Recitation 3

DS-GA 1013 Mathematical Tools for Data Science

1. 1. Let $u, v \in \mathbb{R}^n$ and $A = uv^T$. What is the rank of A? Find eigenvalues and eigenvectors of A.
   2. If the eigenvalues and eigenvectors of a symmetric matrix $A \in \mathbb{R}^{n \times n}$ are $\lambda_1, \lambda_2, \ldots, \lambda_n$ and $u_1, u_2, \ldots u_n$ respectively, then what are the eigenvalues and eigenvectors of $I + 2\alpha A$?
   3. Let $\mathbf{x}$ be a d-dimensional vector with mean $\mu \in \mathbb{R}^d$ and let the covariance matrix be $\Sigma_{\mathbf{x}}$. Can you bound the largest eigenvalue of the matrix $E(\mathbf{x}\mathbf{x}^T)$?

Solution:

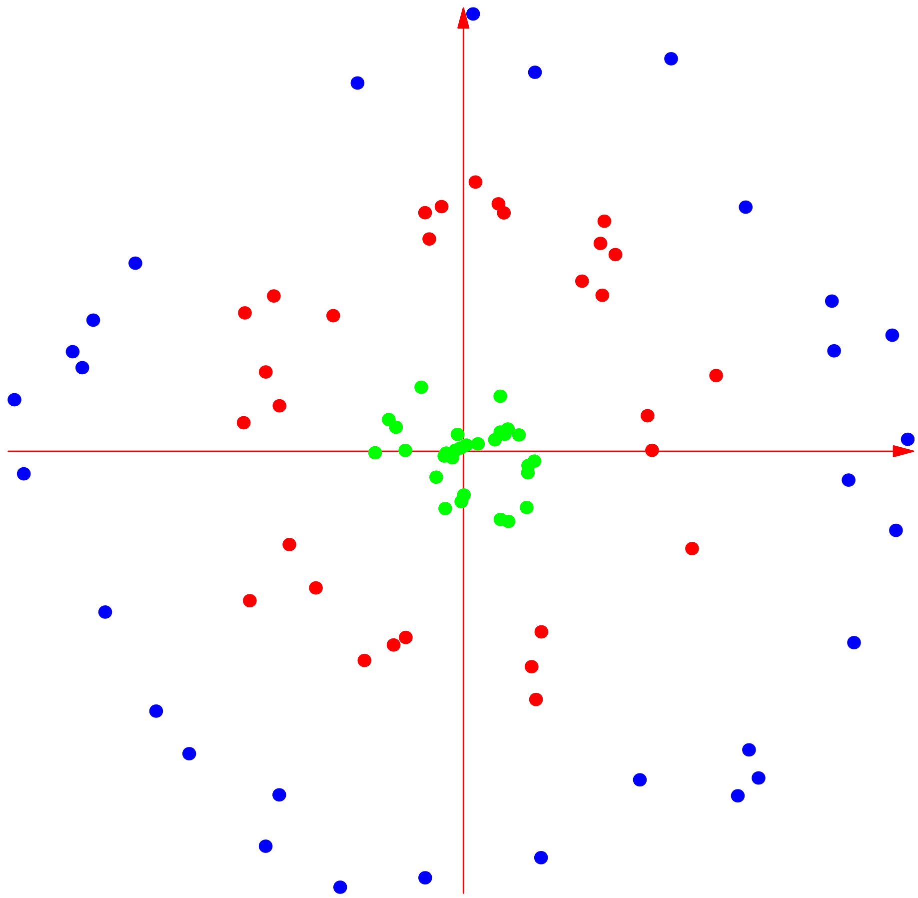
1. A has rank 1.
$$Au = uv^T u = (v^T u)u$$
2. $1 + 2\alpha\lambda_i$
3. From homework:
$$\begin{aligned} E(\mathbf{x}\mathbf{x}^T) &= \Sigma_{\mathbf{x}} + \mu\mu^T \\ \max_{v, ||v||_2=1} v^T E(\mathbf{x}\mathbf{x}^T) v &= \max_{v, ||v||_2=1} v^T (\Sigma_{\mathbf{x}} + \mu\mu^T) v \\ &\leq \max_{v, ||v||_2=1} v^T (\Sigma_{\mathbf{x}} v + \max_{v, ||v||_2=1} v^T \mu\mu^T) v \\ &\leq \lambda_1 + \mu^T\mu \end{aligned}$$
where λ_1 is the largest eigenvalue of Σ_x and the next term is derived by applying first part here.

2. 1. True or False: If you are already working with features that have been normalized to have variance 1, there is no need to whiten your data.
   2. Let Σ_n be the covariance matrix of your data after you've normalized features to have variance 1. Let Σ be the covariance matrix of unnormalized data. Can we say any relationship between the eigenvalues and eigenvectors of both the matrices?
   3. What are the principal directions and variations along principle directions if you perform PCA on whitened data?

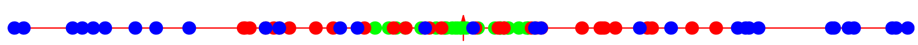
Solution:

1. False. The covariance matrix $\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$ is standardized but not whitened with singular values 1.5 and 0.5.
2. We cannot derive any concrete relationships in general - the eigenvalues and eigenvectors will be different. However, the interpretation of eigenvectors from correlation matrix and covariance matrix is very different. See here for a discussion.
3. All directions are equally good.

3. Given the following dataset determine how to use PCA to get meaningful principal components.



Solution: If we apply standard PCA we obtain the following.



If we add the feature $\|\vec{x}\|^2$ and then perform PCA we obtain the following.

4. 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 $\|\vec{y}\vec{x}^T\|_F^2$.

5. 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.

6. 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.

7. 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 $\vec{x}_1, \ldots, \vec{x}_n$ $\vec{x}_k$ 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).