Math 52 0 - Linear algebra, Spring Semester 2012-2013
Dan Abramovich

Orthogonality

Inner or dot product in $\mathbb{R}^n$:
$\mathbf{u}^T\mathbf{v} = \mathbf{u} \cdot \mathbf{v} = u_1 v_1 + \cdots u_n v_n$ ←examples

Properties:

- $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$
- $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$
- $(c\mathbf{v}) \cdot \mathbf{w} = c(\mathbf{v} \cdot \mathbf{w})$
- $\mathbf{u} \cdot \mathbf{u} \geq 0$,
  and $\mathbf{u} \cdot \mathbf{u} > 0$ unless $\mathbf{u} = \mathbf{0}$.

Norm = length:

$$\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}}$$

$\mathbf{u}$ is a unit vector if $\|\mathbf{u}\| = 1$.

If $\mathbf{v} \neq 0$ then

$$\mathbf{u} = \frac{1}{\|\mathbf{v}\|}\mathbf{v}$$

is a unit vector positively-proportional to $\mathbf{v}$.

examples→

We argue that $\mathbf{u} \perp \mathbf{v}$ if and only if $\|\mathbf{u} - \mathbf{v}\| = \|\mathbf{u} - (-\mathbf{v})\|$.

←perpendicular or orthogonal

This means

$$(\mathbf{u}-\mathbf{v})\cdot(\mathbf{u}-\mathbf{v}) = (\mathbf{u}+\mathbf{v})\cdot(\mathbf{u}+\mathbf{v})$$

Expand and get

$$\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\mathbf{u}\cdot\mathbf{v}$$
$$= \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 + 2\mathbf{u}\cdot\mathbf{v}$$

So

$$\mathbf{u} \perp \mathbf{v} \iff \mathbf{u}\cdot\mathbf{v} = 0$$

Exercise:

$\mathbf{u} \perp \mathbf{v}$

$$\iff \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 = \|\mathbf{u} + \mathbf{v}\|^2$$

Angles: We define the angle cosine between nonzero $\mathbf{u}$ and $\mathbf{v}$ to be

examples→

$$\cos\theta = \frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\,\|\mathbf{v}\|}$$

This fits with the law of cosines!

$W \subset \mathbb{R}^n$ a subspace.

we say

$$\mathbf{u} \perp W$$

if for all $\mathbf{v} \in W$ we have

$$\mathbf{u} \perp \mathbf{v}.$$

$$W^{\perp} = \{\mathbf{u} \in \mathbb{R}^n | \mathbf{u} \perp W\}.$$

Theorem: if $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ spans W then

$$\mathbf{u} \perp W$$

if and only if

$$\mathbf{u} \perp \mathbf{v}_i \quad \text{for all } i.$$

Theorem: $W^\perp \subset \mathbb{R}^n$ is a subspace.

Theorem: $\dim W + \dim W^\perp = n$

Theorem: $(W^\perp)^\perp = W$.

Key examples:

$\mathbf{Row}\ (A)^{\perp} = \mathbf{Null}\ (A).$

$\mathbf{Col}\ (A)^{\perp} = \mathbf{Null}\ (A^T)$

examples - fill in $\mathbb{R}^3$→

$\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is an **orthogonal set** if $\mathbf{u}_i \perp \mathbf{u}_j$ whenever $i \neq j$.

prove it!→

Theorem. If $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is an orthogonal set and $\mathbf{u}_i$ are non-zero, then they are linearly independent.

An orthogonal set of non-zero vectors is a basis for its span.

Definition: An **orthogonal basis** of W is a basis which is an orthogonal set.

←just a change of perspective

Theorem. If $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is an orthogonal basis for W and we want to decompose a vector $\mathbf{y} \in W$ as

$$\mathbf{y} = c_1\mathbf{u}_1 + \cdots + c_k\mathbf{u}_k$$

then

$$c_j = \frac{\mathbf{y} \cdot \mathbf{u}_i}{\mathbf{u}_i \cdot \mathbf{u}_i}.$$

←examples!!

Things get better if we normalize: $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is an **orthonormal set** if it is an orthogonal set of **unit** vectors.

examples→

Theorem. A matrix U has orthonormal columns if and only if

$$U^T U = I.$$

Theorem. If U has orthonormal columns then

$$(U\mathbf{x}) \cdot (U\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$$

Corollary If U has orthonormal columns then

- $\|U\mathbf{x}\| = \|\mathbf{x}\|$
- $\mathbf{x} \perp \mathbf{y} \iff U\mathbf{x} \perp U\mathbf{y}$

If U is a square matrix with orthonormal columns then it is called **an orthogonal matrix**.

Note: $U^T U = I$, and U a square matrix means $UU^T = I$.

meaning the rows are also orthonormal!

Suppose $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is an orthogonal basis for W but $\mathbf{y} \notin W$. What can we do?

Theorem. There is a unique expression

$$\mathbf{y} = \mathbf{y}_W + \mathbf{y}_{W^\perp}$$

where $\mathbf{y}_W \in W$ and $\mathbf{y}_{W^\perp} \in W^\perp$.

In fact

$$\mathbf{y}_W = c_1\mathbf{u}_1 + \cdots + c_k\mathbf{u}_k$$

where

$$c_j = \frac{\mathbf{y} \cdot \mathbf{u}_i}{\mathbf{u}_i \cdot \mathbf{u}_i}.$$

←examples!!

If $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is orthonormal, the formula gets simpler:

$$\mathbf{y}_W = (\mathbf{y} \cdot \mathbf{u}_1)\mathbf{u}_1 + \cdots + (\mathbf{y} \cdot \mathbf{u}_k)\mathbf{u}_k$$

Writing $U = [\mathbf{u}_1 \ldots \mathbf{u}_k]$ this gives

$$\mathbf{y}_W = UU^T\mathbf{y}$$

We call $\mathbf{y}_W$ the perpendicular projection of y on W.

Notation:

$$\mathbf{y}_W = \mathrm{proj}_W \mathbf{y}$$

So

$$\mathbf{y} = \mathrm{proj}_W \mathbf{y} + \mathrm{proj}_{W^\perp} \mathbf{y}.$$

Then $\mathbf{y}_W$ is the vector in W closest to W.

Gram-Schmidt Algorithm. Say $\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ is a basis for W.

Think about $W_i = \textbf{Span}\ \{\mathbf{x}_1, \mathbf{x}_i\}$

We construct inductively $\mathbf{v}_i$ as follows:

examples→

$$\mathbf{v}_1 = \mathbf{x}_1$$

$$\mathbf{v}_2 = \text{Proj}_{W_1^\perp} \mathbf{x}_2 \qquad {\scriptstyle =\mathbf{x}_2 - \text{Proj}_{W_1} \mathbf{x}_2}$$

$$= \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1$$

$$\vdots$$

$$\mathbf{v}_k = \text{Proj}_{W_{k-1}^\perp} \mathbf{x}_k$$

$$= \mathbf{x}_k - \left(\tfrac{\mathbf{x}_k \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 + \cdots + \tfrac{\mathbf{x}_k \cdot \mathbf{v}_{k-1}}{\mathbf{v}_{k-1} \cdot \mathbf{v}_{k-1}}\mathbf{v}_{k-1}\right)$$

We can normalize: $\mathbf{u}_i = \dfrac{1}{\|\mathbf{v}_i\|}\mathbf{v}_i$.

QR factorization

Matrix interpretation of Gram-Schmidt: say A has linearly independent columns

$$A = [\mathbf{x}_1 \cdots \mathbf{x}_n]$$

and say $Q = [\mathbf{u}_1 \cdots \mathbf{u}_n]$.

Since $\mathbf{Span}\ (\mathbf{u}_1, \ldots, \mathbf{u}_i) = \mathbf{Span}\ (\mathbf{x}_1, \ldots, \mathbf{x}_i)$ we have

$$\mathbf{x}_i = r_{1i}\mathbf{u}_1 + \cdots + r_{ii}\mathbf{u}_i$$

meaning $\mathbf{x}_i = Q \begin{bmatrix} r_{1i} \\ \vdots \\ r_{ii} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$

So $A = [\mathbf{x}_1 \cdots \mathbf{x}_n] = Q[\mathbf{r}_1 \cdots \mathbf{r}_n]$.

In other words $A = QR$, where Q has orthonormal columns, and R is invertible upper triangular.

Note: $R = Q^T A$

Least squares approximation.

Say **Col** A does not span $\mathbf{b}$.

You still want to approximate

$$A\mathbf{x} \approx \mathbf{b}.$$

We'll calculate $\hat{\mathbf{x}}$ such that

$$\|A\hat{\mathbf{x}} - \mathbf{b}\|$$

is minimized. Then we'll apply.

The closest an element $A\hat{\mathbf{x}}$ of the column space gets to $\mathbf{b}$ is

Gram-Schmidt? ouch!→

$$\hat{\mathbf{b}} = \text{proj}_{\mathbf{Col}\ A}\,\mathbf{b}.$$

Now $\hat{\mathbf{b}} \in \mathbf{Col}\ A$, so we can solve

$$A\hat{\mathbf{x}} = \hat{\mathbf{b}}.$$

This seems hard. But…

$$A\hat{\mathbf{x}} = \hat{\mathbf{b}} \;=\; \text{proj}_{\mathbf{Col}\ A}\,\mathbf{b}.$$

$$\Longleftrightarrow$$

$$(\mathbf{b} - \hat{\mathbf{b}}) \;\perp\; \mathbf{Col}\ A$$

$$= \mathbf{Row}\ A^T$$

$$\Longleftrightarrow$$

$$A^T(\mathbf{b} - A\hat{\mathbf{x}}) = 0$$

$$\Longleftrightarrow$$

$$\boxed{A^T A\ \hat{\mathbf{x}} = A^T\mathbf{b}}$$

Linear regression

Theoretically a system behaves like

$$y = \beta_0 + \beta_1 x,$$

and you want to find β_i.

You run experiments which give data points (x_i, y_i). They will not actually lie of a line.

You get the approximate equations

$$\begin{array}{ccc} \beta_0 + & x_1\beta_1 & \approx y_1 \\ \beta_0 + & x_2\beta_1 & \approx y_2 \\ \vdots & & \vdots \\ \beta_0 + & x_n\beta_n & \approx y_n \end{array}$$

In matrix notation:

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \approx \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

or

$$X\boldsymbol{\beta} \approx \mathbf{y}.$$

To solve it we take

$$X^T X \boldsymbol{\beta} = X^T \mathbf{y}$$

We can solve directly, or expand:

$$\begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$