Orthogonality and Least Squares

1 Definitions and Terms

1.1 Basics of Vectors

Two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ can be multiplied with each other, using the **dot product**, also called the **inner product**, which produces a scalar value. It is denoted as $\mathbf{u} \cdot \mathbf{v}$, and defined as $\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + \ldots + u_n v_n$. The **length** of a vector $\mathbf{u}$, sometimes also called the **norm**, is denoted by $\|\mathbf{u}\|$. It is defined as $\|\mathbf{u}\| = \sqrt{\mathbf{u} \cdot \mathbf{u}} = \sqrt{u_1^2 + \ldots + u_n^2}$.

A **unit vector** is a vector whose length is 1. For any nonzero vector $\mathbf{u}$, the vector $\frac{\mathbf{u}}{\|\mathbf{u}\|}$ is a unit vector in the direction of $\mathbf{u}$. This process of creating unit vectors is called **normalizing**. The **distance between u and v**, written as $\text{dist}(\mathbf{u}, \mathbf{v})$, is the length of the vector $\mathbf{v} - \mathbf{u}$. That is, $\text{dist}(\mathbf{u}, \mathbf{v}) = \|\mathbf{v} - \mathbf{u}\|$.

1.2 Orthogonal Sets

Two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ are **orthogonal** if $\mathbf{u} \cdot \mathbf{v} = 0$. If $\mathbf{z}$ is orthogonal to every vector in a subspace W, then $\mathbf{z}$ is said to be **orthogonal to** W. The set of all vectors $\mathbf{z}$ that are orthogonal to W is called the **orthogonal complement** of W, and is denoted by $W^\perp$.

A set of vectors $\{\mathbf{u_1}, \ldots, \mathbf{u_n}\}$ in $\mathbb{R}^n$ is said to be an **orthogonal set** if each pair of distinct pair of vectors is orthogonal, that is, if $\mathbf{u_i} \cdot \mathbf{u_j} = 0$ whenever $i \neq j$. An **orthogonal basis** for a subspace W of $\mathbb{R}^n$ is a basis for W of $\mathbb{R}^n$ is a basis for W that is also an orthogonal set.

1.3 Orthonormal Sets

A set of vectors $\{\mathbf{u_1}, \ldots, \mathbf{u_n}\}$ in $\mathbb{R}^n$ is an **orthonormal set** if it is an orthogonal set of unit vectors. If W is the subspace spanned by such a set, then $\{\mathbf{u_1}, \ldots, \mathbf{u_n}\}$ is an **orthonormal basis** for W. An **orthogonal matrix** is a square invertible matrix U such that $U^{-1} = U^T$. Such a matrix always has orthonormal columns.

1.4 Decomposing Vectors

If $\mathbf{u}$ is any nonzero vector in $\mathbb{R}^n$, then it is possible to decompose any vector $\mathbf{y}$ in $\mathbb{R}^n$ into the sum of two vectors, one being a multiple of $\mathbf{u}$, and one being orthogonal to it. The projection $\hat{\mathbf{y}}$ (being the multiple of $\mathbf{u}$) is called the **orthogonal projection of y onto u**, and the component of $\mathbf{y}$ orthogonal to $\mathbf{u}$ is, surprisingly, called the **component of y orthogonal to u**.

Just like it is possible to project vectors on a vector, it is also possible to project vectors on a subspace. The projection $\hat{\mathbf{y}}$ onto the subspace W is called the **orthogonal projection of y onto** W. $\hat{\mathbf{y}}$ is sometimes also called **the best approximation to y by elements of** W.

1.5 The Gram-Schmidt Process

The **Gram-Schmidt Process** is an algorithm for producing an orthogonal or orthonormal basis $\{\mathbf{u_1}, \ldots, \mathbf{u_p}\}$ for any nonzero subspace of $\mathbb{R}^n$. Let W be the subspace, having basis $\{\mathbf{x_1}, \ldots, \mathbf{x_p}\}$. Let $\mathbf{u_1} = \mathbf{x_1}$ and $\mathbf{u_i} = \mathbf{x_i} - \mathbf{\hat{x}_i}$ for $1 < i \leq n$, where $\mathbf{\hat{x}_i}$ is the projection of $\mathbf{x_i}$ on the subspace with basis $\{\mathbf{u_1}, \ldots, \mathbf{u_{i-1}}\}$. In formula: $\mathbf{u_1} = \mathbf{x_1}$ and $\mathbf{u_i} = \mathbf{x_i} - (\frac{x_i \cdot v_1}{v_1 \cdot v_1} v_1 + \ldots + \frac{x_i \cdot v_{i-1}}{v_{i-1} \cdot v_{i-1}} v_{i-1})$.

1.6 Least-Squares Problem

The **general least-squares problem** is to find an $\mathbf{x}$ that makes $\|\mathbf{b} - A\mathbf{x}\|$ as small as possible. If A is $m \times n$ and $\mathbf{b}$ is in $\mathbb{R}^m$, a **least-squares solution** of $A\mathbf{x} = \mathbf{b}$ is an $\mathbf{\hat{x}}$ in $\mathbb{R}^n$ such that $\|\mathbf{b} - A\mathbf{\hat{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\|$ for all $\mathbf{x}$ in $\mathbb{R}^n$. When a least-squares solution $\mathbf{\hat{x}}$ is used to produce $A\mathbf{\hat{x}}$ as an approximation of $\mathbf{b}$, the distance from $\mathbf{b}$ to $A\mathbf{\hat{x}}$ is called the **least-squares error** of this approximation.

1.7 Linear Models

In statistical analysis of scientific and engineering data, there is commonly a different notation used. Instead of $A\mathbf{x} = \mathbf{b}$, we write $X\beta = \mathbf{y}$ and refer to X as the **design matrix**, β as the **parameter vector**, and $\mathbf{y}$ as the **observation vector**.

Suppose we have a certain amount of measurement data which, when plotted, seem to lie close to a straight line. Let $y = \beta_0 + \beta_1 x$. The difference between the observed value (from the measurements) and the predicted value (from the line) is called a **residual**. The **least-squares line** is the line that minimizes the sum of the squares of the residuals. This line is also called a **line of regression of y on** x. The coefficients β_0 and β_1 are called (linear) **regression coefficients**.

The previous system is equivalent to solving the least-squares solution of $X\beta = \mathbf{y}$ if $X = [\,\mathbf{1}\ \ \mathbf{x}\,]$ (where $\mathbf{1}$ has entries 1, 1, …, 1, and $\mathbf{x}$ has entries $x_1, \ldots, x_n$), β has entries β_0 and β_1 and $\mathbf{y}$ has entries $y_1, \ldots, y_n$. A common practice before computing a least-squares line is to compute the average $\bar{x}$ of the original x-values, and form a new variable $x^* = x - \bar{x}$. The new x-data are said to be in **mean-deviation form**. In this case, the two columns of X will be orthogonal.

The **residual vector** ϵ is defined as $\epsilon = \mathbf{y} - X\beta$. So $\mathbf{y} = X\beta + \epsilon$. Any equation in this form is referred to as a **linear model**, in which ϵ should be minimized.

1.8 Inner Product Spaces

An **inner product** on a vector space V is a function that, to each pair of vectors $\mathbf{u}$ and $\mathbf{v}$ in U, associates a real number $\langle \mathbf{u}, \mathbf{v} \rangle$ and satisfies the following axioms, for all $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ in V and all scalars c:

1. $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$
2. $\langle \mathbf{u} + \mathbf{v}, \mathbf{v} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$
3. $\langle c\mathbf{u}, \mathbf{v} \rangle = c\langle \mathbf{u}, \mathbf{v} \rangle$
4. $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$ and $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ if, and only if $\mathbf{u} = \mathbf{0}$

A vector space with an inner product is called an **inner product space**.

2 Theorems

1. Consider the vectors $\mathbf{u}$ and $\mathbf{v}$ as $n \times 1$ matrices. Then, $\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T \mathbf{v}$.

2. If $\mathbf{u} \neq \mathbf{0}$ and $\mathbf{v} \neq \mathbf{0}$ then $\mathbf{u}$ and $\mathbf{v}$ are orthogonal if, and only if $\mathbf{u} \cdot \mathbf{v} = 0$.

3. Two vectors $\mathbf{u}$ and $\mathbf{v}$ are orthogonal if, and only if $\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$.

4. A vector $\mathbf{z}$ is in $W^{\perp}$ if, and only if $\mathbf{z}$ is orthogonal to every vector in a set that spans W.

5. $W^{\perp}$ is a subspace of $\mathbb{R}^n$.

6. If A is an $m \times n$ matrix, then $(\text{ Row}A)^{\perp} = \text{ Nul}A$ and $(\text{ Col}A)^{\perp} = \text{ Nul}A^T$.

7. If A is an $m \times n$ matrix, then $\text{ Row}A = \text{ Col}A^T$.

8. If $S = \{\mathbf{u_1}, \ldots, \mathbf{u_p}\}$ is an orthogonal set of nonzero vectors in $\mathbb{R}^n$, then S is linearly independent and hence is a basis for the subspace spanned by S.

9. Let $\{\mathbf{u_1}, \ldots, \mathbf{u_p}\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$. For each $\mathbf{y}$ in W, the weights in the linear combination $\mathbf{y} = c_1\mathbf{u_1} + \ldots + c_p\mathbf{u_p}$ are given by $c_j = \frac{\mathbf{y}\cdot\mathbf{u_j}}{\mathbf{u_j}\cdot\mathbf{u_j}} = \frac{\mathbf{y}\cdot\mathbf{u_j}}{\|\mathbf{u_j}\|^2}$.

10. An $m \times n$ matrix U has orthonormal columns if, and only if $U^TU = I$.

11. Let U be an $m \times n$ matrix with orthonormal columns, and let $\mathbf{x}$ and $\mathbf{y}$ be in $\mathbb{R}^n$, then:
    (a) $\|U\mathbf{x}\| = \|\mathbf{x}\|$
    (b) $(U\mathbf{x}) \cdot (U\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$
    (c) $(U\mathbf{x}) \cdot (U\mathbf{y}) = 0$ if, and only if $\mathbf{x} \cdot \mathbf{y} = 0$.

12. If U is a square matrix, then U is an orthogonal matrix if, and only if its columns are orthonormal columns. The rows of an orthogonal matrix are also orthonormal rows.

13. If $\mathbf{y}$ and $\mathbf{u}$ are any nonzero vectors in $\mathbb{R}^n$, then the orthogonal projection of $\mathbf{y}$ onto $\mathbf{u}$ is $\hat{\mathbf{y}} = \frac{\mathbf{y}\cdot\mathbf{u_j}}{\mathbf{u_j}\cdot\mathbf{u_j}}\mathbf{u} = \frac{\mathbf{y}\cdot\mathbf{u_j}}{\|\mathbf{u_j}\|^2}\mathbf{u}$, and the component $\mathbf{z}$ of $\mathbf{y}$ orthogonal to $\mathbf{u}$ is $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$.

14. Let W be a subspace of $\mathbb{R}^n$. Then each $\mathbf{y}$ in $\mathbb{R}^n$ can be written uniquely in the form $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$ where $\hat{\mathbf{y}}$ is in W and $\mathbf{z}$ is in $W^{\perp}$. In fact, if $\{\mathbf{u_1}, \ldots, \mathbf{u_p}\}$ is any orthogonal basis of W, then $\hat{\mathbf{y}} = \frac{\mathbf{y}\cdot\mathbf{u_1}}{\mathbf{u_1}\cdot\mathbf{u_1}}\mathbf{u_1} + \ldots + \frac{\mathbf{y}\cdot\mathbf{u_p}}{\mathbf{u_p}\cdot\mathbf{u_p}}\mathbf{u_p}$, and $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$.

15. **The Best Approximation Theorem**. Let W be a subspace of $\mathbb{R}^n$, $\mathbf{y}$ any vector in $\mathbb{R}^n$, and $\hat{\mathbf{y}}$ the orthogonal projection of $\mathbf{y}$ onto W. Then $\hat{\mathbf{y}}$ is the closest point in W to $\mathbf{y}$, in the sense that $\|\mathbf{y} - \hat{\mathbf{y}}\| < \|\mathbf{y} - \mathbf{u}\|$ for all $\mathbf{u} \neq \hat{\mathbf{y}}$ in W.

16. If $\{\mathbf{u_1}, \ldots, \mathbf{u_p}\}$ is an orthonormal basis for a subspace W of $\mathbb{R}^n$, then $\hat{\mathbf{y}} = (\mathbf{y}\cdot\mathbf{u_1})\mathbf{u_1} + \ldots + (\mathbf{y}\cdot\mathbf{u_p})\mathbf{u_p}$. If $U = [\ \mathbf{u_1}\ \ \ldots\ \ \mathbf{u_p}\ ]$, then $\hat{\mathbf{y}} = UU^T\mathbf{y}$ for all $\mathbf{y}$ in $\mathbb{R}^n$.

17. The set of least-squares solutions of $A\mathbf{x} = \mathbf{b}$ coincides with the nonempty set of solutions of the normal equations $A^TA\mathbf{x} = A^T\mathbf{b}$.

18. The matrix A^TA is invertible if, and only if the columns of A are linearly independent. In that case, the equation $A\mathbf{x} = \mathbf{b}$ has only one least-squares solution $\hat{\mathbf{x}}$, and it is given by $\hat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b}$.

3 Calculation Rules

3.1 Algebraic Definitions

The dot product of vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ is defined as:

$$\mathbf{u} \cdot \mathbf{v} = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = u_1 v_1 + \ldots u_n v_n \tag{1}$$

The length of a vector is defined as:

$$\|\mathbf{u}\| = \sqrt{\mathbf{u} \cdot \mathbf{u}} = \sqrt{u_1^2 + \ldots + u_n^2} \tag{2}$$

3.2 Algebraic Rules

The following rules apply for the dot product:

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u} \tag{3}$$

$$(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w} \tag{4}$$

$$(c\mathbf{u}) \cdot \mathbf{v} = c(\mathbf{u} \cdot \mathbf{v}) = \mathbf{u} \cdot (c\mathbf{v}) \tag{5}$$

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T \mathbf{v} \tag{6}$$

The following rules apply for vector lengths:

$$\|c\mathbf{u}\| = |c| \|\mathbf{u}\| \tag{7}$$

$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \|\mathbf{v}\| \cos\theta \tag{8}$$

Where θ is the angle between vectors $\mathbf{v}$ and $\mathbf{u}$.