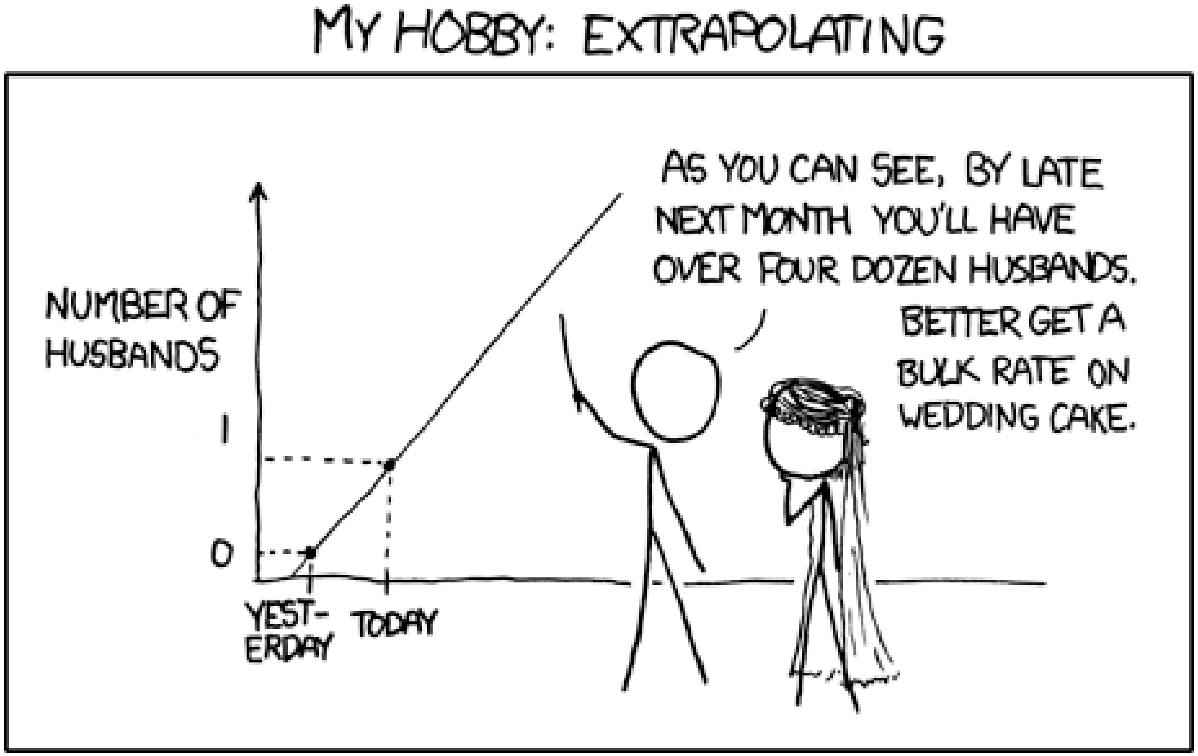
UNIVERSITY of WASHINGTON

Data Science UW Methods for Data Analysis

SVD and more Regression
Lecture 6
Nick McClure


MY HOBBY: EXTRAPOLATING
NUMBER OF HUSBANDS
1
0
YEST-ERDAY
TODAY
AS YOU CAN SEE, BY LATE NEXT MONTH YOU'LL HAVE OVER FOUR DOZEN HUSBANDS.
BETTER GET A BULK RATE ON WEDDING CAKE.

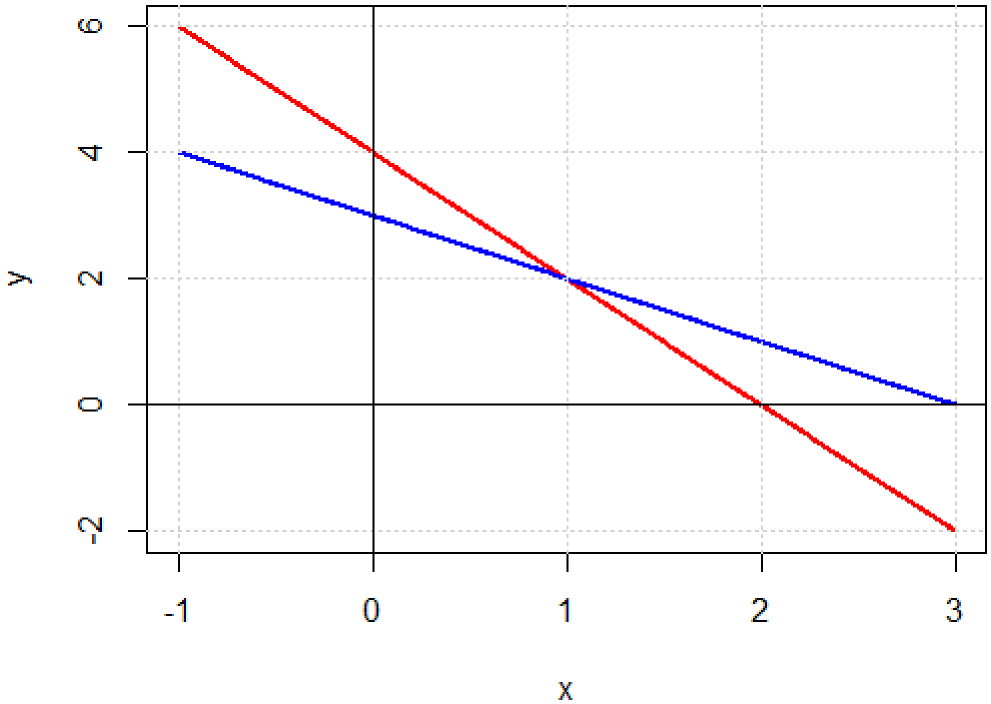
Topics

- Review
- Linear Algebra overview
- Decomposition Methods
- Lasso Regression
- Ridge Regression
- Logistic Regression
- Binary Classification

Motivation

> Consider the system of equations

$$y = -2x + 4$$
$$y = -x + 3$$



Matrix Notation

> When we consider multiple equations, we can write it in short-hand matrix notation.

$$y = -2x + 4$$
$$y = -x + 3$$

$$y = \begin{bmatrix} -2 \\ -1 \end{bmatrix} x + \begin{bmatrix} 4 \\ 3 \end{bmatrix}$$

$$y = \boldsymbol{M}x + \boldsymbol{B}$$

> **M**, and **B** are matrices that both have dimension (2x1).

Linear Algebra

- Matrix: a rectangular array of values, with dimensions n by m (n rows, m columns).
- Vector: a one dimensional array of values (n or m = 1).
- Square matrix: a n x n matrix.
- Identity matrix: a square matrix with 1’s on the diagonal and 0’s elsewhere.

Matrix Operations

> Matrix Transpose: Change rows into columns and vice-versa.

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

$$(A_{n\times m}B_{m\times p})^T = B^T_{p\times m}A^T_{m\times n}$$

> Matrix Trace (square matrices only): sum of the diagonals.

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

$$tr(A) = 1 + 3 + 0 = 4$$

Matrix Operations

> Matrix addition/subtraction

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

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

$$A + B = \begin{bmatrix} 3 & 2 & 0 \\ -3 & 3 & 2 \end{bmatrix}, A - B = \begin{bmatrix} -1 & 4 & -2 \\ 3 & -1 & 2 \end{bmatrix}$$

Matrix Operations

> Matrix multiplication
  – 1st matrix row multiplied element wise with 2nd matrix column, then added up.

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

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

$$A \cdot B^T = \begin{bmatrix} 1 \cdot 2 + 3 \cdot -1 + (-1) \cdot 1 & 1 \cdot -3 + 3 * 2 + (-1) * 0 \\ 0 \cdot 2 + 1 \cdot -1 + 2 \cdot 1 & 0 \cdot -3 + 1 \cdot 2 + 2 \cdot 0 \end{bmatrix}$$

$$A \cdot B^T = \begin{bmatrix} -2 & 3 \\ 1 & 2 \end{bmatrix}$$

$$A_{(2\times3)} \cdot B^T_{(3\times2)} = [A \cdot B^T]_{(2\times2)}$$

Linear Algebra

> Algebraic Properties of Matrices:

– Add/subtract matrices: Must be of the same dimensions

– Multiplication of matrices:

> Inner dimensions must match.

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \times \begin{bmatrix} j & k & l \\ m & n & o \\ p & q & r \end{bmatrix} =$$

$$\begin{bmatrix} \boxed{aj + bm + cp} & ak + bn + cq & al + bo + cr \\ dj + em + fp & dk + en + fq & dl + eo + fr \\ gj + hm + ip & gk + hn + iq & gl + ho + ir \end{bmatrix}$$

– [n x m] * [m x p] = [n x p]

– Note that matrix multiplication is not commutative

$$A \times B \neq B \times A$$

Linear Algebra

> Identity matrix: just like 1 is the multiplicative identity.
  - 5*1=5

$$\begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} \cdot \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}} = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$$

Identity matrix: a square matrix of zeros with 1's on the diagonal. Also written as I_{nxn}

Matrix Operations

> Matrix Identity

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

$$A \cdot I = A$$

> Matrix Inverse

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

$$A \cdot A^{-1} = I$$

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

> Matrix Division is not Defined.

> R examples of Matrix operations.

Linear Algebra

> For a 2x2 matrix:

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

> What if (ad-bc) = 0? That means that ad=bc or a/c = b/d.

> If a/c = b/d, then one of the columns is a multiple of the other!

> These columns are dependent on each other.

- If these were columns in our numerical data frame, then one column would be a multiple of the other.
- Examples: Using meters and Kilometers as separate predictors.

Linear Algebra

$$A \times X$$

- Given a sequence of data points in a matrix, X, which has dimensions 2xn:

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

- If we multiply by another matrix A (2x2):

$$A_{2x2} \times \begin{bmatrix} 1 & 2 & 0 \\ -2 & 3 & 1 \end{bmatrix}$$

- Then we can consider A, a matrix that transforms the points in X.
  - More on this later.

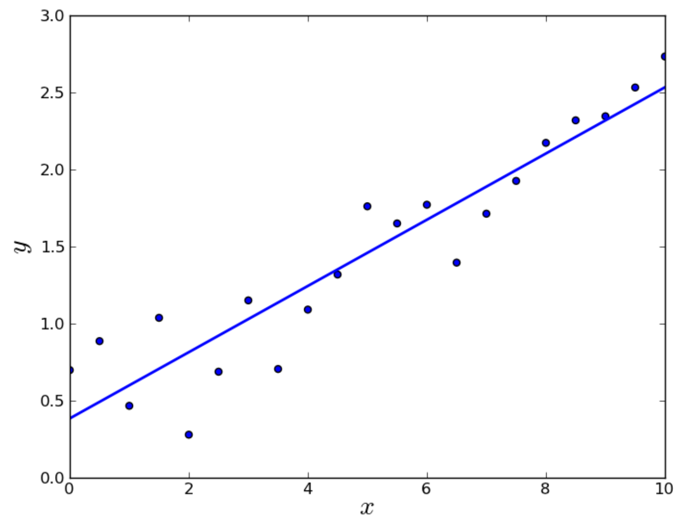
Linear Regression with Matrices

> Consider our linear regression formula.

$$y_i = mx_i + b + \epsilon_i, \epsilon_i \sim \mathrm{N}(0, \sigma^2)$$

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} mx_1 + b \\ \vdots \\ mx_n + b \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} b \\ m \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}$$



> Matrix Dimensions must agree:

(nx1)=(nx2)*(2x1) + (nx1)

Linear Regression with Matrices

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} b \\ m \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

Vector of Responses

The Design Matrix

Vector of Parameters

Vector of Error Terms (Or Residuals)

> In a succinct form:

$$Y = X\beta + \epsilon$$

Linear Regression Example

- Consider two points for linear regression. (3,4),(1,1)

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} b \\ m \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

$$\begin{bmatrix} 4 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} b \\ m \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 3 \\ 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 4 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 3 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} b \\ m \end{bmatrix}$$

$$\begin{bmatrix} 1 & 3 \\ 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 4 \\ 1 \end{bmatrix} = \begin{bmatrix} b \\ m \end{bmatrix}$$

$$\begin{bmatrix} -0.5 & 1.5 \\ 0.5 & -0.5 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \end{bmatrix} = \begin{bmatrix} b \\ m \end{bmatrix}$$

$$\begin{bmatrix} -0.5 \\ 1.5 \end{bmatrix} = \begin{bmatrix} b \\ m \end{bmatrix}$$

Linear Regression Example

- Consider two points for linear regression. (3,4),(3,1)

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} b \\ m \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

$$\begin{bmatrix} 4 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} b \\ m \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

- We run into issues!!!! (ad=bc, which means no inverse).

Linear Regression Example

> Consider multiple points for linear regression.

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} b \\ m \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

$$Y = X\beta + \epsilon$$

$$Y - X\beta = \epsilon$$

> How to solve for β? (Problem when X is not square!)

> Solution: Minimize ϵ^2, because $\epsilon^2 = \epsilon^T\epsilon$ is square.

Linear Regression Example

> Solution: Minimize ϵ^2:

$$\epsilon^2 = (Y - X\beta)^T (Y - X\beta)$$

$$\epsilon^2 = (Y^T - (X\beta)^T)(Y - X\beta)$$

$$\epsilon^2 = Y^T Y - (X\beta)^T Y - Y^T X\beta + (X\beta)^T X\beta$$

$$\epsilon^2 = Y^T Y - \underline{(X\beta)^T Y} - \underline{Y^T X\beta} + (X\beta)^T X\beta$$

> Note that the middle two terms end up being *1d-vectors.*

$$\epsilon^2 = Y^T Y - 2(X\beta)^T Y + (X\beta)^T X\beta$$

$$\epsilon^2 = Y^T Y - 2\beta^T X^T Y + \beta^T X^T X\beta$$

> To minimize, 'derivative' set to zero.

$$0 = 2X^T Y + 2X^T X\beta$$

$$2X^T X\beta = 2X^T Y$$

$$\beta = (X^T X)^{-1} X^T Y$$

Linear Regression Example

> Numerical Verification

$$\beta = (X^T X)^{-1} X^T Y$$

$$\begin{bmatrix} b \\ m \end{bmatrix} = \left(\begin{bmatrix} 1 & 3 \\ 1 & 1 \end{bmatrix}^T \begin{bmatrix} 1 & 3 \\ 1 & 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 3 \\ 1 & 1 \end{bmatrix}^T \begin{bmatrix} 4 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} b \\ m \end{bmatrix} = \begin{bmatrix} 2 & 4 \\ 4 & 10 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} b \\ m \end{bmatrix} = \begin{bmatrix} 2.5 & -1 \\ -1 & 0.5 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} b \\ m \end{bmatrix} = \begin{bmatrix} -0.5 & 1.5 \\ 0.5 & -0.5 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \end{bmatrix} = \begin{bmatrix} -0.5 \\ 1.5 \end{bmatrix}$$

Linear Regression Example

> This works with higher dimensions as well.

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} b \\ m \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

$$Y_{n\times 1} = X_{n\times 2}\beta_{2\times 1} + \epsilon_{n\times 1}$$

$$\beta_{2\times 1} = (X^T{}_{2\times n}X_{n\times 2})^{-1}X^T{}_{2\times n}Y_{n\times 1}$$

Multiple Linear Regression

- This works with Multiple Linear Regression as well.
- n points, F features.

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_F x_F + \epsilon_i$$

- Remember that a data point for this regression is:

$$(x_1, x_2, \ldots, x_F, y)$$

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{F1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \cdots & x_{Fn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_F \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

$$Y_{n\times 1} = X_{n\times F}\beta_{F\times 1} + \epsilon_{n\times 1}$$

$$\beta_{F\times 1} = ({X^T}_{F\times n} X_{n\times F})^{-1} {X^T}_{F\times n} Y_{n\times 1}$$

Why Independent Columns?

> Consider the 2x2 inverse formula:

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

> Even if $ad \neq bc$, and $ad \sim bc$, then we run into very large numbers and the inverse is hard to compute.

> This holds true for inverses of any size square matrix, when the columns are linearly dependent on each other, the inverse doesn't exist.

The Determinant

- Consider the 2x2 inverse formula:

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

- The number: $ad - bc$ is called the determinant.
- The determinant is very special and important.
  - It governs when a matrix has an inverse.
  - It tells us how linearly related columns are.
  - It is physically linked to the volume of the ‘unit’ box which we apply the matrix as a transformation.
    - The identity matrix is exactly the unit box and has a determinant of 1.
    - If we multiply a row by a constant, the determinant is multiplied by that constant. (physically expanding or shrinking the box).
    - This is also used in physics to keep conservation laws by
    - forcing a volume to stay at 1.

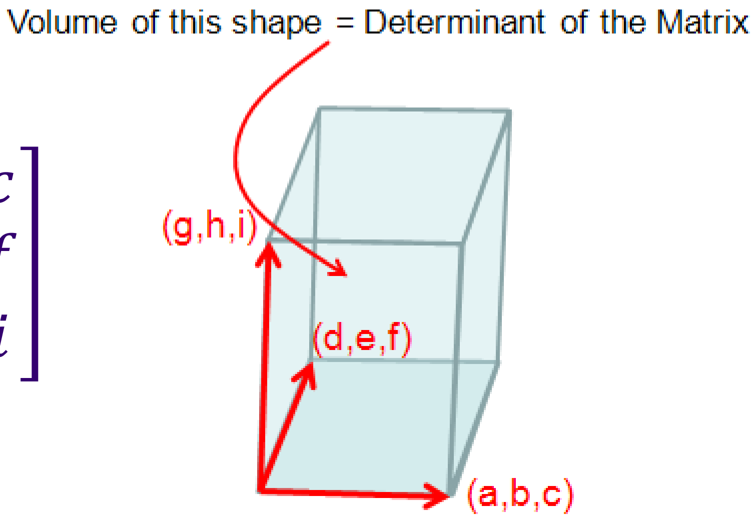
The Determinant

> Consider the 2x2 inverse formula:

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

> The number: $ad - bc$ is called the determinant.

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$$



Linear Algebra

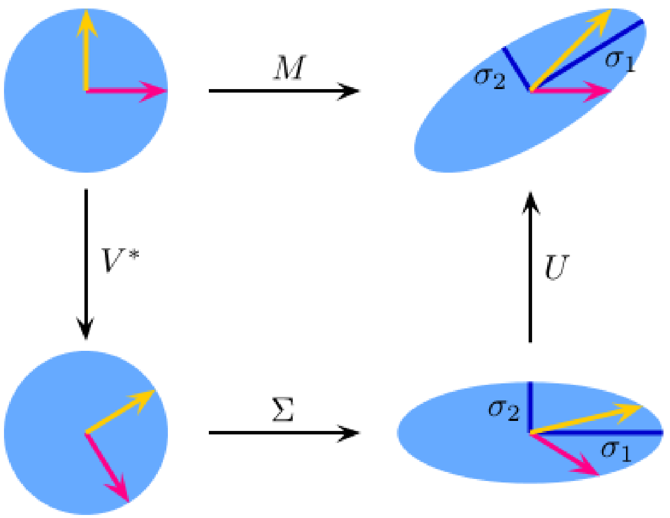
- Eigenvalues: Given a nxn matrix, A, λ is an eigenvalue if there exists a vector X such that:

$$AX = \lambda X$$

- Finding the eigenvectors of A involves lots of computation.
- If A rotates and shifts a vector X, then we can think of eigenvalues as a geometric hinge on which the ‘A’ operation acts.
- Eigenvalues have corresponding eigenvectors.
- This may seem insignificant at the moment, but eigenvalues and eigenvectors play an important role in manipulating our data.

Linear Algebra

- Matrix Decompositions allow us to write a matrix, M, in many different forms.
- The one that is the most used, is Singular Value Decomposition (SVD).
- The SVD is a way to express a transformation from one nxn space (the space M lies in) to another nxn space by writing M as a product of three matrices, say M=VSU.



$$M = U \cdot \Sigma \cdot V^*$$

Linear Algebra

- These three matrices, say, V,S,U, (M = V*S*U), have very specific properties that we can use to our advantage when describing a data set.
- R-demo.

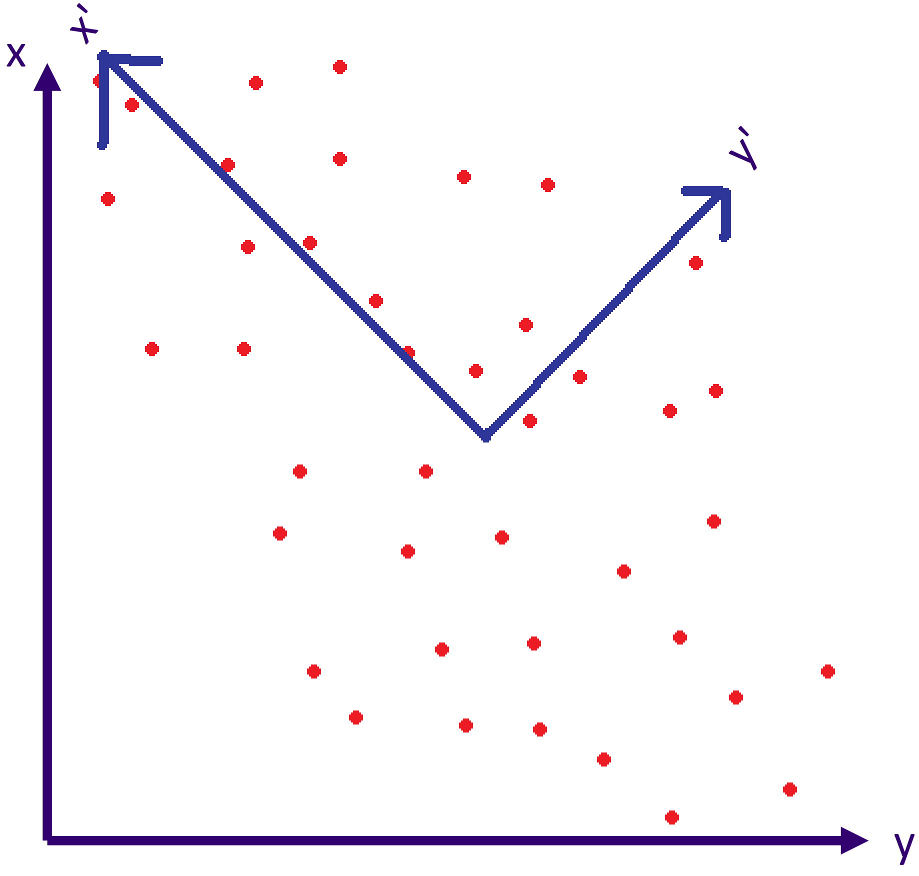
Deriving Independent Features from Dependence

- With larger data sets, we've seen that no matter the quality, we can find a explanatory feature.
- If we consider our data as a matrix, we know that having dependent columns is a problem.
- Solutions:
  - Remove columns that do not contain enough 'information'.
    - Too much missing data.
    - Low Variance.
  - Remove columns that are correlated
  - Maybe we can transform our data such that our data is more independent?

Possible Data Transformations

> If two variables are correlated, we can transform the data to directions in which they are not correlated.

> These new axes are called the Principal Components.



SVD

- This transformation is called the Singular Value Decomposition, or SVD.
- It holds true for as many features (dimensions) as we wish to choose, up to the number of original dimensions.
- Each of the new axes is some function of all the old axes.
- The SVD assures us that:
  - The first axis explains the most variation, the second axis the most variation after the first, and so on.
  - All axes are right-angled to each other (orthogonal).
- Usually, we keep less than the original amount of axes, so that we can reduce the amount of dimensions we have to keep track of.

SVD

- Know that instead of our original system:

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots$$

- We now have the system:

$$y_i = \beta_0 + \beta_1 f_1(x_1, x_2, \dots) + \beta_2 f_2(x_1, x_2, \dots) + \cdots$$

- The *f* functions are called our principle components.
- The *f* function outputs are guaranteed to be independent of each other.
- We can no longer interpret our linear model coefficients!

SVD

- SVD returns the same amount of components as our number of features.
- Since these are *all* orthogonal, the first few will explain much more variance than the last few axes. How do we decide how many to keep?
- We look at the magnitude of the associated eigenvalues for each principal component.
- R-demo.

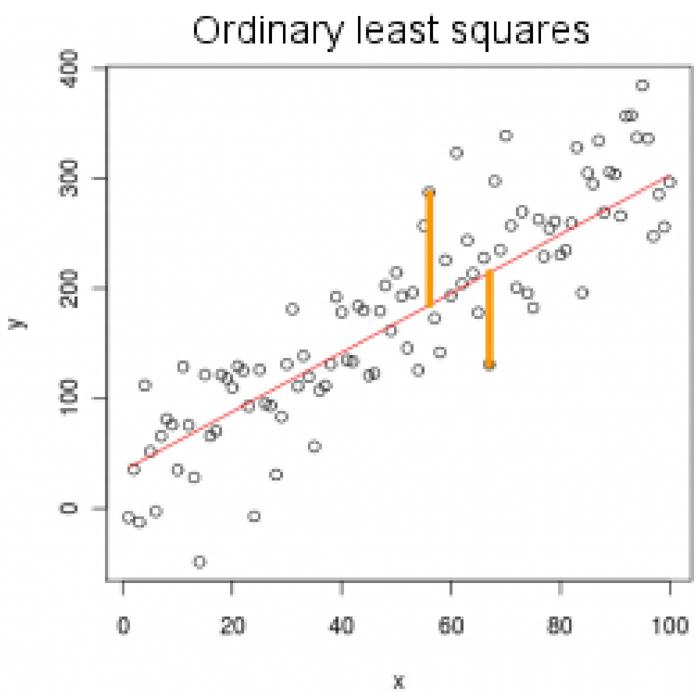
SVD

- This seems like an awful lot of work for little improvement and loss of interpretability.
- But note that we lost the dependence in the data set!
- There are other applications as well…

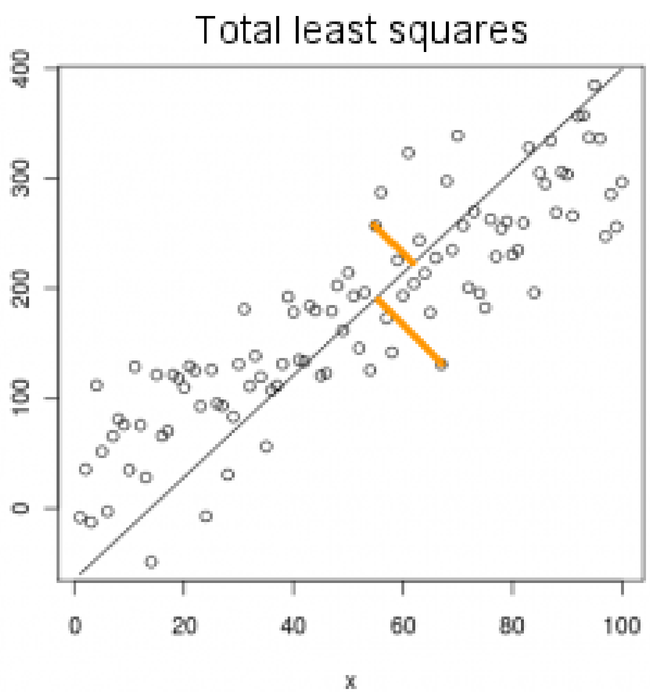
SVD, as a type of regression

> Also, looking at the first principal component, we can consider SVD as a new type of regression, which is called total least squares. (Also called Deming regression or PCA Regression)

Regressing y on x



SVD Primary Principal Component



> R demo

SVD, as a type of regression

- When to use total least squares:
  - If we want to control for error in x as well as y.
  - We are minimizing the distance from the point to the line as opposed to the distance between the y-values.
- R-squared doesn’t really apply here, at least in the way we have defined it.

SVD, as a way to compress information

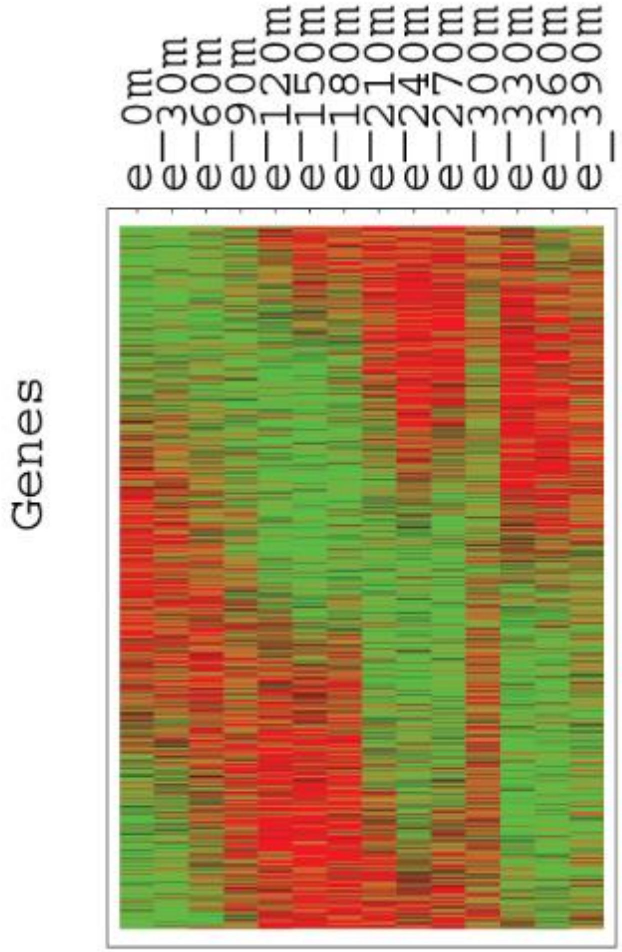
- We can group together similar points via SVD and store them as multiples of principal components.

- R-demo.

SVD, as a way to cluster data

> We can group together similar points via which SVD component is closest to representing original point.

- One of the most common uses is clustering individuals or genes as it pertains to RNA expression.
- In the microarray to the right, red represents absence of expression and green represents over expression.
- Each row is a gene (thousands of them) and each column is a sample (or patient).



Ridge Regression

- Ridge regression is a way to limit the amount of independent variables in the regression.
- Our regular least squares criterion minimizes the least squares of the error plus a regularization term that is a product of a constant and the sum of squared coefficients :

$$\min \sum (y - y_i)^2 + \alpha \sum \beta^2$$

- Essentially this is preventing the partial slope terms from getting too large.

Lasso Regression

- Lasso regression is another way to limit the amount of independent variables in the regression.
- Our regular least squares criterion minimizes the least squares of the error:

$$\min \sum (y - y_i)^2$$

- Lasso regression minimizes the same with the addition of a 'regularization' term:

$$\min \sum \left(y - y_j\right)^2 \quad \text{Such that} \quad \sum |\beta_i| < \lambda$$

- Here, y is the predicted for j points. There are i terms with beta coefficients. Lambda is a fixed value that limits the betas.

Using Linear Regression to Predict Limited Dependent Variables

> Let’s say we wanted to predict if someone evacuated their home during hurricane Katrina.

> R demo.

Logistic Regression

- The purpose of logistic regression is to use linear regression to predict a limited dependent variable.
- Usually our dependent variable has 2 outcomes (1 or 0) or occurrence.
- Examples:
  - Bank gives a yes (1) or no (0) outcome to loan applications.
  - Success/Failures of clinical trials.
  - Morbidity outcomes.
  - Marketing outcomes (will a user click on an add).
- Logistic predictions will result in a probability of success.

Logistic Regression

> Logistic regression is also called the 'logit' model:
> Original model:

$$y_i = \beta_0 + \beta_1 x_1 + \varepsilon_0$$

> Logit model:

$$\underbrace{\ln\left[\frac{p_i}{1-p_i}\right]}_{\text{Log-odds-ratio}} = \beta_0 + \beta_1 x_1 + \varepsilon_0$$

> So estimated probabilities follow: (solving for p)

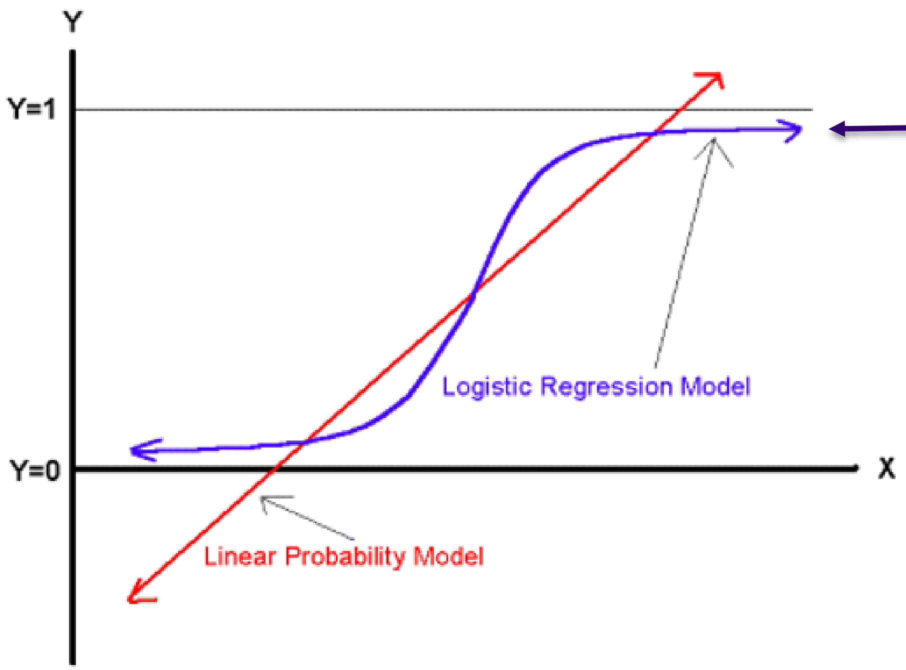
$$p_i = \frac{1}{1 + \exp(-(\beta_0 + \beta_1 x_1))}$$

Logistic Regression

$$p_i = \frac{1}{1 + \exp(-(\beta_0 + \beta_1 x_1))}$$

> As $(\beta_0 + \beta_1 x_1)$ gets really big, p approaches 1.

> As $(\beta_0 + \beta_1 x_1)$ gets really small, p approaches 0.



$$\frac{1}{1 + \exp(-x)}$$

Logistic Regression

- Differences between linear and logistic regression.
- Predictions
  - Linear regression outcomes are unbounded.
  - Logistic regression outcomes are bounded between 0 and 1.

$$p_i = \frac{1}{1 + \exp(-(\beta_0 + \beta_1 x_1))}$$

- Error distribution
  - Linear regression errors are normally distributed.
  - Logistic regression errors are Bernoulli distributed.

- R demo

Assignment

- Complete Homework 6:
  - Perform SVD regression on Crime-community data.
  - You should submit:
    - A R-script (PRODUCTION) and a text/log write up.
  - Read Introduction to Data Science, Chapter 16.
  - Read two articles about p-values and reproducible research.
  - http://blogs.plos.org/publichealth/2015/06/24/p-values/
  - http://www.science20.com/the_conversation/half_of_biomedical_studies_arent_reproducible_and_what_we_need_to_do_about_that-156696