CS 189/289 Math Prerequisites

Isabelle Guyon, September 2015

Linear Algebra

Vector

Patterns or data samples will be represented as vectors in a d dimensional space:

$\mathbf{x} = [x_1, x_2, \ldots, x_d]$

The elements of the vector (coefficients) $x_1, x_2, \ldots, x_d$ correspond to input variables or features.

Operations on vectors

Given 2 vectors $\mathbf{x} = [x_1, x_2, \ldots, x_d]$ and $\mathbf{x'} = [x'_1, x'_2, \ldots, x'_d]$

Sum: $\mathbf{x + x'} = [x_1+x'_1, x_2+x'_2, \ldots , x_d+x'_d]$

Difference: $\mathbf{x - x'} = [x_1-x'_1, x_2-x'_2, \ldots , x_d-x'_d]$

Multiplication by a scalar "λ": $\lambda\mathbf{x} = [\lambda x_1, \lambda x_2, \ldots, \lambda x_d]$

Division by a scalar "λ": $\mathbf{x}/\lambda = [x_1/\lambda, x_2/\lambda, \ldots, x_d/\lambda]$

Dot product (scalar product)

Given 2 vectors $\mathbf{x} = [x_1, x_2, \ldots, x_d]$ and $\mathbf{x'} = [x'_1, x'_2, \ldots, x'_d]$ the dot product is defined as:

$\mathbf{x} \,.\, \mathbf{x'} = x_1 x'_1 + x_2 x'_2 + \ldots + x_d x'_d$

We often use the sum notation: $\mathbf{x} \,.\, \mathbf{x'} = \sum_{i=1:d} x_i x'_i$

The dot product is a ***measure of similarity between 2 vectors***. Two orthogonal vectors have a dot product of 0.

Euclidean norm and distance

The norm of $\mathbf{x}$ is:

$\|\mathbf{x}\| = \mathrm{sqrt}(\mathbf{x} \,.\, \mathbf{x'})$

This is the length of vector **x.**

With the norm, we can define the distance between 2 vectors: $d(\mathbf{x}, \mathbf{x'}) = \|\mathbf{x}-\mathbf{x'}\|$

Normalizing a vector

$\mathbf{x}/\|\mathbf{x}\|$ is a vector of norm 1. It often makes sense to normalize feature vectors before using the dot product as a measure of similarity because the information may lie more in the relative values of the coefficients rather than the average intensity. For instance, an image may have been captured with more or less light; normalization would focus more on the contrast between pixels that the overall luminosity of the image.

Cosine between two vectors

The cosine is the dot product between normalized vectors. It is also a measure of similarity between patterns, taking out the influence of the norm.

$\cos(\mathbf{x}, \mathbf{x'}) = (\mathbf{x}/\|\mathbf{x}\|) \,\mathbf{.}\, (\mathbf{x'}/\|\mathbf{x'}\|)$

The cosine is between -1 and 1. The cosine of orthogonal vectors is 0.

Standardizing a vector

The mean μ_x and the standard deviation σ_x of a vector $\mathbf{x}$ are defined as:

$\mu_x = (1/d) \sum_{i=1:d} x_i$

$\sigma_x = \mathrm{sqrt}\big((1/d) \sum_{i=1:d} (x_i - \mu_x)^2 \big) = \|\mathbf{x} - \mu_x\|$

σ_x is the norm of the centered vector. It is often meaningful to first remove the mean value of a vector, which may correspond to a simple offset, before normalizing it. This leads to the centered vector: $\mathbf{x} - \mu_x$.

The standardized vector is:

$s(\mathbf{x}) = (\mathbf{x} - \mu_x)/\sigma_x = (\mathbf{x} - \mu_x)/\|\mathbf{x} - \mu_x\|$

Like the normalized vector $\mathbf{x}/\|\mathbf{x}\|$ it has norm 1.

Covariance and correlation between two vectors

The covariance between two vectors is the dot product of its centered versions:

$\mathrm{corr}(\mathbf{x}, \mathbf{x'}) = (\mathbf{x} - \mu_x) \,\mathbf{.}\, (\mathbf{x'} - \mu_{x'})$

The correlation between two vectors is the dot product of its standardized versions:

$\mathrm{corr}(\mathbf{x}, \mathbf{x'}) = s(\mathbf{x}) \mathbf{.}\, s(\mathbf{x'})$

Linearity

A function $F(\mathbf{x})$ is linear if and only if, given 2 scalars λ and μ:

$F(\lambda \mathbf{x} + \mu \mathbf{x'}) = \lambda F(\mathbf{x}) + F(\mu \mathbf{x'})$

Given a vector **w** of same dimension as **x**, the following function is linear in **x** and **w**.

$F(\mathbf{x}, \mathbf{w}) = \mathbf{w} . \mathbf{x} = w_1 x_1 + w_2 x_2 + \ldots + w_d x_d$

Consider a feature transformation $\Phi(\mathbf{x}) = [\phi_1(\mathbf{x}), \phi_2(\mathbf{x}), \ldots, \phi_D(\mathbf{x})]$ into a space of dimension D, the following function is linear in **w** but NOT necessarily in **x**.

$F(\mathbf{x}, \mathbf{w}) = \mathbf{w} . \Phi(\mathbf{x}) = w_1 \phi_1(\mathbf{x}), + w_2 \phi_2(\mathbf{x}) + \ldots + w_D \phi_D(\mathbf{x})$

(note: **w** in that case is of dimension D).

Hyperplane

Consider a vectorial space of vectors **x** of dimension d. A subspace of dimension (d-1) is called hyperplane. The equation of a hyperplane is:

$\mathbf{w} . \mathbf{x} = 0$

or

$w_1 x_1 + w_2 x_2 + \ldots + w_d x_d = 0$

Vector **w** is orthogonal to the hyperplane.