JL Lemma - Extensions & Applications

David Veitch

University of Toronto

December 8 2020

Agenda

1. JL Lemma - Reminder
2. Database Friendly Random Projections
3. Compressed Least Squares Regression
4. Application - Modelling Computer Prices
5. Application - Visualizing Countries' Development

JL Lemma - Reminder

Formal Statement

Let $\epsilon \in (0, \frac{1}{2})$. Let $Q \subset \mathbb{R}^d$ be a set of n points and $k = \frac{20 \log n}{\epsilon^2}$. There exists a Lipschitz mapping $f : \mathbb{R}^d \to \mathbb{R}^k$ such that for all $u, v \in Q$

$$(1-\epsilon)||u-v||^2 \leq ||f(u)-f(v)||^2 \leq (1+\epsilon)||u-v||^2$$

What it Means

If we choose a matrix A where the entries of A are $\mathcal{N}(0,1)$ then for vectors $u, v \in \mathbb{R}^d$ then $||u-v||^2 \approx ||Au-Av||^2$

Problem!

- For $x_1, \ldots, x_n \in \mathbb{R}^d$, and projection matrix $A \in \mathbb{R}^{d \times k}$ where $k < d$, matrix multiplication $O(ndk)$.
- SQL can easily generate Uniform $[0, 1]$ random variables, harder to generate normal random variables

Problem! - SQL Code Sample

Generate Uniform $[0, 1]$ Random Variable

```
select rand()
```

Generate Uniform $\{-1, 1\}$ Random Variable

```
select ceiling(rand()-.5)*2-1
```

Generate Normal Random Variable

```
select randNum1 = rand()
select randNum2 = rand()
select value1 = round( (sqrt( -2.0*log(randNum1) )
    *cos (2pi*randNum2) )*stdDev, precision)+ mean
select value2 = round((sqrt(-2.0*log(randNum1))
    *sin(2pi*randNum2))*stdDev, precision)+ mean
```

Idea

- What if A constructed without using Gaussian random variables?

Idea

- What if A constructed without using Gaussian random variables?
- Can we do this?

Idea

- What if A constructed without using Gaussian random variables?
- Can we do this?
- Let's revisit the proof of the JL Lemma. What facts did we use about Gaussian random variables?

Idea

- What if A constructed without using Gaussian random variables?
- Can we do this?
- Let's revisit the proof of the JL Lemma. What facts did we use about Gaussian random variables?
  - Moments
  - Bound on χ^2 random variable

Idea

- What if A constructed without using Gaussian random variables?
- Can we do this?
- Let's revisit the proof of the JL Lemma. What facts did we use about Gaussian random variables?
  - Moments
  - Bound on χ^2 random variable
- How about we make A with nicer random variables!

How about we make A with nicer random variables! Let A's entries $\{a_{ij}\}_{1\leq i\leq d, 1\leq j\leq k}$ be as follows

$$a_{ij} = \begin{cases} 1 & \text{with probability } \frac{1}{2} \\ -1 & \text{with probability } \frac{1}{2} \end{cases} \tag{1}$$

or

$$a_{ij} = \begin{cases} \sqrt{3} & \text{with probability } \frac{1}{6} \\ 0 & \text{with probability } \frac{2}{3} \\ -\sqrt{3} & \text{with probability } \frac{1}{6}. \end{cases} \tag{2}$$

- For random variable 1 we only generate random variables $\in \{-1, +1\}$!
- For random variable 2 we can throw away 2/3 of the data!

Let $Q \subset \mathbb{R}^d$ be a set of n points. For a given $\epsilon, \beta > 0$ let

$$k_0 = \frac{4 + 2\beta}{\epsilon^2/2 - \epsilon^3/3} \log n$$

and for some integer $k \geq k_0$ let $A \in \mathbb{R}^{d \times k}$ be a random matrix where its entries are independent random variables from the prveiously mentioned distributions. Let

$$E = \frac{1}{\sqrt{k}} QA$$

and $f : \mathbb{R}^d \to \mathbb{R}^k$ map the i^{th} row of Q to the i^{th} row of E. Then with probability of at least $1 - n^{-\beta}$ for all $u, v \in Q$

$$(1 - \epsilon)||u - v||^2 \leq ||f(u) - f(v)||^2 \leq (1 + \epsilon)||u - v||^2.$$

Moral of the story: JL Lemma still works!

Classic Regression

$\{x_j, y_j\}_{1\leq j\leq n}$, $x_j \in \mathbb{R}^d$, $y_j \in \mathbb{R}$, want to find $\hat{\beta} \in \mathbb{R}^{d+1}$ such that

$$\hat{\beta} = \underset{(\beta_0,\ldots,\beta_d)}{\arg\min} \sum_{j=1}^{n} (y_j - (\beta_0 + \beta_1 x_{j,1} + \cdots + \beta_d x_{j,d}))^2$$

Classic Regression

$\{x_j, y_j\}_{1\leq j\leq n},\ x_j \in \mathbb{R}^d,\ y_j \in \mathbb{R}$, want to find $\hat{\beta} \in \mathbb{R}^{d+1}$ such that

$$\hat{\beta} = \underset{(\beta_0,\ldots,\beta_d)}{\arg\min} \sum_{j=1}^{n} (y_j - (\beta_0 + \beta_1 x_{j,1} + \cdots + \beta_d x_{j,d}))^2$$

Problem!

Suppose $d > n$. Where $X \in \mathbb{R}^{n\times d}$ is a matrix of the x's stacked. To solve for $\hat{\beta}$ we use

$$\hat{\beta} = (X^T X)^{-1} X^T Y.$$

But if $X^T X$ invertible $\Rightarrow X$ has linearly independent columns. But X has more columns than rows, therefore this is impossible!

Leads to solution not being unique.

Solution

Multiply X by a random matrix to reduce its dimensionality.

For example if $X \in \mathbb{R}^{100\times 10000}$, $A \in \mathbb{R}^{10000\times 10} \Rightarrow XA \in \mathbb{R}^{100\times 10}$.

$$\hat{\beta} = \underset{(\beta_0,\ldots,\beta_{10})}{\arg\min} \sum_{j=1}^{n} (y_j - (\beta_0 + \beta_1 [XA]_{j,1} + \cdots + \beta_d [XA]_{j,d}))^2$$

A tractable problem!

Bound on Squared Loss of Reduced Function (Remark 1)

$$\underbrace{\mathbb{E}[||\hat{g}_L - f^*||_P^2]}_{\substack{\text{Expected squared error}\\ \text{of reduced dimension function}}} = O\left(||\alpha^+||\sqrt{\mathbb{E}[||\varphi(X)||^2}\frac{\log n/\delta}{\sqrt{n}} + \underbrace{\inf_{f\in\mathcal{F}_d}||f - f^*||_P^2}_{\substack{\text{Expected squared error}\\ \text{of ideal function without}\\ \text{reducing dimension}}}\right)$$

So even if we are doing regression in this reduced dimension setting, as n grows we can still achieve something close to optimal.

Problem

Predict sales of 231 tablet computers each week for 24 weeks period ($n = 5544$).

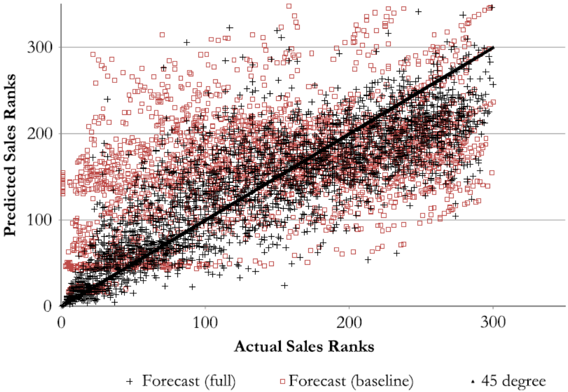
Features

- 14 dimesions: features that are characteristics of each tablet (e.g. RAM, storage, weight)
- 20,068 dimensions: Bag of words - word counts from customer reviews of each product

Solution

Using a random projection, reduce dimension of word vector to $k = 300$.

	Baseline Model (no words)	Full Model (with words)
R^2	34.2%	86.2%
Forecast Error	163.5%	37.2 %



The forecasts of sales ranks from the full model are closer to actual sales ranks than the baseline model. One can also see the baseline model produces quite a few forecasts which are wildly off.

Application - Visualizing Countries' Development

Question

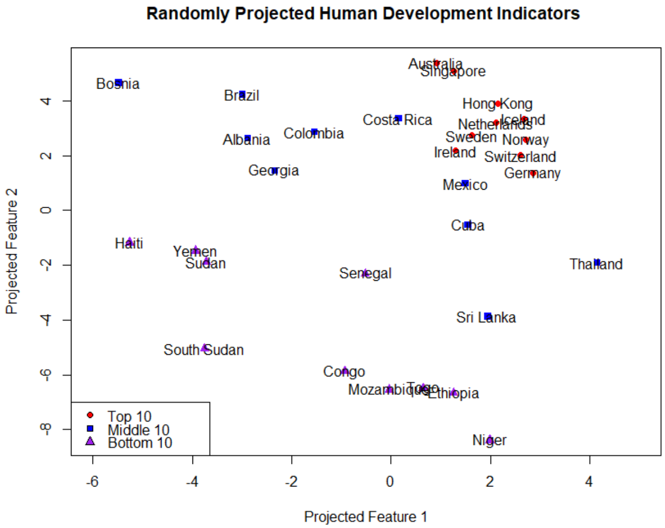
Can we use random projections to better visualize high dimensional data.

Dataset

- Human Development Indicators from the United Nations for Top 10, Middle 10, Bottom 10 based on Human Development Index
- For each country retrieve most recent data on: dependency ratio, proportion urban population, mean years of school completed, CO^2 per capita, life expectancy, proportion internet users, proportion with basic drinking water, and unemployment rate.
- Normalized Data & Removed NAs

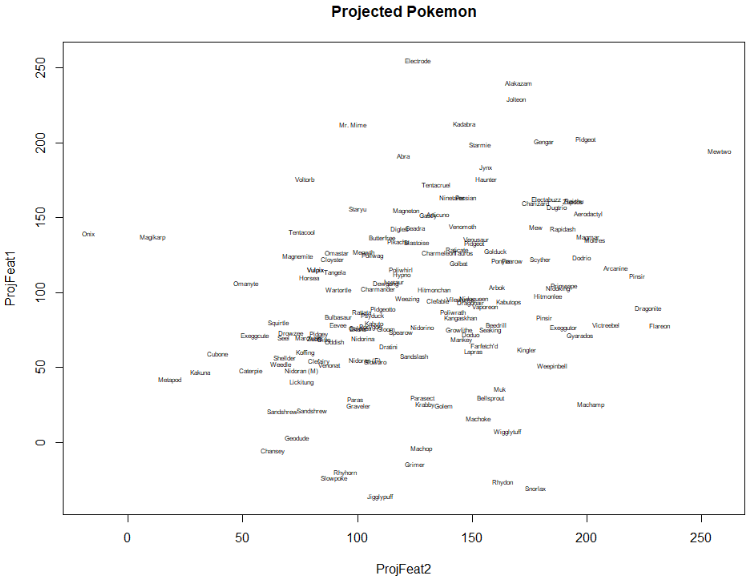
Solution

Randomly project this dataset $(n = 30,\ p = 8)$ to two dimensions and plot.


Randomly Projected Human Development Indicators
Australia
Singapore
Bosnia
Brazil
Hong Kong
Costa Rica
Netherlands
Iceland
Sweden
Norway
Albania
Colombia
Ireland
Switzerland
Germany
Georgia
Mexico
Cuba
Haiti
Yemen
Sudan
Thailand
Senegal
Sri Lanka
South Sudan
Congo
Mozambique
Togo
Ethiopia
Niger
Top 10
Middle 10
Bottom 10
Projected Feature 1
Projected Feature 2
-6
-4
-2
0
2
4
-8
-6
-4
-2
0
2
4

Thank You!

Just for Fun - Visualizing 150 Pokemon Stats (All)



Just for Fun - Visualizing 150 Pokemon Stats (Select)