Appendix A — Linear Algebra

A.1 Vectors and Matrices

To successfully understand linear regression, we will require some basic notations regrading the manipulation of vectors and matrices. First, we will denote a n dimensional vector as $Y \in \mathbb{R}^n$ and an $n \times p$ matrix as $X \in \mathbb{R}^{n \times p}$.

In linear regression, we are ultimately faced with the problem of solving a linear system of equations. That is, let $A \in \mathbb{R}^{p \times p}$ be an invertible matrix, $z \in \mathbb{R}^p$ a vector, and $b \in \mathbb{R}^p$ another vector. If we know A and b, then we want to solve the following system for Z.

$$\left.\begin{matrix} a_{11}z_1 + \ldots a_{1p}z_p = b_1 \\ \vdots \\ a_{p1}z_1 + \ldots a_{pp}z_p = b_p \end{matrix}\right\} \Rightarrow Az = b.$$

This solution can be written as $z = A^{-1}b$ assuming, again, that A is invertible. An invertible matrix must necessarilly be square; i.e. number of rows = number of columns.

For a matrix $X \in \mathbb{R}^{n \times p}$, its transpose is $X^{\mathrm{T}} \in \mathbb{R}^{p \times n}$. This matrix has ijth entry $(X^{\mathrm{T}})_{i,j} = X_{j,i}$. If $X = X^{\mathrm{T}}$, then we say that X is symmetric. Symmetric matrices always have real-valued eigenvalues.

A square matrix $A \in \mathbb{R}^{p \times p}$ is said to be *positive definite* if $x^{\mathrm{T}}Ax > 0$ for all choices of vector $x \in \mathbb{R}^p$ such that $x \neq 0$. If we replace the $>$ with a $\geq$, then we say that A is *positive semi-definite*; i.e. $x^{\mathrm{T}}Ax \geq 0$ for all choices of vector $x \in \mathbb{R}^p$.

A matrix is said to be *idempotent* if $A^2 = A$. Idempotent matrices define orthogonal projections in $\mathbb{R}^p$. These are of critical importance in linear regression as least squares regression is simply a projection in $\mathbb{R}^n$ of the observed output Y onto the column space of $X \in \mathbb{R}^{n \times p}$, which is the p-dimensional subspace spanned by the columns of X.

A.2 Eigenvalues

A very important aspect of a square matrix A is its eigenvalues. For a much more general discussion, see the Spectral Theorem. For our purposes, assume A is a symmetric matrix. Then, we can write $A = UDU^{\mathrm{T}}$ where D is the diagonal matrix of eigenvalues, and U is an orthogonal matrix; i.e. $UU^{\mathrm{T}} = U^{\mathrm{T}}U = I$, the identity matrix. The eigenvalues will be denoted by $\lambda_1, \ldots, \lambda_p \in \mathbb{R}$.
There are some certain special cases we will need to consider:

- if all $\lambda_i \neq 0$, then A is invertible.
- if at least one $\lambda_i = 0$, then A is singular.
- if all $\lambda_i > 0$, then A is positive definite.
- if all $\lambda_i \geq 0$, then A is positive semi-definite.
- if all $\lambda_i < 0$, then A is negative definite.
- if all $\lambda_i \leq 0$, then A is negative semi-definite.

A.3 Covariance Matrices

For a random vector $X \in \mathbb{R}^n$, its covariance matrix is the $n \times n$ matrix with ijth entry $\text{cov}\left(X_i, X_j\right)$. Covariance matrices are necessarily symmetric and positive definite. Furthermore, any symmetric positive definite matrix can be thought of as a covariance matrix. Strictly speaking, we may have a covariance matrix that is positive *semi*-definite. In this case, we say that the covariance matrix is *degenerate*.

A.4 Quick List of Facts

1. For a matrix $A \in \mathbb{R}^{n \times p}$ with row i and column j entry denoted $a_{i,j}$, then the transpose of A is $A^{\mathrm{T}} \in \mathbb{R}^{p \times n}$ with row i and column j entry $a_{j,i}$. That is, the indices have swapped.
2. For matrices $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times p}$, we have that $(AB)^{\mathrm{T}} = B^{\mathrm{T}} A^{\mathrm{T}}$
3. For an invertible matrix $A \in \mathbb{R}^{n \times n}$, we have that $\left(A^{-1}\right)^{\mathrm{T}} = \left(A^{\mathrm{T}}\right)^{-1}$.
4. A matrix A is *square* if the number of rows equals the number of columns. That is, $A \in \mathbb{R}^{n \times n}$.
5. A square matrix $A \in \mathbb{R}^{n \times n}$ is *symmetric* if $A = A^{\mathrm{T}}$.
6. A symmetric matrix $A \in \mathbb{R}^{n \times n}$ necessarily has real eigenvalues.
7. A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is *positive definite* if for all $x \in \mathbb{R}^n$ with $x \neq 0$, we have that $x^{\mathrm{T}} A x > 0$.
8. A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is also positive definite if all of its eigenvalues are positive real valued.
9. A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is *positive semi-definite* (also non-negative definite) if for all $x \in \mathbb{R}^n$ with $x \neq 0$, we have that $x^{\mathrm{T}} A x \geq 0$. Or alternatively, all of the eigenvalues are non-negative real valued.
10. Covariance matrices are always positive semi-definite. If a covariance matrix has some zero valued eigenvalues, then it is called *degenerate*.
11. If $X, Y \in \mathbb{R}^n$ are random vectors, then

$$\text{cov}\left(X, Y\right) = \mathrm{E}\left((X - \mathrm{E}X)(Y - \mathrm{E}Y)^{\mathrm{T}}\right) \in \mathbb{R}^{n \times n}.$$

12. If $X, Y \in \mathbb{R}^n$ are random vectors and $A, B \in \mathbb{R}^{m \times n}$ are non-random real valued matrices, then

$$\text{cov}\left(AX, BY\right) = A\text{cov}\left(X, Y\right) B^{\mathrm{T}} \in \mathbb{R}^{m \times m}.$$

13. If $Y \in \mathbb{R}^n$ is multivariate normal–i.e. $Y \sim \mathcal{N}\left(\mu, \Sigma\right)$–and $A \in \mathbb{R}^{m \times n}$ then AY is also multivariate normal with $AY \sim \mathcal{N}\left(A\mu, A\Sigma A^{\mathrm{T}}\right)$.
14. A square matrix $A \in \mathbb{R}^{n \times n}$ is *idempotent* if $A^2 = A$.