QUESTION 1

(a) Let $\{x_1, \ldots, x_n\} \subset \mathbb{R}^D$ be data with mean 0. Write the formula for the $D \times D$ covariance matrix Σ.

(b) Prove that for all $y \in \mathbb{R}^{D\times 1}, y^T \Sigma y \geq 0$.

(c) Is Σ necessarily invertible? Prove or give a counterexample.

(a). $\Sigma = \frac{1}{n} \sum_{i=1}^{n} x_i^T x_i$, where $x_i \in \mathbb{R}^{D\times 1}$ is understood as a column vector

(b). Proof: Let $y \in \mathbb{R}^{D\times 1}$ be arbitrary. Then:

$$y^T \Sigma y$$

$$= \frac{1}{n} y^T \left(\sum_{i=1}^{n} x_i^T x_i\right) y$$

$$= \frac{1}{n} \sum_{i=1}^{n} y^T (x_i^T x_i) y$$

$$= \frac{1}{n} \sum_{i=1}^{n} \underbrace{y^T x_i^T}_{= \langle x_i, y \rangle} \underbrace{x_i y}_{= \langle x_i, y \rangle}$$

Note: $y^T x_i^T = (x_i y)^T$
$= x_i y$
$= \sum_{j=1}^{D} [x_i]_j y_j$

$$= \frac{1}{n} \sum_{i=1}^{n} |\langle x_i, y \rangle|^2$$

≥ 0. ∎

(c.) False: counterexample is given by any data that has a covariance matrix with a 0 eigenvalue. For example, if $n = 1$ and x_1 is anything.

Question 2

Suppose $\{x_1, \ldots, x_n\} \subset \mathbb{R}^D$ is formed into a data matrix $X \in \mathbb{R}^{n\times D}$ with the i^{th} row corresponding to x_i.

(a) Suppose X has singular value decomposition $X = U\Lambda V^T$, where $UU^T = I, VV^T = I$, and Λ is a diagonal matrix with diagonal entries $\sigma_1 > \sigma_2 > \cdots > \sigma_n \geq 0$ (note the strict inequality between successive singular values). In terms of this decomposition, what is the direction of maximum variance in the data? Explain.

(b) In terms of the singular value decomposition, what proportion of the variance in the data is retained when projecting onto the first principal component? Explain.

(a.) We note that if we consider $X^T X$

$$= \sum_{i=1}^{n} X_i X_i^T,$$ with x_i understood as a column vector,

then $X^T X$ is (up to a scalar) the covariance matrix. Notice moreover that by applying the SVD to X, $X^T X$

$$= [U\Lambda V^T]^T [U\Lambda V^T]$$

$$= V\Lambda^T U^T U \Lambda V^T$$

$$= V(\Lambda^T \Lambda)V^T \quad , \text{ Since } U^T U = I \text{ by } \text{~~under~~ orthogonality}$$

So, the covariance matrix of X^{th} has eigenvectors given by the right singular vectors of X and eigenvalues given by the singular values of X squared. In particular, the direction of maximum variance is V_1.

(b.) We know the proportion of variance in the direction of maximum variance is

$$\frac{\lambda_1}{\sum_{j=1}^{D} \lambda_j},$$ where $\{\lambda_j\}_{j=1}^{D}$ are the eigenvalues of $X^T X$. In terms of SVD, it is $\frac{\sigma_1^2}{\sum_{j=1}^{D} \sigma_j^2}$.

QUESTION 3

Let $\{\theta_i\}_{i=0}^{n}$ be an equi-spaced partition of $[0, 2\pi]$, i.e. $\theta_i = i(2\pi/n)$.

(a) What visual shape do the data $\{(\cos(\theta_i), \sin(\theta_i))\}_{i=0}^{n}$ form when plotted in $\mathbb{R}^2$?

(b) Compute the covariance matrix of the data $\{(\cos(\theta_i), \sin(\theta_i))\}_{i=0}^{n} \subset \mathbb{R}^2$.

(c) Recall that for a continuous function $f(\theta)$ on $[0, 2\pi]$, the *Riemann integral* of $f(\theta)$ may be computed using *Riemann sums* as

$$\frac{1}{2\pi}\int_0^{2\pi} f(\theta)d\theta = \lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n} f(\theta_i),$$

where $\{\theta_i\}_{i=0}^{n}$ is as above. Use this fact (without proving it) to argue that as $n \to \infty$, the covariance matrix in (b) becomes of the form $\begin{pmatrix}\alpha & 0\\ 0 & \alpha\end{pmatrix}$, for some $\alpha > 0$.

(d) Part (c) shows that, as $n \to \infty$, the covariance matrix has two eigenvectors with equal eigenvalues. Does this make sense, given your answer in (a)? Can you explain the conclusion of (c) geometrically?

(a.) Circle

(b.) $\Sigma = \frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}\cos(\theta_i)\\ \sin(\theta_i)\end{pmatrix}\begin{pmatrix}\cos(\theta_i) & \sin(\theta_i)\end{pmatrix}$

$= \frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}\cos^2(\theta_i) & \sin(\theta_i)\cos(\theta_i)\\ \sin(\theta_i)\cos(\theta_i) & \sin^2(\theta_i)\end{pmatrix}$

(c.) Note that $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sin(\theta_i)\cos(\theta_i)$

$= \frac{1}{2\pi}\int_0^{2\pi} \sin(\theta)\cos(\theta)\,d\theta$

$= 0$ by symmetry or by using u-sub

On the other hand, $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sin^2(\theta_i) = \frac{1}{2\pi}\int_0^{2\pi}\sin^2(\theta)d\theta > 0$,

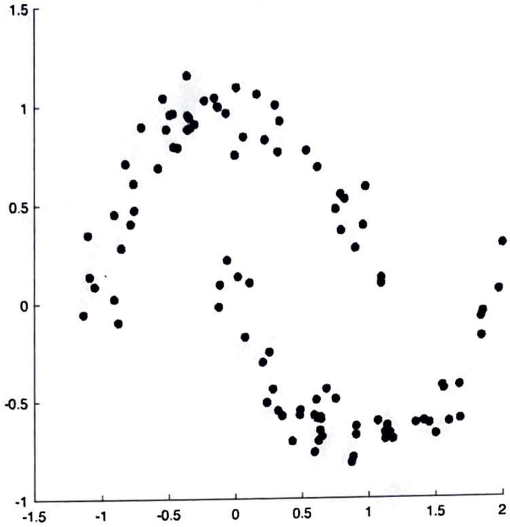
Same with $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\cos^2\theta_i$. The result follows by plugging into (b).

(d.) $\begin{pmatrix}\alpha & 0\\ 0 & \alpha\end{pmatrix}$ has two eigenvectors $(1,0), (0,1)$ with eigenvalues α. Makes sense — circle has no dominant direction.

Question 4

(a) Let $x_1, \dots, x_n \subset \mathbb{R}^D$ be data to be clustered. Write down the functional that the K-means clustering algorithm attempts to minimize.

(b) Consider the "two moons" data. Will K-means with $K = 2$ learn the two moons exactly? Why or why not?



(a.) $F(C_1, \dots, C_K) = \sum_{i=1}^{K} \left(\sum_{x \in C_i} \|x - \bar{x}_i\|^2 \right)$, where $C_1, \dots, C_K$ is a partition of the data and $\bar{x}_i$ is the center of the i^{th} cluster

(b.) There is no choice of centers $(\bar{x}_1, \bar{x}_2)$ s.t. the moons will be separated by spheres with centers $\bar{x}_1$ and $\bar{x}_2$. So no, K-means (even given the computational resources to perfectly minimize F) will not learn the two moons exactly.