Statistical Operations and Matrices (I)

Matrix Algebra 4 Statistical Learning

Gaston Sanchez

Vector-Matrix Notation for Statistical Operations

Motivation

I want to discuss how we can use vector-matrix notation to represent some basic statistical operations and summaries.

First we need to quickly review some concepts around inner products.

Inner Product

The inner product of two vectors $\mathbf{x}$ and $\mathbf{y}$—of the same size—is defined as:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^{\mathsf{T}} \mathbf{y} = \sum_i x_i y_i$$

basically the inner product consists of the element-by-element product of $\mathbf{x}$ and $\mathbf{y}$, and then adding everything up. The result is not another vector but a single number, a scalar.

Inner Product

We can also write the inner product $\langle \mathbf{x}, \mathbf{y} \rangle$ in vector notation as:

$$\mathbf{x}^{\mathsf{T}}\mathbf{y} = (x_1 \dots x_n) \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \sum_{i=1}^{n} x_i y_i$$

Inner Product

Keep in mind that inner products can be generalized to any type of metric $\mathbf{M}$ matrix:

$$\langle \mathbf{x}, \mathbf{y} \rangle_M = \mathbf{x}^\mathsf{T} \mathbf{M} \mathbf{y}$$

Inner Product

Having an inner space endowed with an inner product, we can derive other concepts such as:

- **Length** of a vector (and *norms* in general)
- **Distance** between points
- **Angle** between vectors
- **Projection** of vectors

Length

Another important usage of the inner product is that it allows us to define the **length** of a vector $\mathbf{x}$ denoted by:

$$\|\mathbf{x}\| = \sqrt{\mathbf{x}^\mathsf{T}\mathbf{x}}$$

which is typically known as the (Euclidean) **norm** of a vector.
(There are actually other types of norms)

Length

The inner product of a vector with itself is equal to its squared norm: $\mathbf{x}^{\mathsf{T}}\mathbf{x} = \|\mathbf{x}\|^2$

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

Distance

The square (Eculidean) distance between two vectors $\mathbf{x}$ and $\mathbf{y}$ can be obtained as:

$$d^2(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|^2 = (\mathbf{x} - \mathbf{y})^\mathsf{T}(\mathbf{x} - \mathbf{y})$$

Distance

The square (Eculidean) distance between two vectors $\mathbf{x}$ and $\mathbf{y}$ can be obtained as:

$$d^2(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|^2 = (\mathbf{x} - \mathbf{y})^\mathsf{T}(\mathbf{x} - \mathbf{y})$$

The distance between two vectors $\mathbf{x}$ and $\mathbf{y}$ can be obtained as:

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\| = \sqrt{(\mathbf{x} - \mathbf{y})^\mathsf{T}(\mathbf{x} - \mathbf{y})}$$

Angle between Vectors

The **angle** θ between two nonzero vectors $\mathbf{x}, \mathbf{y}$ can also be expressed using inner products:

$$cos(\theta) = \frac{\mathbf{x}^{\mathsf{T}}\mathbf{y}}{\sqrt{\mathbf{x}^{\mathsf{T}}\mathbf{x}}\sqrt{\mathbf{y}^{\mathsf{T}}\mathbf{y}}} = \frac{\mathbf{x}^{\mathsf{T}}\mathbf{y}}{\|\mathbf{x}\|\,\|\mathbf{y}\|}$$

Angle

Equivalently, we can reexpress the formula of the inner product using

$$\mathbf{x}^{\mathsf{T}}\mathbf{y} = \|\mathbf{x}\| \, \|\mathbf{y}\| \, cos(\theta)$$

Orthogonality

Besides calculating lengths of vectors and angles between vectors, an inner product allows us to know whether two vectors are orthogonal.

In a two dimensional space, orthogonality is equivalent to perpendicularity; so if two vectors are perpendicular to each other—the angle between them is a 90 degree angle—they are orthogonal.

Two vectors vectors $\mathbf{x}$ and $\mathbf{y}$ are orthogonal if their inner product is zero:

$$\mathbf{x}^{\mathsf{T}}\mathbf{y} = 0 \quad \Longleftrightarrow \quad \mathbf{x} \perp \mathbf{y}$$

Projection

The last aspect I want to touch related with the inner product is the so-called projections. More specifically: orthogonal projection of a vector $\mathbf{y}$ onto another vector $\mathbf{x}$.

The basic notion of projection requires two ingredients: two vectors, $\mathbf{x}$ and $\mathbf{y}$. To obtain the projection of $\mathbf{y}$ onto $\mathbf{x}$, we need to express $\mathbf{x}$ in unit norm.

Unit Vector

Recall that a **unit vector** is a vector of length 1, sometimes also called a *direction* or *unitary* vector.

In other words, if $\mathbf{v}$ is a unit vector, this means that $\|\mathbf{v}\| = 1$.

Given a non-zero vector $\mathbf{v}$, how do we get a unit vector?

Unit Vector

Given a non-zero vector $\mathbf{v}$, you can get a unit vector $\mathbf{u}$ by dividing $\mathbf{v}$ by its norm, that is:

$$\mathbf{u} = \frac{\mathbf{v}}{\|\mathsf{v}\|} = \frac{\mathbf{v}}{\sqrt{\mathbf{v}^{\mathsf{T}}\mathbf{v}}}$$

Some say that we "normalize $\mathbf{v}$"

Projection

Having two nonzero vectors $\mathbf{x}$ and $\mathbf{y}$, we can project $\mathbf{y}$ on $\mathbf{x}$, denoted by $\hat{\mathbf{y}}$ as:

$$\hat{\mathbf{y}} = \mathbf{x}\left(\frac{\mathbf{y}^\top \mathbf{x}}{\mathbf{x}^\top \mathbf{x}}\right)$$

Note that the term in parenthesis is just a scalar.

Projection

Note that

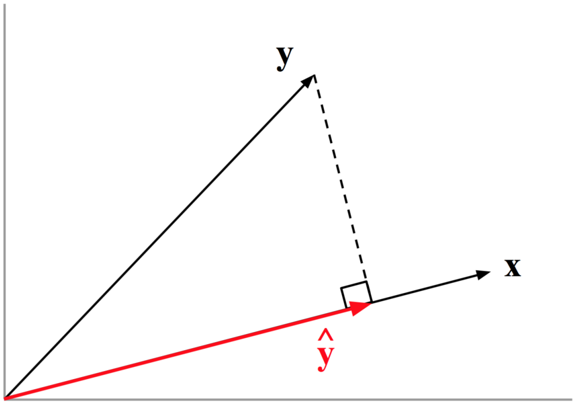
$$\hat{\mathbf{y}} = \mathbf{x}\left(\frac{\mathbf{y}^\top\mathbf{x}}{\mathbf{x}^\top\mathbf{x}}\right) = \mathbf{x}\left(\frac{\mathbf{x}^\top\mathbf{y}}{\mathbf{x}^\top\mathbf{x}}\right)$$

can also be written as:

$$\hat{\mathbf{y}} = \mathbf{x}(\mathbf{x}^\top\mathbf{x})^{-1}\mathbf{x}^\top\mathbf{y}$$

Does it look familiar?

Orthogonal Projection



Projection

We can actually express $\hat{\mathbf{y}}$ as $a\mathbf{x}$. This means that a projection implies multiplying $\mathbf{x}$ by some number a, such that $\hat{\mathbf{y}} = a\mathbf{x}$ is a stretched or shrinked version of $\mathbf{x}$.

Inner Products ... So what?

We'll see in a moment how many statistical summaries can be represented with inner products.

Review of Some Statistical Operations

Statistical Summaries

I want to review basic statistical summaries and see how we can express them in vector-matrix notation:

- Sum of values
- Sum of squared values
- Mean
- Variance

Sum of Values

Consider a variable $X \in \mathbb{R}^n$ represented with a vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)$

A common operation consists of adding the elements of the vector

$$\sum_{i=1}^{n} x_i$$

Sum of Values

Consider a variable $X \in \mathbb{R}^n$ represented with a vector $\mathbf{x} = (x_1, x_2, \dots, x_n)$

A common operation consists of adding the elements of the vector

$$\sum_{i=1}^{n} x_i$$

This sum can be expressed in vector notation as:

$$\sum_{i=1}^{n} x_i = \mathbf{x}^\mathsf{T}\mathbf{1} = \mathbf{1}^\mathsf{T}\mathbf{x}$$

where $\mathbf{1}$ is a vector of n elements equal to 1

Mean

The (arithmetic) mean of X is

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Mean

The (arithmetic) mean of X is

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Using vector notation the mean is expressed as:

$$\bar{x} = \frac{1}{n} \mathbf{x}^\mathsf{T} \mathbf{1}$$

Sum of Square dValues

Another common operation is the sum of squared values

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

Sum of Square dValues

Another common operation is the sum of squared values

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

Using vector notation this sum can be written as an inner product:

$$\sum_{i=1}^{n} x_i^2 = \mathbf{x}^{\mathsf{T}}\mathbf{x} = \|\mathsf{x}\|^2$$

Variance

The variance of X (in its population version)

$$var(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

- Let $\bar{\mathbf{x}}$ be an n-element vector constant of $\bar{x}$ values
- Let $\tilde{\mathbf{x}} = \mathbf{x} - \bar{\mathbf{x}}$

Variance

The variance of X (in its population version)

$$var(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

- Let $\bar{\mathbf{x}}$ be an n-element vector constant of $\bar{x}$ values
- Let $\tilde{\mathbf{x}} = \mathbf{x} - \bar{\mathbf{x}}$

The variance can be expressed as:

$$var(X) = \frac{1}{n}\tilde{\mathbf{x}}^{\mathsf{T}}\tilde{\mathbf{x}}$$

Variance

If you consider the “sample” variance

$$var(X) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Then:

$$var(X) = \frac{1}{n-1}\tilde{\mathbf{x}}^{\mathsf{T}}\tilde{\mathbf{x}}$$

Variance

- I will sometimes consider $\mathbf{x}$ to be mean-centered.
- This means that $\bar{x}$ has already been subtracted from each element x_i

In this case the variance can be compactly expressed as:

$$var(X) = \frac{1}{n}\mathbf{x}^\mathsf{T}\mathbf{x}$$

Covariance

Consider two variables X and Y. The covariance $cov(X, Y)$

$$cov(X, Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

Covariance

Consider two variables X and Y. The covariance $cov(X,Y)$

$$cov(X,Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

Assuming that X and Y are mean-centered, and represented by $\mathbf{x}$ and $\mathbf{y}$, respectively, the covariance can be expressed as:

$$cov(\mathbf{x},\mathbf{y}) = \frac{1}{n}\mathbf{x}^{\mathsf{T}}\mathbf{y}$$

Correlation

Consider two variables X and Y. The correlation $cor(X, Y)$

$$cor(X, Y) = \frac{cov(X, Y)}{\sqrt{var(X)}\sqrt{var(Y)}}$$

Correlation

Consider two variables X and Y. The correlation $cor(X, Y)$

$$cor(X, Y) = \frac{cov(X, Y)}{\sqrt{var(X)}\sqrt{var(Y)}}$$

Assuming that X and Y are mean-centered, and represented by $\mathbf{x}$ and $\mathbf{y}$, respectively, the covariance can be expressed as:

$$cor(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{x}^\mathsf{T}\mathbf{y}}{\sqrt{\mathbf{x}^\mathsf{T}\mathbf{x}}\sqrt{\mathbf{y}^\mathsf{T}\mathbf{y}}} = \frac{\mathbf{x}^\mathsf{T}\mathbf{y}}{\|\mathbf{x}\|\ \|\mathbf{y}\|}$$

Correlation: Geometric Interpretation

What does it mean geometrically:

$$cor(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{x}^\mathsf{T}\mathbf{y}}{\|\mathbf{x}\| \; \|\mathbf{y}\|}$$

Hint: recall the angle between two vectors.

You can think of the correlation between two variables as the cosine of the angle between two vectors.