Exercise 1.4.69

Let v be a vector whose entries represent some statistical data. For example, v could be a vector with 365 entries representing the high temperature in Seattle on 365 consecutive days.

We can normalize this vector by computing its mean and subtracting the mean from each of the entries to obtain a new vector with mean zero.

Suppose now we have such a normalized vector. Then the variance of v is

$$\sum_{i=1}^{n} v_i^2$$

.

This nonnegative number gives a measure of the variation in the data.

Notice that if we think of v as a column vector, the variance can be expressed as $v^T v$.

Now let v and w be two vectors with mean zero and variance (normalized to be) one. Then the correlation of v and w is defined to be

$$\sum_{i=1}^{n} v_i w_i = w^T v = v^T w$$

This number, which can be positive or negative, measures whether the data in v and w vary with or against one another. For example, the temperatures in Seattle and Tacoma should have a positive correlation, while the temperatures in Seattle and Hobart, Tasmania, should have a negative correlation.

Now consider k vectors $v_1, ..., v_k$ with mean zero and variance one. The correlation matrix C of the data $v_i, ..., v_k$ is the $k \times k$ matrix whose (i, j) entry is $c_{ij} = v_j^T v_i$ the correlation of v_i with v_j.

(a) Show that C is a symmetric matrix whose main-diagonal entries are all ones.

(b) Show that $C = V^T V$ for some appropriately constructed (nonsquare) matrix V.

(c) Show that C is positive definite if the vectors $v_i, ..., v_k$ are linearly independent.

(d) Show that if $v_i, ..., v_k$ are linearly dependent, then C is not positive definite, but it is positive semidefinite, i.e. $x^T C x \geq 0$ for all x.

Answers:

(a.1) C is symmetric

$$c_{ij} = v_j^T v_i$$

Which is a line-column vector multiplication. Which is commutative. So:

$$c_{ij} = v_j^T v_i = v_i^T v_j = c_{ji}$$

$\square$

(a.2) $c_{ii} = 1$

$$c_{ii} = v_i^T v_i$$

But $v_i^T v_i$ is the variance of v. Which we know is 1 because the dataset was normalized to be so.

$$c_{ii} = v_i^T v_i = \mathrm{var}(v) = 1$$

$\square$

(b)

We can define "Matrix Product" as:

$$C = A * B$$
$$c_{ij} = \mathrm{row}_i(a) * \mathrm{col}_j(b)$$

Given that covariance matrix C is:

$$c_{ij} = v_j^T v_i$$

We can arrange a matrix V such that:

$$\begin{aligned}
C &= V^T V \\
c_{ij} &= v_j^T v_i \\
c_{ij} &= v_i^T v_j \\
c_{ij} &= \mathrm{row}_i(V^T) * \mathrm{col}_j(V) \\
c_{ij} &= \mathrm{col}_i(V) * \mathrm{col}_j(V)
\end{aligned}$$

Which gives me that the covariance is just the dot-product of columns of V. So I can arrange the V matrix as:

$$V = \begin{bmatrix} \begin{bmatrix} a_1 \\ b_1 \\ \vdots \end{bmatrix} & \begin{bmatrix} a_2 \\ b_2 \\ \vdots \end{bmatrix} & ... & \begin{bmatrix} a_3 \\ b_3 \\ \vdots \end{bmatrix} \end{bmatrix}$$

□

(c) and (d)

$$\begin{aligned} C &= V^T V \\ x^T C x &= x^T V^T V x \end{aligned}$$

$$\begin{aligned} x^T V^T &= \begin{bmatrix} x_1 a_1 + x_2 a_2 + ... + x_n a_n & x_1 b_1 + x_2 b_2 + ... + x_n b_n & ... \end{bmatrix} \\ V x &= \begin{bmatrix} a_1 x_1 + a_2 x_2 + ... + a_n x_x \\ b_1 x_1 + b_2 x_2 + ... + b_n x_x \\ \vdots \end{bmatrix} \end{aligned}$$

$$\begin{aligned} x^T V^T V x = &(x_1 a_1 + x_2 a_2 + ... + x_n a_n) * (a_1 x_1 + a_2 x_2 + ... + a_n x_x) \\ &+ (x_1 b_1 + x_2 b_2 + ... + x_n b_n) * (b_1 x_1 + b_2 x_2 + ... + b_n x_x) \\ &... \\ &= (x_1 a_1 + x_2 a_2 + ... + x_n a_n)^2 \\ &+ (x_1 b_1 + x_2 b_2 + ... + x_n b_n)^2 \\ &... \end{aligned}$$

Which will be positive because is a summation of positive values.
With the addition that C is square and symmetric, we can affirm that C is "Positive Definite".
The unique caveat is that if the v_i are linear dependent, there are at least one x that

$$Vx = 0$$

In this particular case the summation above is not greater than zero, it is exactly zero. That is why we need that v_is to be linear independent to C be "Positive "Definite" and if v_is are linear dependent C is "Positive Semidefinite".