COMP 585 **Spring 2026**
Probabilistic Toolkit for Learning and Computing *Instructor: Maryam Aliakbarpour*
March 3, 2026 *Scriber: Junze Yin & Maryam Aliakbarpour*

Lecture 14

Johnson– Lindenstrauss lemma

An *embedding* is a transformation that maps vectors from one space to another—typically from a high-dimensional to a low-dimensional space—such that the essential geometric properties are nearly maintained. Embeddings are crucial in many applications, including machine learning, data mining, and signal processing, because they enable efficient computation and storage by reducing the dimensionality of data without significantly distorting its underlying structure.

The Johnson-Lindenstrauss (JL) lemma is a fundamental result in dimensionality reduction, stating that a set of points in a high-dimensional space can be *embedded* into a significantly lower-dimensional space while approximately preserving the *pairwise Euclidean distances*. See Figure 1.

Lemma 1 (Johnson–Lindenstrauss lemma)**.** *Suppose we are given n points in $\mathbb{R}^d$: $u_1, u_2, \ldots, u_n$, and two arbitrary parameters $\epsilon, \delta \in (0, 1]$. For an integer d', if*

$$d' \geq \left\lceil \frac{\log(n^2/\delta)}{2\,\epsilon^2} \right\rceil = \Theta\left(\frac{\log(n/\delta)}{\epsilon^2} \right),$$

then there exists a randomized linear map $F : \mathbb{R}^d \to \mathbb{R}^{d'}$ such that with probability at least $1 - \delta$, for all pairs of i, j in $[n]$, we have:

$$(1-\epsilon)\|u_i - u_j\|_2^2 \leq \|F(u_i) - F(u_j)\|_2^2 \leq (1+\epsilon)\|u_i - u_j\|_2^2.$$

Application: k-Means Clustering with Dimension Reduction

k-means clustering is a fundamental unsupervised learning algorithm that aims to partition n data points into k distinct clusters. It achieves this by iteratively assigning each data point to its nearest cluster centroid and subsequently updating the centroids to minimize the within-cluster variance. It turns out this problem is equivalent to seeking a partition $S = \{s_1, s_2, \ldots, s_k\}$ that minimizes the sum of squared distances within each cluster:

$$\arg\min_{S} \sum_{s_i \in S} \sum_{x,y \in s_i} \|x - y\|_2^2.$$

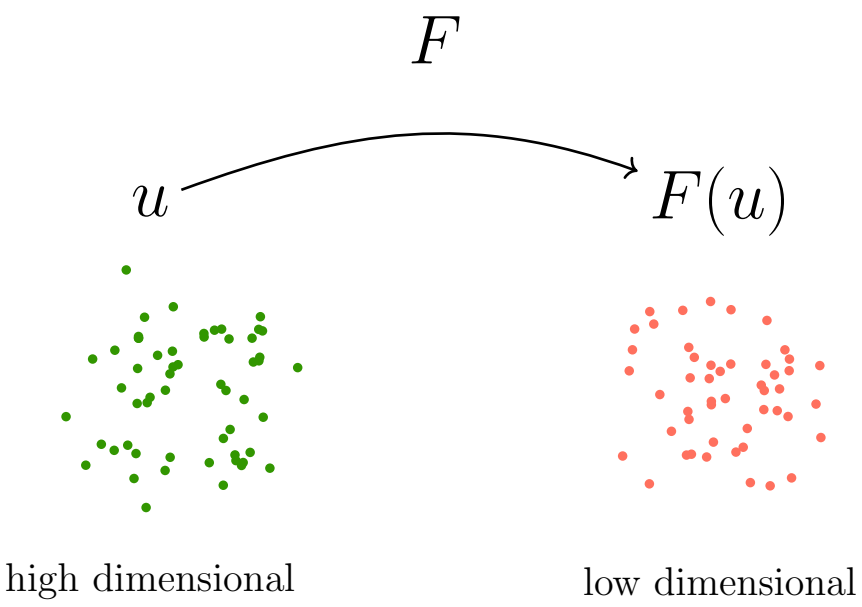

Figure 1: Dimension reduction: F maps high-dimensional points like u (green) to lower-dimensional points $F(u)$ (red), with $\dim F(u) \ll \dim u$.

While this clustering problem is known to be NP-hard, approximation algorithms exist, often with time complexity proportionate to d, the dimension.

In scenarios involving high-dimensional data, directly applying k-means clustering can become computationally prohibitive. To mitigate this challenge, a common strategy is to first embed the data into a lower-dimensional space, while approximately preserving the pairwise Euclidean distances between the data points. Subsequently, the approximation algorithms are applied to the data points within this reduced-dimensional space. Since the dimensionality reduction step preserves the essential structural information of the data, specifically the Euclidean distance, the quality of the clustering is maintained. However, by operating in a lower dimension, the approximation algorithms execute significantly faster, thereby enhancing the overall efficiency of the clustering process.

Proposed Embedding

Suppose we have a $d' \times d$ matrix $M \in \mathbb{R}^{d' \times d}$ such that every entry of M is drawn from a normal distribution $\mathcal{N}(0, \frac{1}{d'})$, namely for all $i \in [d']$ and $j \in [d]$, we have

$$M_{ij} \sim \mathcal{N}\left(0, \frac{1}{d'}\right) .$$

We define $F(u)$ to be:

$$F(u) \coloneqq Mu\,.$$

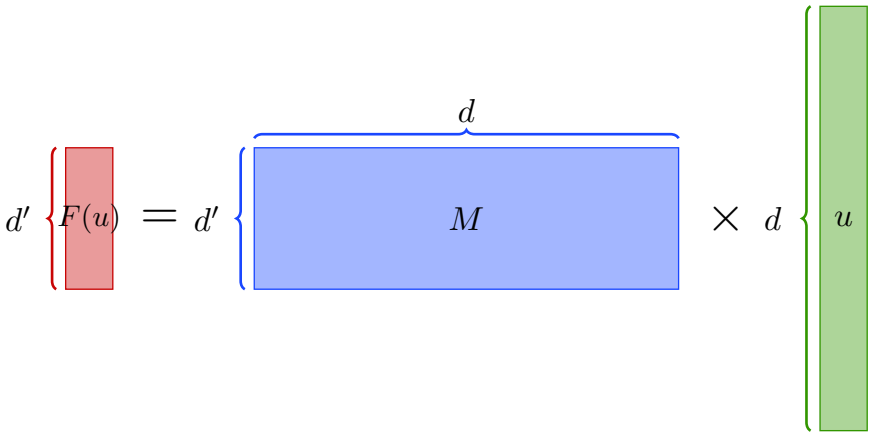

Figure 2: A linear embedding $F(u) = Mu$ with $u \in \mathbb{R}^d$, $M \in \mathbb{R}^{d' \times d}$, and $F : \mathbb{R}^d \to \mathbb{R}^{d'}$.

Proof of JL-Lemma

The proof of the Johnson-Lindenstrauss Lemma is achieved by leveraging concentration inequalities for sub-exponential random variables applied to the squared norms of the projected vectors.

Proof. We prove the lemma in the following steps:

Step 1: Entries of $F(v)$ for a unit vector v: We begin by considering a fixed unit vector v in $\mathbb{R}^d$ with $\|v\|_2 = 1$ and focus on the entries of $F(v)$. Let M_i denote the i-th row of the projection. Note that each entry $M_{ij} \sim \mathcal{N}(0, \frac{1}{d'})$. We claim that the inner product

$$Z_i := \langle M_i, v \rangle = \sum_{j=1}^{d} M_{ij} v_j \,.$$

is a Gaussian random variable.

Since Z_i is a linear combination of independent Gaussian variables, it follows that Z_i is a Gaussian random variable. The mean of Z_i is zero. And, we compute the variance:

$$\begin{aligned} \mathbf{Var}[Z_i] = \mathbf{Var}\left[\sum_{j=1}^{d} M_{ij} v_j\right] &= \sum_{j=1}^{d} v_j^2 \cdot \mathbf{Var}[M_{ij}] && \text{(using independence of } M_{ij}\text{'s)} \\ &= \frac{\|v\|_2^2}{d'} = \frac{1}{d'} \,. && \text{(using } \|v\|_2^2 = 1\text{)} \end{aligned}$$

Thus, we conclude:

$$\langle M_i, v \rangle \sim \mathcal{N}(0, 1/d') \,.$$

Step 2: ℓ_2-norm of $F(v)$ Next, we show that for a unit vector v, then $\|F(v)\|_2^2$ is a sub-exponential random variable. Note that, by the definition of the ℓ_2 norm, we have:

$$\|F(v)\|_2^2 = \|Mv\|_2^2 = \sum_{i=1}^{d'} \langle M_i, v\rangle^2 .$$

From Step 1, we know that $\langle M_i, v\rangle \sim \mathcal{N}(0, 1/d')$. Hence, $\langle M_i, v\rangle$ has the same distribution as $Z_i/\sqrt{d'}$, where Z_i is a standard normal distribution:

$$\langle M_i, v\rangle \stackrel{d}{=} \frac{1}{\sqrt{d'}} Z_i, \qquad \text{where } Z_i \sim \mathcal{N}(0,1) .$$

Thus, we have

$$\|F(v)\|_2^2 = \sum_{i=1}^{d'} \frac{Z_i^2}{d'}.$$

Note that in a previous lecture, we have shown that

$$\mathbf{E}\left[\exp\left(Z_i^2 - 1\right)\right] \le e^{2\lambda^2} \qquad \text{for } \lambda \text{ with } |\lambda| \le \frac{1}{4} .$$

Clearly, this implies the sub-exponentiality of Z_i^2, and in particular $\mathsf{subE}\left((2,4)\right)$. We aim to use Bernstein's inequality, for which we need zero-mean, sub-exponential random variables in $\mathsf{subE}\left((1,1)\right)$. Hence, we work with the centered and scaled version of this random variable as defined below:

$$W_i := \frac{Z_i^2 - 1}{4} .$$

Clearly, the W_i's have zero mean. In addition, by the scaling property of sub-exponentials, W_i is in $\mathsf{subE}\left(2/16, 1\right) \subseteq \mathsf{subE}\left((1,1)\right)$.

Now, using Bernstein's inequality (See Theorem 1 in this lecture note), we obtain:

$$\begin{aligned} \mathbf{Pr}\left[\left|\|F(v)\|_2^2 - 1\right| > \epsilon\right] =& \mathbf{Pr}\left[\left|\frac{1}{d'}\sum_{i=1}^{d'} Z_i^2 - 1\right| > \epsilon\right] = \mathbf{Pr}\left[\left|\frac{1}{d'}\sum_{i=1}^{d'} W_i\right| > 4\,\epsilon\right] \\ &\le 2\exp\left(-d' \min\left(\frac{(4\,\epsilon)^2}{2},\ \frac{4\,\epsilon}{2}\right)\right) \le 2\,\exp\left(-2\,d'\,\epsilon^2\right) , \end{aligned} \tag{1}$$

where in the last inequality we use the fact that $\epsilon \in (0, 1]$.

Step 3: Generalizing to any non-zero vectors u Next, we show that for a vector u that is non-zero, then $\|F(u)\|_2^2$ satisfies a similar tail bound. Define $v := \frac{u}{\|u\|_2}$ so that $\|v\|_2^2 = 1$. We have:

$$\frac{1}{\|u\|_2^2} \cdot F(u) = \frac{1}{\|u\|_2^2} \cdot M \cdot u = M \cdot v = F(v)\,.$$

Therefore, for all $\epsilon \in (0, 1]$, we have:

$$\mathbf{Pr}\left[\left|\frac{\|F(u)\|_2^2}{\|u\|_2^2} - 1\right| > \epsilon\right] = \mathbf{Pr}\big[\big|\|F(v)\|_2^2 - 1\big| > \epsilon\big] \leq \; 2\exp\left(-2\,d'\,\epsilon^2\right)\,,$$

where the inequality follows from Equation (1).

Step 4: Proof of the JL Lemma Our goal is to show that $\|F(u_i) - F(u_j)\|_2^2$ remains within a factor of $(1 \pm \epsilon)$ of $\|u_i - u_j\|_2^2$. More precisely, we need to show:

$$\mathbf{Pr}\big[\exists i, j \text{ such that } \|F(u_i) - F(u_j)\|_2^2 \notin \left[(1-\epsilon)\cdot\|u_i - u_j\|_2^2\,,\;\; (1+\epsilon)\cdot\|u_i - u_j\|_2^2\right]\big] \leq \delta.$$

Fix any $i, j \in [n]$. If $u_i = u_j$, then we have $F(u_i) = F(u_j)$, and the statement is trivial. Assume $u_i \neq u_j$, and define: $u_{ij} \coloneqq u_i - u_j$. Clearly, due to linearity of matrix multiplication, we see that:

$$F(u_{ij}) = Mu_{ij} = \; F(u_i) - F(u_j).$$

Using Step 3, we get:

$$\begin{aligned}
&\mathbf{Pr}\big[\|F(u_i) - F(u_j)\|_2^2 \notin \left[(1-\epsilon)\cdot\|u_i - u_j\|_2^2\,,\;\; (1+\epsilon)\cdot\|u_i - u_j\|_2^2\right]\big] \\
&\quad= \; \mathbf{Pr}\left[\left|\frac{\|F(u_i) - F(u_j)\|_2^2}{\|u_i - u_j\|_2^2} - 1\right| > \epsilon\right] \\
&\quad= \; \mathbf{Pr}\left[\left|\frac{\|F(u_{ij})\|_2^2}{\|u_{ij}\|_2^2} - 1\right| > \epsilon\right] \;\leq\; 2\exp\left(\exp\left(-2\,d'\,\epsilon^2\right)\right)\,.
\end{aligned}$$

Applying the union bound over all $\binom{n}{2}$ pairs, we have:

$$\begin{aligned}
&\mathbf{Pr}\big[\exists i, j \text{ such that } \|F(u_i) - F(u_j)\|_2^2 \notin \left[(1-\epsilon)\cdot\|u_i - u_j\|_2^2\,,\;\; (1+\epsilon)\cdot\|u_i - u_j\|_2^2\right]\big] \\
&\quad\leq 2\binom{n}{2}\,\exp\left(-\frac{d'\epsilon^2}{2}\right) \leq n^2\,\exp\left(-2\,d'\,\epsilon^2\right)\,.
\end{aligned}$$

To ensure this is at most δ, we set:

$$d' \geq \left\lceil\frac{\log(n^2/\delta)}{2\,\epsilon^2}\right\rceil = \Theta\left(\frac{\log(n/\delta)}{\epsilon^2}\right)\,.$$

This completes the proof of the Johnson-Lindenstrauss Lemma. □

Optimality of JL lemma: Larsen and Nelson have shown the optimality of the JL lemma. In particular, they have shown there exists a set of n points in the space that any embedding that preserves their pairwise distances must have $d' \geq (\log n)/\epsilon^2$ [LN17].

Bibliographic Note

The JL lemma was originally proved in [JLS86]. The content of this lecture was derived from Section 5.3 of [Ver18], and the lecture notes of Prof. Sasha Rakhlin for "Mathematical Statistics: A Non-Asymptotic Approach", which can be found here [Rak22].

References

[JLS86] William B Johnson, Joram Lindenstrauss, and Gideon Schechtman. Extensions of lipschitz maps into banach spaces. *Israel Journal of Mathematics*, 54(2):129–138, 1986.

[LN17] Kasper Green Larsen and Jelani Nelson. Optimality of the johnson-lindenstrauss lemma. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 633–638. IEEE, 2017.

[Rak22] Alexander Rakhlin. Mathematical statistics: A non-asymptotic approach, 2022. Lecture notes for MIT course IDS.160, Spring 2022.

[Ver18] Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge University Press, 2018.