Optimization-Based Data Analysis

Recitation 4

1. Let $\vec{x}, \vec{y} \in \mathbb{R}^n$ and let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be a random matrix with iid standard Gaussian entries. What is the distribution of $\vec{x}^T \mathbf{A} \vec{y}$?

   Solution. Note that
   $$\vec{x}^T \mathbf{A} \vec{y} = \sum_{i,j} \vec{x}[i] \vec{y}[j] \mathbf{A}_{i,j}.$$
   This is a Gaussian random variable with mean 0 and variance $\sum_{i,j} \vec{x}[i]^2 \vec{y}[j]^2$. The variance can also be written as $\|\vec{y}\vec{x}^T\|_F^2$.

2. Show how to generate a vector uniformly at random from the sphere $S^{n-1} \subseteq \mathbb{R}^n$.

   Solution. Let $\vec{\mathbf{x}} \sim \mathcal{N}(\vec{0}, I)$ have dimension n and let $\vec{\mathbf{y}} = \vec{\mathbf{x}}/\|\vec{\mathbf{x}}\|_2$. Then $\vec{\mathbf{y}}$ is uniformly distributed on S^{n-1} since $\vec{\mathbf{x}}$ is isotropic.

3. Let $\mathbf{A}$ be a random $m \times n$ matrix with iid standard Gaussian entries. Give the means and variances of the entries of $\mathbf{A}^T \mathbf{A}$. Let $\mathbf{B} = \frac{1}{\sqrt{m}} \mathbf{A}$. What does this say about $\mathbf{B}$ as $m \to \infty$?

   Solution. $E[A_{ii}] = m$, $E[A_{ij}] = 0$, $\mathrm{Var}[A_{ii}] = 2m$, and $\mathrm{Var}[A_{ij}] = m$, for $i \neq j$. We expect the columns of $\mathbf{B}$ to be approximately orthonormal. This can also be seen using the fact that $\mathbf{B}^T \mathbf{B}$ is an estimator for the covariance matrix of $\mathbf{A}$, i.e., the identity.

4. Let $\mathbf{A}$ be an $m \times n$ random matrix with iid standard Gaussian entries. Under what conditions will it have rank $\min(m, n)$ with high probability?

   Solution. In all cases it will have rank $\min(m, n)$ with probability 1. Assume $m \geq n$ and let $\vec{x}_1, \ldots, \vec{x}_n$ denote the columns of $\mathbf{A}$. We will show the probability that $\vec{x}_k$ lies in the span of $\vec{x}_1, \ldots, \vec{x}_{k-1}, \vec{x}_{k+1}, \ldots, \vec{x}_{n-1}$ is zero. Then by the union bound over k the result follows. By symmetry, let's assume $k = n$, and define E to be the event that $\vec{x}_n$ is in the span of $\vec{x}_1, \ldots, \vec{x}_{n-1}$. Then we have
   $$\begin{aligned}
   \int_E p(\vec{x}_1) \cdots p(\vec{x}_n)\, d\vec{x}_1 \cdots d\vec{x}_n &= \int p(\vec{x}_1) \cdots p(\vec{x}_n)\, \mathbf{1}_E \; d\vec{x}_1 \cdots d\vec{x}_n \\
   &= \int_{\vec{x}_1, \ldots, \vec{x}_{n-1}} p(\vec{x}_1) \cdots p(\vec{x}_{n-1}) \int_{\vec{x}_n} p(\vec{x}_n)\, \mathbf{1}_E \; d\vec{x}_n \, d\vec{x}_1 \cdots d\vec{x}_{n-1} \\
   &= 0,
   \end{aligned}$$
   since for $\vec{x}_1, \ldots, \vec{x}_{n-1}$ fixed we have
   $$\int_{\vec{x}_n} p(\vec{x}_n)\, \mathbf{1}_E \; d\vec{x}_n = \int_E p(\vec{x}_n)\, d\vec{x}_n = 0.$$

This is the probability that a length m Gaussian vector lies in a subspace of of $\mathbb{R}^m$ of dimension strictly less than m. In other words, this a set of zero volume, and thus has zero probability (since the Gaussian distribution has a density).

5. Suppose you have two datasets of p vectors in $\mathbb{R}^n$ for two different data science problems you are analyzing: $\vec{x}_1, \ldots, \vec{x}_p$ and $\vec{y}_1, \ldots, \vec{y}_p$. You would like to apply J-L to each separately to randomly project to a lower dimensional space with some fixed accuracy parameter $\epsilon \in (0,1)$. Assuming you use the version of J-L from the homework that chooses the dimensions of $\mathbf{A}$ to give success with probability α on a dataset of size p, answer the following:

   (a) Give a bound on the probability of success if you use the same random matrix $\mathbf{A}$ for both datasets.

   (b) Give a bound on the probability of success if you use two independent matrices for the datasets.

   Solution.

   (a) In the proof of J-L from the homework we obtain

   $$\bigcup_{i,j} \mathbb{P}(\mathcal{E}_{i,j}^c) \leq 1 - \alpha.$$

   If there are two datasets, we must union over twice as many events and get a bound of $2(1-\alpha)$ giving an answer of $2\alpha - 1$.

   (b) Since the events are independent, we obtain a probability of α^2. This is larger than $2\alpha - 1$ since $\alpha^2 - (2\alpha - 1) = (1-\alpha)^2$.

6. Suppose $A \in \mathbb{R}^{m \times n}$ has $m \ll n$ and is approximately rank k. Give an efficient algorithm for computing the SVD of A.

   Solution. Apply the randomized SVD algorithm from the notes to A^T.