Marcus Rose
Math 123 HW 1
Due 1/26/23

---

1. Linearity of Matrix-Vector Multiplication

   Proof To prove linearity, WTS

   (a) $\forall \mathbf{x},\ \mathbf{y} \in \mathbb{R}^{n\times 1}, A(\mathbf{x},\ \mathbf{y}) = A(\mathbf{x}) + A(\mathbf{y})$
   (b) $\forall \mathbf{x} \in \mathbb{R}^{n\times 1}$ and $\forall \alpha \in \mathbb{R}, A(\alpha\mathbf{x}) = \alpha A(\mathbf{x})$.

   Let $A = (A_{ij})_{i,j=1}^{n}$. Then $A\mathbf{x}$ is

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots a_{2n} \\ \vdots & \vdots & \vdots & \ddots \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i = \sum_{j=1}^{n} A_{ij}x_j.$$

   Now consider $\mathbf{x},\ \mathbf{y} \in \mathbb{R}^{n\times 1}$,

$$A(\mathbf{x}+\mathbf{y}) = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots a_{2n} \\ \vdots & \vdots & \vdots & \ddots \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix} = \sum_{i}^{n} \begin{bmatrix} a_{i1} & a_{i2} & \dots & a_{in} \end{bmatrix} \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix}.$$

   As we showed for $A\mathbf{x}$, this can be further expressed as

$$\sum_{i}^{n}\sum_{j}^{n} a_{ij}(x_i + y_i) = \sum_{i}^{n}\sum_{j}^{n} a_{ij}x_i + a_{ij}y_i = \sum_{j}^{n} A_{ij}x_j + \sum_{j}^{n} A_{ij}y_j = A(\mathbf{x}) + A(\mathbf{y}),$$

   with the middle equality derived from the fact that $(Ax)_i = \sum_i^n A_{ij}x_j$.
   To prove (b),

$$A(\alpha\mathbf{x}) = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots a_{2n} \\ \vdots & \vdots & \vdots & \ddots \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} \alpha x_1 \\ \alpha x_2 \\ \vdots \\ \alpha x_n \end{bmatrix} = \sum_{i}^{n}\sum_{j}^{n} a_{ij}(\alpha x_i) = \sum_{j=1}^{n} A_{ij}(\alpha x_j) = \alpha \sum_{j=1}^{n} A_{ij}x_j = \alpha A(\mathbf{x})$$

   with the second-to-last equality derived from scalar linearity. Therefore, matrix-vector multiplication is linear.

   □

2. Suppose $A \in \mathbb{R}^{n\times n}$ is a symmetric. Prove that if A has eigenvalue $\lambda = 0$ with eigenvector $\mathbf{x} \neq \mathbf{0}$, then A is not invertible.

Proof Suppose A has $\lambda = 0$ and $\mathbf{x} \neq \mathbf{0}$. Then, $A\mathbf{x} = 0\mathbf{x} = 0$. If $A\mathbf{x} = 0$, then the columns of A are linearly dependent, meaning there exists a non-trivial solution. Therefore, A is not full rank and hence not invertible.

□

3. Prove $\|\cdot\|_2$ is a norm.

The l^2 norm of $\mathbf{x} \in \mathbb{R}^n$ is defined as

$$\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2}.$$

Proof To prove something is a norm, WTS

(a) $\forall \mathbf{x} \in \mathbb{R}^n, \|\mathbf{x}\|_2 \geq 0$

(b) $\|\mathbf{x}\|_2 = 0 \iff \mathbf{x} = \mathbf{0}$

(c) $\forall \alpha \in \mathbb{R}, \mathbf{x} \in \mathbb{R}^n \|\alpha\mathbf{x}\|_2 = \alpha\|\mathbf{x}\|_2$

(d) $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $\|\mathbf{x} + \mathbf{y}\|_2 \leq \|\mathbf{x}\|_2 + \|\mathbf{y}\|_2$

Starting with (a),

$$\|\mathbf{x}\|_2 \geq 0 \implies \sqrt{\sum_{i=1}^{n} |x_i|^2} \geq 0 \implies \sum_{i=1}^{n} |x_i|^2 \geq 0 \implies |x_1|+|x_2|+\ldots+|x_n| = |x_1+x_2+\ldots+x_n| \geq 0$$

for $x_i \in \{0, \mathbb{R}^+\}$. Because a sum of the absolute value of real-valued numbers cannot be negative, (a) must be true. Similarly for (b) and proving the forward direction first,

If $\|\mathbf{x}\|_2 = 0$, then $\mathbf{x} = \mathbf{0}$. Then,

$$\|\mathbf{x}\|_2 = 0 \implies \sqrt{\sum_{i=1}^{n} |x_i|^2} = 0 \implies \sum_{i=1}^{n} |x_i| = 0 \implies |x_1| + |x_2| + \ldots + |x_n| = 0,$$

which is only possible if $x_1 = x_2 = \ldots = x_n = 0$. Next proving the opposite direction,

If $\mathbf{x} = \mathbf{0}$, then $\|\mathbf{x}\|_2 = 0$.

$$\mathbf{x} = \mathbf{0} \implies x_1 = x_2 = \ldots = x_n = 0 \implies \sum_{i=1}^{n} |x_i|^2 = 0,$$

or the sum of a set of zero-valued numbers is necessarily 0, which is exactly what we needed to show. Now proving (c),

$$\|\alpha\mathbf{x}\|_2 = \sqrt{\sum_{i}^{j} |\alpha x_i|^2} = |\alpha| \sqrt{\sum_{i}^{j} |x_i|^2} = |\alpha| \|\mathbf{x}\|_2,$$

with the second equality from the fact that constants are linear. Finally proving (d), note that

$$\|\mathbf{x}+\mathbf{y}\|_2 = \|\mathbf{x}\|_2^2 + 2(\mathbf{x}\cdot\mathbf{y}) + \|\mathbf{y}\|_2^2 \leq \|\mathbf{x}\|_2^2 + 2\|\mathbf{x}\|_2\dot{\|}\mathbf{y}\|_2 + \|\mathbf{y}\|_2^2 = (\|\mathbf{x}\|_2 + \|\mathbf{y}\|_2)^2,$$

with the above inequality true by the Cauchy-Schwarz Inequality. Therefore,

$$(\|\mathbf{x}+\mathbf{y}\|_2)^2 \leq (\|\mathbf{x}\|_2 + \|\mathbf{y}\|_2) \implies \|\mathbf{x}+\mathbf{y}\|_2 \leq \|\mathbf{x}\|_2 + \|\mathbf{y}\|_2$$

□

4. Given $(x_1^1, x_1^2), (x_2^1, x_2^2) \subset \mathbb{R}^2$ sampled from the line $x_2 = \alpha x_1$ for $\alpha \in \mathbb{R}$. Prove that the empirical covariance matrix is rank of at most 1, i.e. one of its eigenvalues is necessarily 0.

   Proof Variance is a measure of variation of a single variable. Covariance measures how much two variables vary with respect to one another. For two random variables X^1 and X^2 that sample from $x^1, x^2 \in \mathbb{R}^2$, respectively, the covariance is defined as

   $$Cov(X^1, X^2) = \mathbb{E}[(X^1 - \mu_{X^1})(X^2 - \mu_{X^2})],$$

   where μ is the respective sample mean for each of the random variables and $\mathbb{E}$ is the expectation. The variance-covariance matrix Σ in two dimensions becomes,

   $$\Sigma = \begin{bmatrix} Var(X^1) & Cov(X^1, X^2) \\ Cov(X^2, X^1) & Var(X^2) \end{bmatrix}.$$

   To prove that this matrix has rank 1, we must show that it has only one linearly independent column (which is also the number of linearly independent rows). By contradiction, assume the matrix has rank 2, that is, the matrix in two dimensions is full rank. This means that each of the columns C_1, C_2 in the space are linearly independent. In other words, there exists scalars $a, b \in \mathbb{R}$ such that

   $$C_1 = aVar(X^1) + bCov(X^1, X^2)$$
   $$C_2 = aCov(X^1, X^2) + bVar(X^2)$$

   and only the trivial solution exists, i.e. $a = b = 0$.

   Taking this determinant yields

   $$|\Sigma| = ab(Cov(X^1, X^2) - Var(X^2)^2) = 0$$

   because we assumed each of the columns fully spans the matrix. But the determinant is 0, which is not possible with a matrix that is full rank. In other words, there is a column of the matrix that is already fully determined. Geometrically in two dimensions, you are squishing all of the space executed from the linear transformation into a smaller dimension. This means that the matrix can have at most a rank of 1.

   □

5. The goal of the final question is to compute the two principal components of a circle and an ellipse from the definition in terms of the covariance matrix. Both the circle and the ellipse take uniformly sampled points, and the following steps will be used to solve for both:

   (a) Mean center the data

(b) Calculate covariance matrix

(c) Find the eigenvalues of the covariance matrix

(d) Sort eigenvalues by size

(e) Select the top n eigenvectors, where n is the number of dimensions to project onto

(f) Use the n projections above to define a new coordinate system and project the data onto a new subspace by dotting the centered data with the eigenvectors

Below is a sample of my code:

```
1
2  % Mean center  the data
3  %   This is done by subtracting the univariate means from the two columns,
4  %   i.e. the two variables of data
5
6  %
7  data_c = X_Circle;
8  data_c(:,1) = data_c(:,1) - mean(data_c(:,1));
9  data_c(:,2) = data_c(:,2) - mean(data_c(:,2));
10 display(data_c)
11
12 % Calculate covariance matrix
13 %
14 %
15 %
16 C = cov(data_c);
17 display(C);
18
19 % Find the eigenvalues of the covariance matrix
20 %   det(cov - B * E)
21 %
22 %
23 [A, B] = eig(C);
24 % display([A, B]);
25 fprintf("A: %d, \t\t B: %d\n", [A,B]);
26
27 n1 = norm(A(:,1));
28 n2 = norm(A(:,2));
29 display(dot(n1, n2));
30
31 % Calculate data with the new coordinates
32 %
33 %
34 %
35
36 newdata = A' * data_c';
37 newdata = newdata';
38 newdata = fliplr(newdata);
39
40 newdata
41
42 % Calculate relative variance for each new var
43 %
44 %
45 %
46 new_var = var(newdata);
47 ans = var(newdata)/sum(var(newdata));
48 disp(new_var);
49 disp(ans);
50
51 variance = B / sum(B(:));
```

```
52 disp(variance);
```

Listing 1: Code for Circle

The code for both is roughly the same, so the output covariance matrices for both is outputted below, along with what the actual principal components are using the MATLAB built-in pca function.

Circle Covariance (from code)= $\begin{pmatrix} 0.4870 & 0 \\ 0 & 0.5130 \end{pmatrix}$. Circle Covariance (from built-in pca) = $\begin{pmatrix} 51.6018 \\ 48.3982 \end{pmatrix}$.

Ellipse Covariance (from code) = $\begin{pmatrix} 0.0583 & 0 \\ 0 & 0.9417 \end{pmatrix}$. Ellipse Covariance (from built-in pca) = $\begin{pmatrix} 94.0930 \\ 5.9070 \end{pmatrix}$.

Note the results of the manually computed principal components are found in the form of a Covariance Matrix while the built-in MATLAB pca function provides a column vector corresponding to the non-zero entries of the covariance matrix.

Note that each principal component is a linear combination of the original features. The first principal component has the most variance. The second PC has the second highest variance (direction orthogonal to the dimension of the first component generally). Note the eigenvectors: in the ellipse, the eigenvectors relate to the major and minor axes of the ellipse. In the circle, the eigenvectors are the tangent points to the point of maximum variance.

To see further details of the code, please see my GitHub.