for } x > 0 \]

Useful corollary

\[ \Pr \left\{ \left( \min_{i \in [k]} X_i \right) > \mu C \right\} \leq \left( \frac{1}{C} \right)^k \]

(will use this for count min sketch)

Quantile trick

Define \( F(x) = \int_{-\infty}^{x} \Pr(X) \) for \( x \in \text{dom}(\Pr) \)

and

\[ \text{define } F_k(x) = \Pr \left\{ \min_{i \in [k]} X_i \leq x \right\} \]

Useful result: (will not prove this)

\[ \Pr \left\{ \max_{i \in [k]} (\pi(X)) = \pi(X) \right\} = \frac{1}{|X|} \]

for min-wise independent families.