Stat 608 – Chapter 1

Theme of The Class

- It makes sense to base inferences or conclusions only on valid models.

- A key step in any regression model, then, is to identify and address model weaknesses.

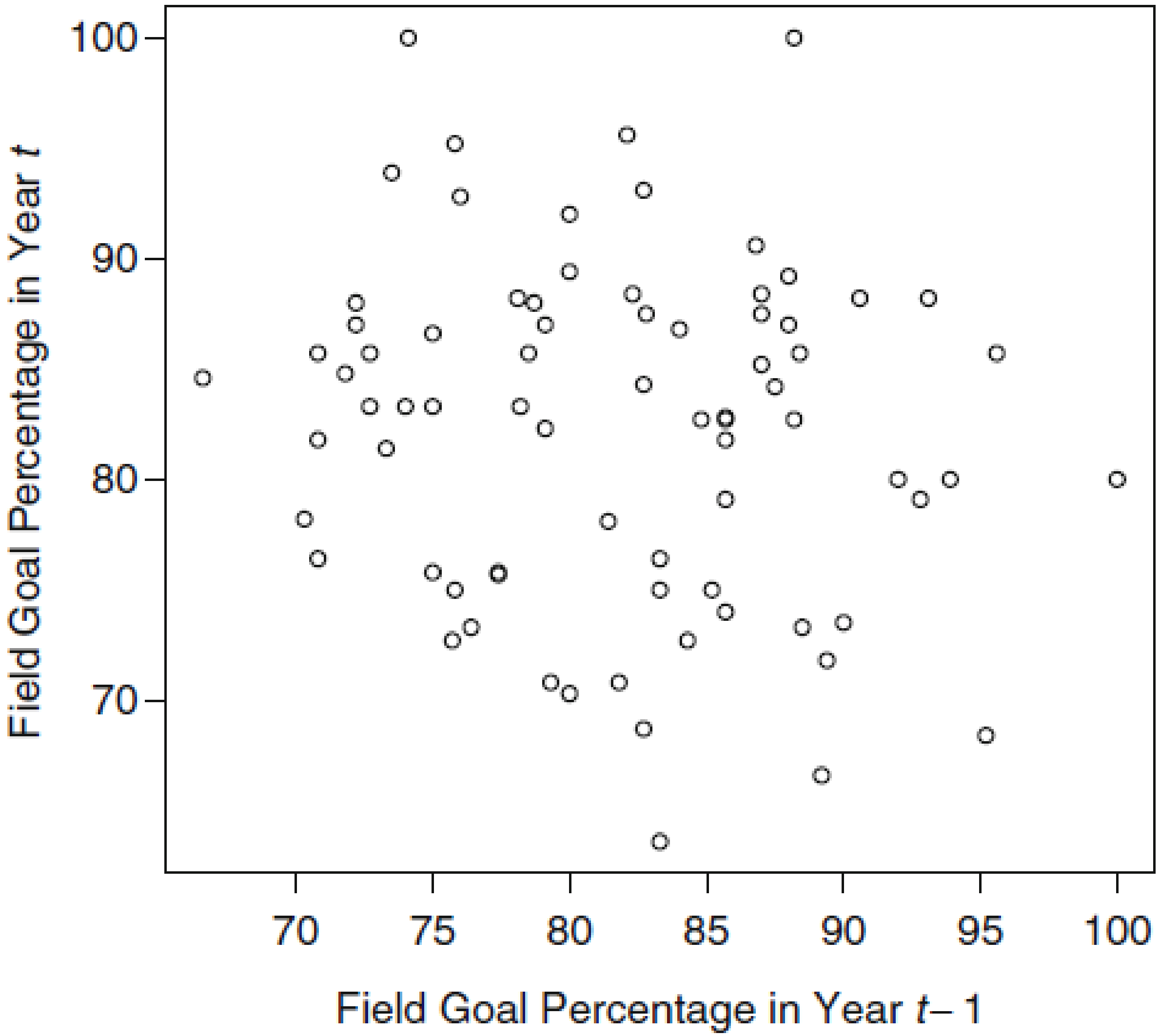
- There are two main parts of the class:
  1. Choose appropriate diagnostic procedures for building and assessing validity of regression models.
  2. Understand underlying mathematical properties of regression models in order to make appropriate decisions.

Example

In the Keeping Score column by Aaron Schatz in the Sunday November 12, 2006 edition of the *New York Times entitled "N.F.L. Kickers Are Judged on the* Wrong Criteria" the author makes the following claim:

There is effectively no correlation between a kicker's field goal percentage one season and his field goal percentage the next.

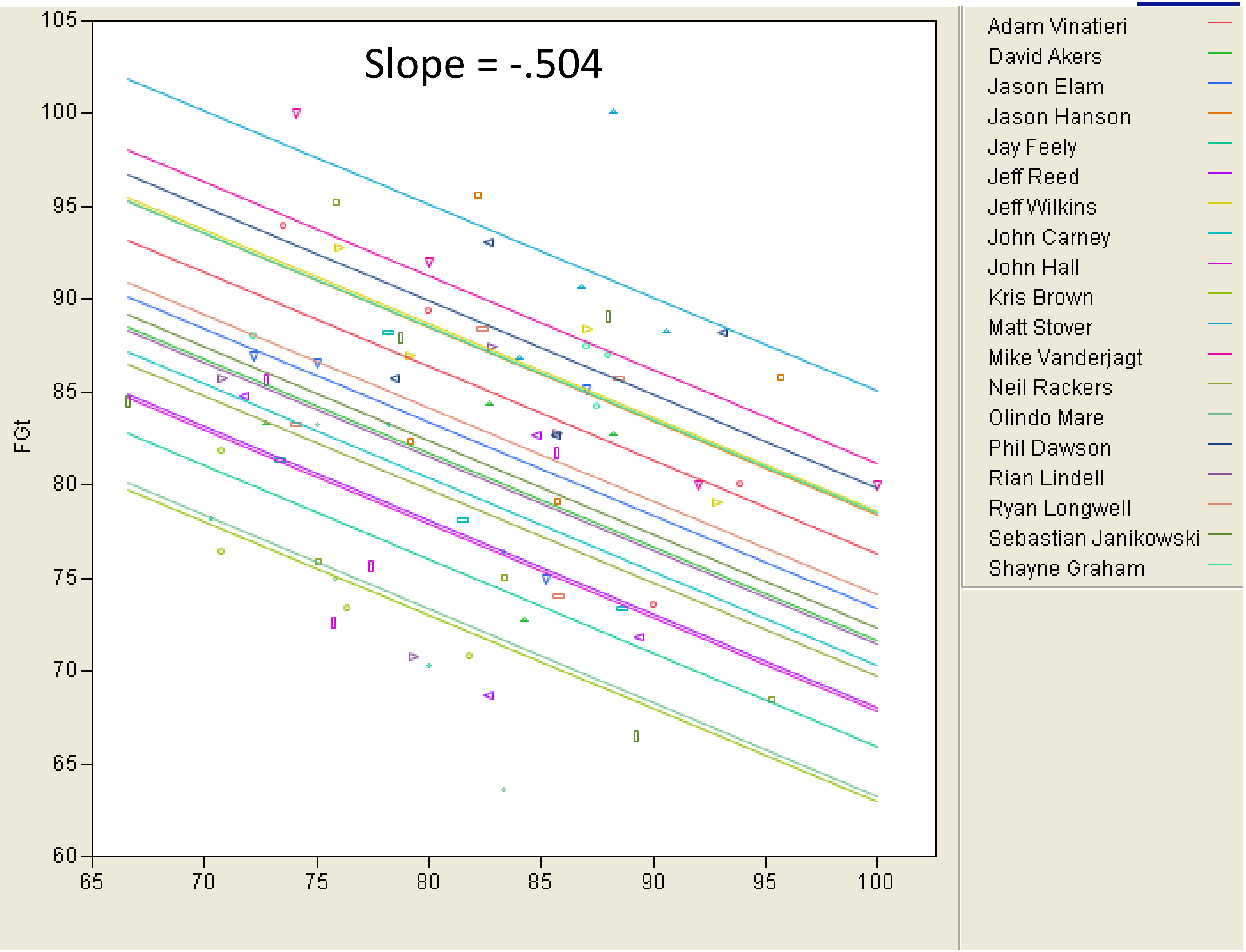
Slope = - .15



100
90
80
70
Field Goal Percentage in Year t
70
75
80
85
90
95
100
Field Goal Percentage in Year t– 1

Valid Model?

However, this approach is **fundamentally flawed as it fails to take into account the potentially** different abilities of the 19 kickers. In other words this approach is based on an **invalid model.**



Slope = -.504
FGt
105
100
95
90
85
80
75
70
65
60
65
70
75
80
85
90
95
100
Adam Vinatieri
David Akers
Jason Elam
Jason Hanson
Jay Feely
Jeff Reed
Jeff Wilkins
John Carney
John Hall
Kris Brown
Matt Stover
Mike Vanderjagt
Neil Rackers
Olindo Mare
Phil Dawson
Rian Lindell
Ryan Longwell
Sebastian Janikowski
Shayne Graham

Level of Mathematics

From former classes:

- Transpose, trace, determinant
- Addition / Subtraction
- Multiplication / Inverse
- Vector spaces, bases
- Logs
- Partial Derivatives

New material:

- Matrix Derivatives
- Expectation & Variance of Matrices

Square Matrices

- A square matrix has the same number of rows and columns.

Square

$$\begin{bmatrix} 3 & 0 & -1 \\ -1 & 2 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

Not Square

$$\begin{bmatrix} 3 & 0 \\ -1 & 2 \\ 0 & 1 \\ 2 & 4 \\ -2 & 0 \end{bmatrix}$$

Symmetric Matrices

- A symmetric matrix is symmetric about the *diagonal*. That is, $a_{12} = a_{21}$, and $a_{23} = a_{32}$, etc. Symmetric matrices must be square.

- Which one is symmetric?

$$\begin{bmatrix} 1 & -2 & 0 \\ -2 & 3 & -5 \\ 0 & -5 & 4 \end{bmatrix} \qquad \begin{bmatrix} 1 & -2 & 1 \\ 2 & -5 & 2 \\ 3 & -4 & 3 \end{bmatrix}$$

Symmetric Matrices

- The variance-covariance matrix is symmetric:
  - Variances for variables 1, 2, …, p are found in the matrix at locations (1,1), (2,2), …, (p,p).
  - The covariance for variables i and j can be found at locations (i, j) and (j, i).
- The correlation matrix is also symmetric, having 1’s on the diagonal and correlations on the off-diagonal.

Transpose

- To transpose A, use the rows of A for the columns of A’. (Or A^T, but the textbook uses A’.)

$$A = \begin{bmatrix} 3 & 0 & -1 \\ -1 & 2 & 1 \\ 0 & 1 & 0 \end{bmatrix}, A' = \begin{bmatrix} 3 & -1 & 0 \\ 0 & 2 & 1 \\ -1 & 1 & 0 \end{bmatrix}$$

- The equivalent of x^2 for a matrix X is X’X.

Trace

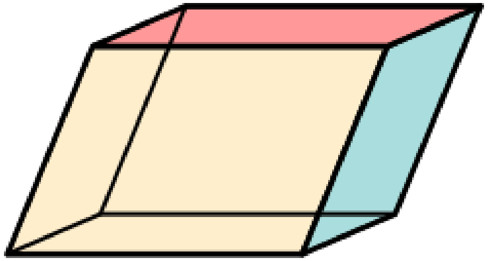
- The trace of a matrix A, tr(A), is the sum of the diagonal elements of A.

$$A = \begin{bmatrix} 3 & 0 & -1 \\ -1 & 2 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

Tr(A) = 3 + 2 + 0 = 5

Determinant

- A determinant is a single value that characterizes a square matrix.
- The determinant is the volume of a parallelepiped formed by the vectors of the matrix.
- The determinant of a covariance matrix is the generalized variance of a set of variables.



Determinant

- The determinant of a 2X2 matrix with elements as follows is ad – bc:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

- For example:

$$A = \begin{bmatrix} 2 & 3 \\ 0 & 1 \end{bmatrix}$$

det(A) = |A| = 2(1) – 3(0) = 2

Determinant

- Determinant of a 3X3 matrix:

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix}$$

Determinant

- Example:

$$A = \begin{bmatrix} 3 & 0 & -1 \\ -1 & 2 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

- $|A| = 3\{2(0) - 1(1)\} - 0\{-1(0) - 1(0)\} +$

$(-1)\{(-1)(1) - 2(0)\} = -3 - 0 + 1 = -2$

Addition

- To add, just add each element:

$$\begin{bmatrix} 3 & 0 & -1 \\ -1 & 2 & 1 \\ 0 & 1 & 0 \end{bmatrix} + \begin{bmatrix} 2 & -1 & 0 \\ 1 & 0 & 1 \\ 0 & 2 & -2 \end{bmatrix} = \begin{bmatrix} 5 & -1 & -1 \\ 0 & 2 & 2 \\ 0 & 3 & -2 \end{bmatrix}$$

- Matrices or vectors must be the same size to be added.

Subtraction

- To subtract, simply subtract each element:

$$\begin{bmatrix} 3 & 0 & -1 \\ -1 & 2 & 1 \\ 0 & 1 & 0 \end{bmatrix} - \begin{bmatrix} 2 & -1 & 0 \\ 1 & 0 & 1 \\ 0 & 2 & -2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & -1 \\ -2 & 2 & 0 \\ 0 & -1 & 2 \end{bmatrix}$$

- Matrices and vectors must be the same size to be subtracted.

Multiplication

- If A is a matrix, and c is a scalar, cA is found by multiplying c by every element of A:

$$c = 2$$

$$A = \begin{bmatrix} 3 & 0 & -1 \\ -1 & 2 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

$$cA = \begin{bmatrix} 6 & 0 & -2 \\ -2 & 4 & 2 \\ 0 & 2 & 0 \end{bmatrix}$$

Multiplication

- The dot product of two vectors is the sum of the products of the corresponding elements:

$$a = \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}, b = \begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix}, a \cdot b = -3 + 0 + 2 = -1$$

- Both vectors must be the same length for dot products to be defined.
- When the dot product of two vectors = 0, they are orthogonal.

Multiplication

- Matrix multiplication is more complicated than addition. To find the i^{th}, j^{th} element of matrix AB, take the dot product of the i^{th} row of A and the j^{th} column of B.

$$A = \begin{bmatrix} 3 & 0 & -1 \\ -1 & 2 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad B = \begin{bmatrix} 2 & -1 & 0 \\ 1 & 0 & 1 \\ 0 & 2 & -2 \end{bmatrix} \qquad AB = \begin{bmatrix} 6 & -5 & 2 \\ 0 & 3 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

- The number of columns of A must equal the number of rows of B.

- *Matrix multiplication is not commutative!*

~~Division~~ Inversion

- Actually, division for real numbers is multiplication by the inverse: instead of dividing by 2, multiply by ½.
- Inverses of diagonal matrices are easy:

$$\begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 5 \end{bmatrix}^{-1} = \begin{bmatrix} 1/3 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/5 \end{bmatrix}$$

~~Division~~ Inversion

- And inverting a 2X2 matrix isn't bad:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

+

Singular Matrices

- If the determinant of a matrix is 0, the inverse cannot be calculated, and the matrix is called *singular.*
- If two variables are perfectly correlated, one is a scalar multiple of the other, meaning there is a linear "combination" of one vector that gives the other. That is, perfectly correlated variables are not linearly independent, and X'X is singular.
- Indicator variables for *every* category of a categorical variable, plus an intercept, will make X'X singular.
- If there are more variables than observations, X'X will be singular.
- In linear models, if X'X is singular, the generalized inverse can be obtained, but the parameter estimates are not unique.

Commonly Used Matrices

- The multiplicative identity:

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- The additive identity:

$$0 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

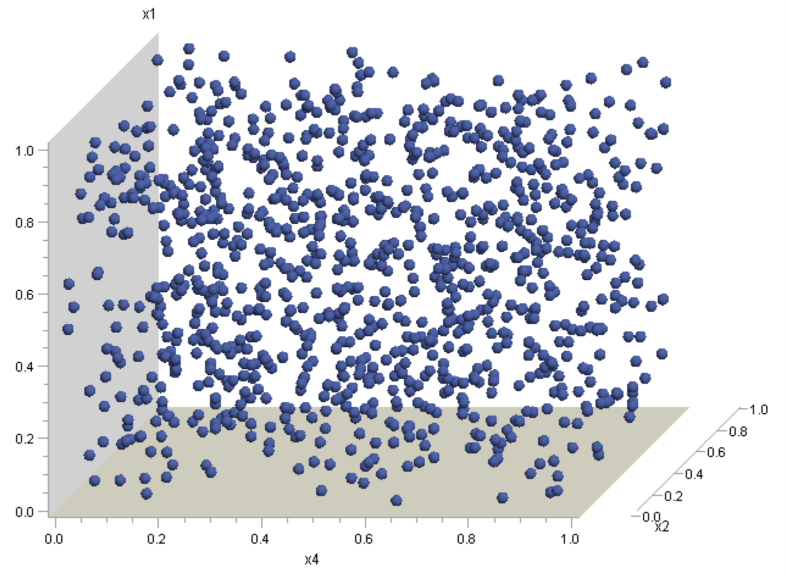
Orthogonal Bases in Vector Space

Recall that the determinant is the volume of a parallelepiped of dimension *p*.

- The volume of a parallelepiped for a given *p* is maximized when the axes are mutually perpendicular.
- Equivalently, the matrix is orthogonal (inner or dot product is 0) when the axes are mutually perpendicular (variables are independent).
- Data typically do not yield orthogonal covariance matrices without some transformation.
- It is useful to find an orthogonal basis to work with matrices in statistics.

Basis in Vector Space

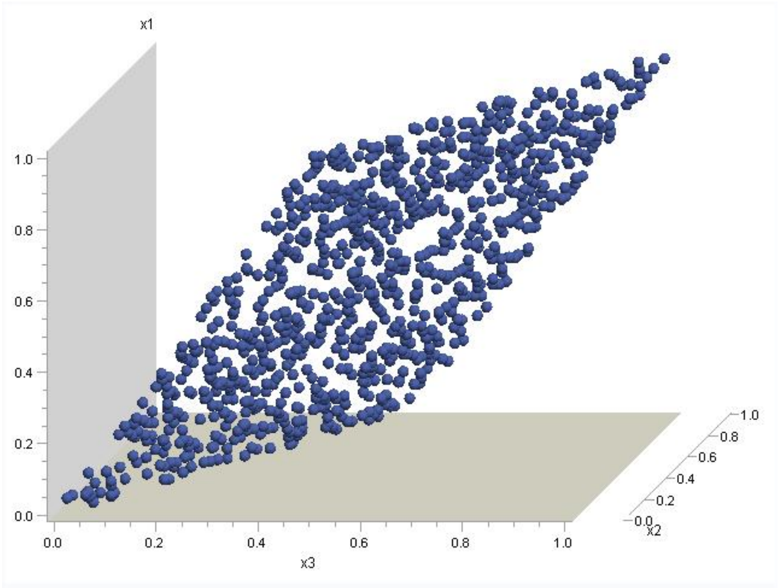
- Find an orthogonal basis for a data set with 1,000 observations (each observations is represented as a point on the graph, and can be thought of as a vector).



- Because the cube I_3 is uniformly populated, it is a good approximation for these three variables.

Basis in Vector Space

- But what about these observations?



- There are only 2 dimensions with variation.
- A square would be a better approximation.

Basis in Vector Space

- But what about many dimensions?
- If you have 15 variables, approximately how many orthogonal dimensions are there?
- Simple graphs of raw variables fail to provide insight.
- Matrix decomposition is useful in this case, and graphs can be constructed from the results of decomposition.

Expectation and Variance of Random Variables

Recall the definition of variance:

$$
\begin{aligned}
Var(X) &= E[(X - \mu_X)^2] \\
&= E[X^2] - \mu_X^2
\end{aligned}
$$

Covariance is similar. It is the numerator of correlation:

$$Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

+

Expectation and Variance of Random Variables

Assume X and Y are random variables, and a and b are constants. Then:

$$E[aX + b] = a\,E[X] + b$$

$$Var(aX) = a^2\,Var(X)$$

$$Cov(aX, bY) = ab\,Cov(X, Y)$$

$$Corr(X, Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

Expectation and Variance of Matrices

Assume **x** is a vector of random variables. Then:

$$E[\mathbf{x}] = E\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} = \boldsymbol{\mu}$$

Covariance Matrix

Assume $\mathbf{x}$ is a vector of random variables. Let σ_1^2 be the variance of random variable x_1, σ_2^2 the variance of variable x_2, and so on. Also let $\sigma_{1,2}$ be the covariance of variables x_1 and x_2, and so on. Then:

$$\mathrm{Var}(\mathbf{x}) = \mathrm{Cov}(\mathbf{x}) = \Sigma$$

$$= \begin{bmatrix} \sigma_1^2 & \sigma_{1,2} & \cdots & \sigma_{1,n} \\ \sigma_{2,1} & \sigma_2^2 & \cdots & \sigma_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n,1} & \sigma_{n,2} & \cdots & \sigma_n^2 \end{bmatrix}$$

Correlation Matrix

- To calculate the correlation matrix, divide by the appropriate standard deviations elementwise.

$$\mathbf{R} = \begin{bmatrix} 1 & r_{1,2} & \cdots & r_{1,n} \\ r_{2,1} & 1 & \cdots & r_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n,1} & r_{n,2} & \cdots & 1 \end{bmatrix}$$

Expectation and Variance of Matrices

Assume **X** and **Y** are matrices of random variables; **x** and **y** are vectors of random variables; and **A**, **B**, and **C** are constant matrices. Then:

$$E[\mathbf{AXB} + \mathbf{C}] = \mathbf{A}\,E[\mathbf{X}]\,\mathbf{B} + \mathbf{C}$$

$$Var(\mathbf{Ax}) = \mathbf{A}\,Var(\mathbf{x})\,\mathbf{A}'$$

$$Cov(\mathbf{AX}, \mathbf{BY}) = \mathbf{A}\,Cov(\mathbf{X}, \mathbf{Y})\,\mathbf{B}'$$

Quadratic Form

If $\mathbf{x}$ is a vector of random variables, and $\mathbf{A}$ is an n-dimensional symmetric matrix, then the scalar quantity $\mathbf{x}'\mathbf{A}\mathbf{x}$ is known as a quadratic form in $\mathbf{x}$.

Expectation of Quadratic Form

It can be shown that

$$E[\mathbf{x}'\mathbf{A}\mathbf{x}] = tr(\mathbf{\Sigma}\mathbf{A}') + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}$$

where **μ** and **Σ** are the expected value and variance-covariance matrix of **x**, respectively, and tr denotes the trace of a matrix.

This result only depends on the existence of **μ** and **Σ**; in particular, normality of **x** is *not* required.

Derivatives of Matrices

Derivatives of matrices are defined as partials with respect to the variable vector. For example:

$$f(\mathbf{x}) = \mathbf{a}'\mathbf{x} = \sum_{i=1}^{n} a_i x_i$$

$$\frac{\partial f}{\partial x_i} = \frac{\partial \left(\sum_{i=1}^{n} a_i x_i \right)}{\partial x_i} = a_i$$

$$\nabla_x f = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = \mathbf{a}$$

Derivatives of Matrices

The previous example was a derivative of a scalar-by-vector on this wikipedia page:

http://en.wikipedia.org/wiki/Matrix_calculus#Derivatives_with_matrices

It can also be found in the section on First Order derivatives in the Matrix Cookbook. (61) p. 9 (see eCampus)

Derivatives of Quadratic Forms

The gradient and Hessian:

$$f(\mathbf{x}) = \mathbf{x}'\mathbf{A}\mathbf{x} + \mathbf{b}'\mathbf{x} =$$

$$\nabla_x f = \frac{\partial f}{\partial \mathbf{x}} = (\mathbf{A} + \mathbf{A}')\mathbf{x} + \mathbf{b}$$

$$\frac{\partial^2 f}{\partial \mathbf{x} \partial \mathbf{x}'} = \mathbf{A} + \mathbf{A}'$$