Linear Algebra

October 26, 2015

Dot product of two vectors

The dot product of $\mathbf{a}$ and $\mathbf{b}$, two vectors of length n, is given by

$$\mathbf{a} \cdot \mathbf{b} = a_1 \cdot b_1 + a_2 \cdot b_2 + \cdots + a_n \cdot b_n$$

If $\mathbf{a}$ and $\mathbf{b}$ are centered at zero (i.e., de-meaned), their dot product divided by n is the covariance of $\mathbf{a}$ and $\mathbf{b}$. Recall that the covariance of $\mathbf{a}$ and $\mathbf{b}$ is given by

$$\text{Cov}(\mathbf{a},\ \mathbf{b}) = \frac{\sum_{i=1}^{n}(a_i - \bar{a})(b_i - \bar{b})}{n-1}$$

Euclidean distance between two vectors

The straight-line distance between two points in the n-dimensional space.

$$\text{d}(\mathbf{a},\ \mathbf{b}) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \cdots + (a_n - b_n)^2}$$

Cosine similarity

Another metric of the similarity between two vectors is the cosine of the angle θ between them, given by

$$\cos\theta = \frac{\mathbf{a} \cdot \mathbf{b}}{||\mathbf{a}|| \cdot ||\mathbf{b}||}$$

If both $\mathbf{a}$ and $\mathbf{b}$ are demeaned, the cosine similarity is the correlation between the two vectors.

Vector norm

The norm of a vector is a measure of its *size* or its *length* in the n-dimensional space. It is effectively the distance from the origin.

$$||\mathbf{a}|| = \sqrt{a_1^2 + a_2^2 + \cdots + a_n^2}$$

Similar statistics are critical for when we discuss *regularization*, specifically L_1 and L_2 regularization for a vector of parameter estimates θ. The L_1 penalty is the sum of the absolute values of the elements in the vector.

$$||\theta||_1 = \sum_{i=1}^{n} |\theta_i|$$

And the L_2 penalty is the sum of the squared values of the elements in the vector.

$$||\theta||_2^2 = \sum_{i=1}^{n} \theta_i^2$$

Note that this is simply the dot product of the vector with itself. And therefore the variance of this vector can be obtained by demeaning and dividing by its length, as above. The generalizable form of the above is the L_p norm, defined as

$$||\theta||_p = \left(\sum_i^n |\theta_i| \right)^{\frac{1}{p}}$$

But we will most commonly run into the three forms shown above.

Linear combinations

Often we wish to multiply vectors by constants and then add the vectors. This procedure draws on two fundamental concepts. First, the product of a vector $\mathbf{v}$ and a constant a is a vector of the elements $v_i \in \mathbf{v}$ multiplied by the constant.

$$\begin{vmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{vmatrix} \cdot a = \begin{vmatrix} a \cdot v_1 \\ a \cdot v_2 \\ \vdots \\ a \cdot v_n \end{vmatrix}$$

Second, the sum of two vectors $\mathbf{v}$ and $\mathbf{v}$ is a vector of the element-wise sums of the vectors.

$$\begin{vmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{vmatrix} + \begin{vmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{vmatrix} = \begin{vmatrix} v_1 + w_1 \\ v_2 + w_2 \\ \vdots \\ v_n + w_n \end{vmatrix}$$

Then we can express a linear combination of vectors $v_1, \ldots, v_m$ as the sum of vectors each multiplied by a constant.

$$a_1 \cdot \mathbf{v}_1 + a_2 \cdot \mathbf{v}_2 + \cdots + a_m \cdot \mathbf{v}_m$$

In linear regression, for example, we choose the weights $\mathbf{a}$ to minimize the difference between the linear combination and some quantitative output vector.

Transpose of a matrix

The *transpose* of a matrix $\mathbf{A}$ is obtained by interchanging its rows and columns.

$$\mathbf{A} = \begin{vmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{vmatrix} \qquad \mathbf{A}^T = \begin{vmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{vmatrix}$$

Matrix multiplication

The dot product of two matrices $\mathbf{A}$ and $\mathbf{B}$ is possible when the the number of columns in $\mathbf{A}$ matches the number of rows in $\mathbf{B}$. The result is a matrix with the number of rows equal to the number of rows in $\mathbf{A}$ and columns equal to the number of columns in $\mathbf{B}$.

Generally speaking, $\mathbf{AB} \neq \mathbf{BA}$. Why is that?

Inverse of a matrix

For a square matrix $\mathbf{A}_{n \times n}$, find $\mathbf{A}^{-1}_{n \times n}$ such that the dot product of $\mathbf{A}$ and $\mathbf{A}^{-1}$ is equal to the identity matrix $\mathbf{I}_n$ (i.e., 1s on the diagonal). One place where we will see this is in the closed form solution for the coefficient estimates in a linear regression. Specifically,

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$$

Questions

1. What is the norm the following vectors?

- A 1×8 zero vector?
- The vector $\{3, 4\}$?
- The vector $\{1, 1, 1, 1\}$

2. What is the shape of the result of the following operations?

- $A_{1\times n} \cdot B_{1\times n}$
- $A^T_{5\times 4}$
- $A_{10\times 3} \cdot B_{3\times 5}$
- $A_{3\times 10} \cdot B_{5\times 3}$
- $(A_{5\times 4} \cdot B_{4\times 3})^{-1}$
- Given $A_{5\times 4}$, $\left(A^T A\right)^{-1}$