Linear Algebra Bootcamp

The 90 Minute Primer

Linear Algebra

- Study of functions/surfaces/spaces that do not bend or curve.
- Scalar multiplication and addition.

Matrices and Vectors

- Arrays or lists of numbers.
- Indexed first by row (i) then by column (j) $\mathbf{X}_{ij}$ $\mathbf{v}_i$

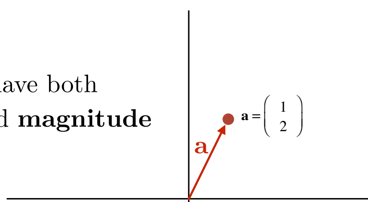
$$\mathbf{X} = \begin{pmatrix} 1 & 8 & 7 & -1 \\ 4 & 9 & 6 & 9 \\ -3 & -4 & 9 & 8 \\ -2 & -1 & 10 & 3 \\ 3 & -3 & 1 & 7 \end{pmatrix} \quad \mathbf{v} = \begin{pmatrix} 0.3 \\ -1 \\ 1.2 \\ -1 \end{pmatrix}$$

X_{23} v_2

Vectors/Points

(Geometrically)

Vectors have both **direction** and **magnitude**



Direction arrow points from origin to the coordinate *point*

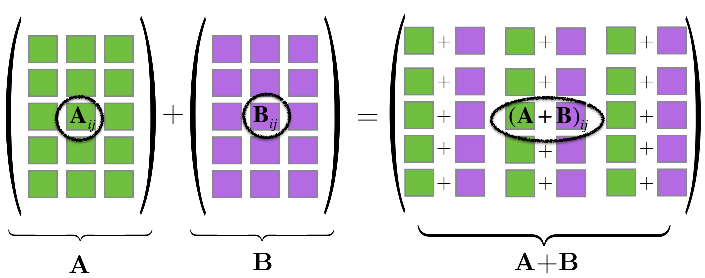
Magnitude is the length of that arrow #pythagoras

Matrix Arithmetic

(multi-dimensional math)

Addition

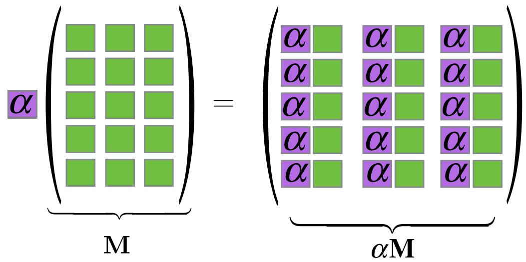
Element-wise



$$(\mathbf{A}+\mathbf{B})_{ij} = \mathbf{A}_{ij} + \mathbf{B}_{ij}$$

Scalar Multiplication

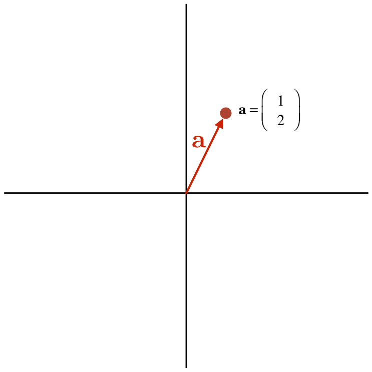
Element-wise



$$(\alpha \mathbf{M})_{ij} = \alpha \mathbf{M}_{ij}$$

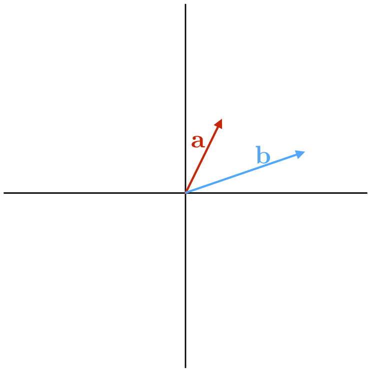
Scalar Multiplication

(Geometrically)



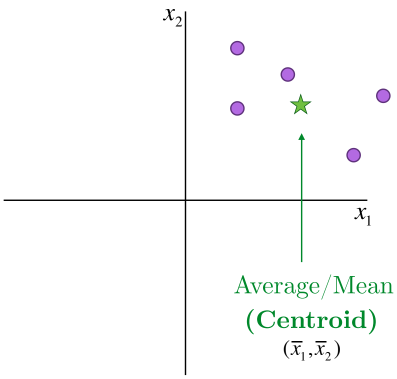
Vector Addition

(Geometrically)



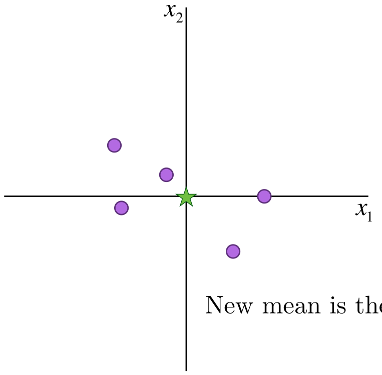
Example: Centering the data

(Geometrically)



Example: Centering the data

(Geometrically)

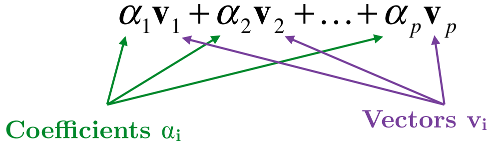


New mean is the origin (0,0)

Linear Combinations

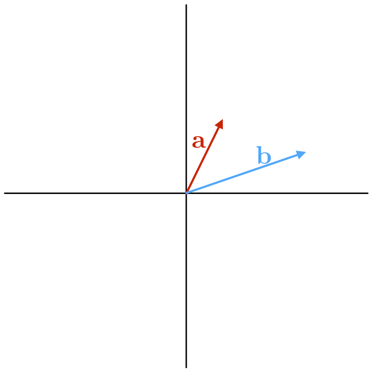
A linear combination of vectors is a just weighted sum:

$$\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \ldots + \alpha_p \mathbf{v}_p$$



Linear Combinations

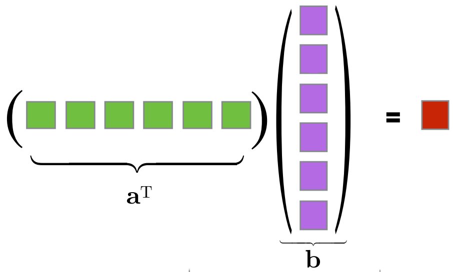
(Geometrically)



Multiplication

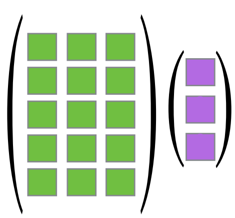
Inner Product

(row x column)

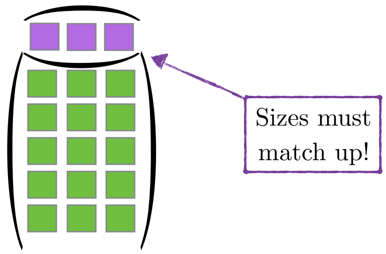


$$\mathbf{a}^T\mathbf{b} = \sum_{i=1}^{n} a_i b_i$$

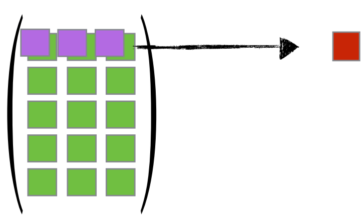
Matrix-Vector Multiplication (Inner-product view)



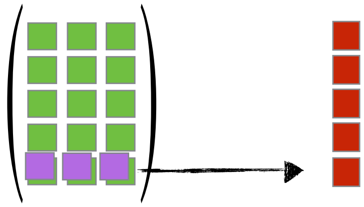
Matrix-Vector Multiplication (Inner-product view)



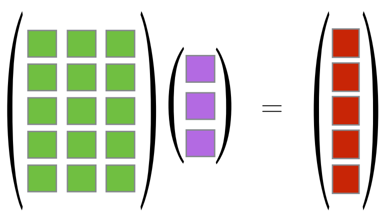
Matrix-Vector Multiplication (Inner-product view)



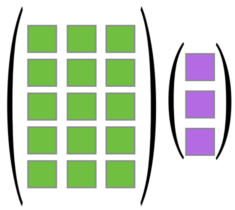
Matrix-Vector Multiplication (Inner-product view)



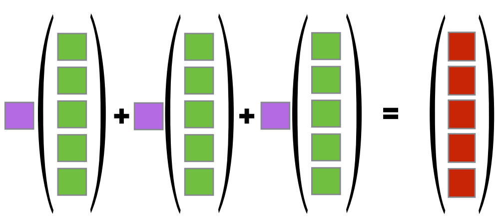
Matrix-Vector Multiplication (Inner Product View)



Matrix-Vector Multiplication (Linear Combination View)

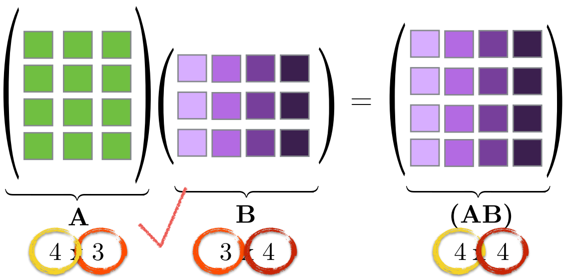


Matrix-Vector Multiplication (Linear Combination View)



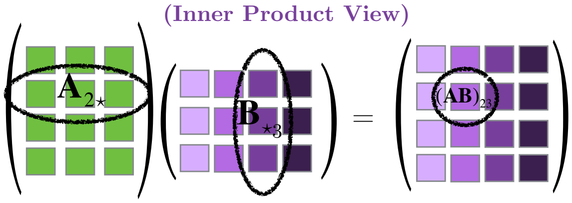
Matrix-Matrix Multiplication

Just a collection of matrix-vector products
(linear combinations) with different coefficients.



Matrix-Matrix Multiplication

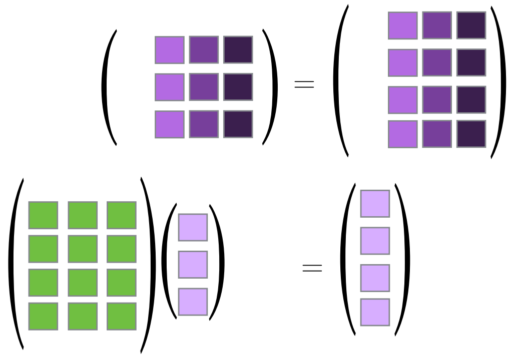
(Inner Product View)



$$(\mathbf{AB})_{ij} = \mathbf{A}_{i\star}\mathbf{B}_{\star j}$$

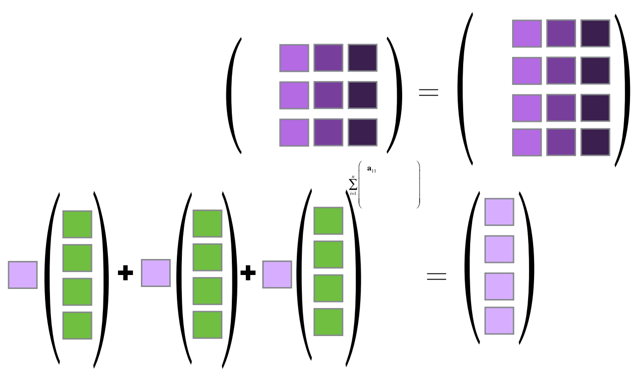
Matrix-Matrix Multiplication

(Linear Combination View)



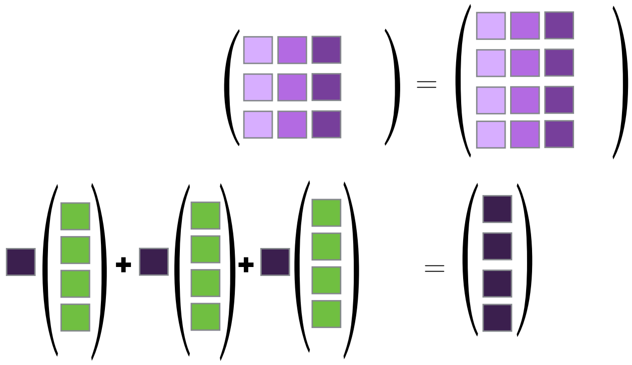
Matrix-Matrix Multiplication

(Linear Combination View)



Matrix-Matrix Multiplication

(Linear Combination View)



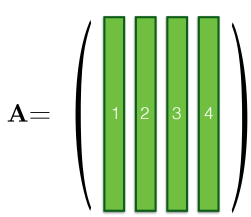
Matrix-Matrix Multiplication

- MATRIX MULTIPLICATION IS NOT COMMUTATIVE! $\mathbf{AB} \neq \mathbf{BA}$
- Just a collection of matrix-vector products (linear combinations) with different coefficients.
- Each linear combination involves the same set of vectors (the green columns) with different coefficients (the purple columns).
- This has important implications!

More Matrix Operations and Special Matrices

Transpose Operator

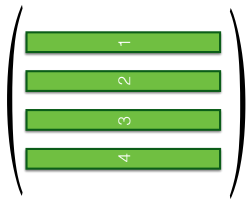
The transpose of a matrix **A**, written $\mathbf{A}^T$ is the matrix whose columns are the rows of **A**



Transpose Operator

The transpose of a matrix $\mathbf{A}$, written $\mathbf{A}^{\mathrm{T}}$ is the matrix whose columns are the rows of $\mathbf{A}$

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



The transpose is useful for forming meaningful matrix products, typically of the form $\mathbf{A}^{\mathrm{T}}\mathbf{A}$.

The Identity Matrix

The **identity matrix**, denoted $\mathbf{I}$ is to matrix algebra what the number 1 is to scalar algebra. The multiplicative identity.

When multiplied by the identity, a matrix remains unchanged.

$$\mathbf{AI} = \mathbf{A}$$
$$\mathbf{IA} = \mathbf{A}$$

The Identity Matrix

The identity matrix is a matrix of zeros with 1's on the main diagonal.

$$\boldsymbol{I_1} = [\mathbf{1}],\ \boldsymbol{I_2} = \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{bmatrix},\ \boldsymbol{I_3} = \begin{bmatrix} \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix},\ \cdots,\ \boldsymbol{I_n} = \begin{bmatrix} \mathbf{1} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{1} \end{bmatrix}.$$

The Inverse Matrix

The **inverse** of a matrix $\mathbf{A}$, should it exist, is denoted $\mathbf{A}^{-1}$, is a matrix for which multiplication by $\mathbf{A}$ results in the identity matrix.

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

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

The Inverse Matrix

All operations involving "cancelling" terms must be done with an inverse matrix.

$$\mathbf{A}\cancel{\mathbf{x}} = \lambda\cancel{\mathbf{x}} \quad ?$$

No.

Systems of Equations

Systems of Equations

$$\begin{cases} \quad\;\; 2x_2 + 3x_3 = 8 \\ 2x_1 + 3x_2 + 1x_3 = 5 \\ x_1 - \;\; x_2 - 2x_3 = -5 \end{cases}$$

$$\begin{pmatrix} 0 & 2 & 3 \\ 2 & 3 & 1 \\ 1 & -1 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 8 \\ 5 \\ -5 \end{pmatrix}$$

Systems of Equations

(Three types)

- In some applications, systems of equations have an **exact solution** - but this is rare.
- The system of equations may be a set of constraints (≤ , = , ≥). **Infinitely many solutions** within the constraints and must optimize some other quantity.
- In most applications, there is **no exact solution**. We introduce an error term and try to minimize it.

Systems of Equations

(Least Squares)

Obs	Weight	Width	Length	Time
1	3	5.4	6.3	10.11
2	1.1	1.2	2.1	4.25
3	2.4	3.4	5	8.09
4	1.9	2.8	8.1	7.20
5	3.2	6.1	4.5	9.90
6	2.7	3.7	4.6	7.75

Time $= \beta_0 + \beta_1\text{Weight} + \beta_2\text{Width} + \beta_3\text{Length}$

Systems of Equations

(Least Squares)

Intercept Weight Width Length Time

$$\beta_0 \begin{pmatrix}1\\1\\1\\1\\1\\1\end{pmatrix} + \beta_1 \begin{pmatrix}3\\1.1\\2.4\\1.9\\3.2\\2.7\end{pmatrix} + \beta_2 \begin{pmatrix}5.4\\1.2\\3.4\\2.8\\6.1\\3.7\end{pmatrix} + \beta_3 \begin{pmatrix}6.3\\2.1\\5\\8.1\\4.5\\4.6\end{pmatrix} \approx \begin{pmatrix}10.11\\4.25\\8.09\\7.20\\9.90\\7.75\end{pmatrix}$$

Time $\cancel{=} \beta_0 + \beta_1\text{Weight} + \beta_2\text{Width} + \beta_3\text{Length} + \varepsilon$

Systems of Equations

(Least Squares)

Intercept	Weight	Width	Length	Time
1	3	5.4	6.3	10.11
1	1.1	1.2	2.1	4.25
1	2.4	3.4	5	8.09
1	1.9	2.8	8.1	7.20
1	3.2	6.1	4.5	9.90
1	2.7	3.7	4.6	7.75

X (Intercept, Weight, Width, Length columns) **y** (Time column)

Time = $\beta_0 + \beta_1$Weight + β_2Width + β_3Length + ε

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \varepsilon$$

Systems of Equations

(Least Squares)

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}$$ (has no solutions. "inconsistent")

Want to find $\boldsymbol{\beta}$ that gets the modeled values ($\hat{\mathbf{y}}=\mathbf{X}\boldsymbol{\beta}$) on the left as close as possible to the true values ($\mathbf{y}$) on the right.

Minimize squared error

$$\min_{\beta} \sum_i \varepsilon_i^2$$

$$\varepsilon = y - \mathbf{X}\boldsymbol{\beta}$$

Systems of Equations

(Least Squares)

Minimize squared error

$$\min_{\beta} \sum_{i} \varepsilon_i^2$$

$$\varepsilon = y - \mathbf{X}\beta$$

or equivalently

$$\min_{\beta} (\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta)$$

or equivalently

$$\min_{\beta} \| (\mathbf{y} - \mathbf{X}\beta) \|_2^2$$

Systems of Equations

(Least Squares)

HOW to find the least squares solution?

The Normal Equations

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

As long as X is full rank (no perfect multicollinearity), $\mathbf{X}^T\mathbf{X}$ has an inverse and this system has an exact solution.

That solution **IS** the least squares solution.

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

We're DONE talking about regression in Linear Algebra class.

From now on, our focus is on *unsupervised* problems that do not have a target variable.

Norms, Distances, and Similarity

Norms

- **Norms** are functions that measure the *magnitude* or *length* of a vector.
- Written $||\mathbf{x}||$
- 2-Norm (Euclidean norm) is the most common.

$$\|\mathbf{x}\|_2 = \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2} = \sqrt{\mathbf{x}^T\mathbf{x}}$$

Norms

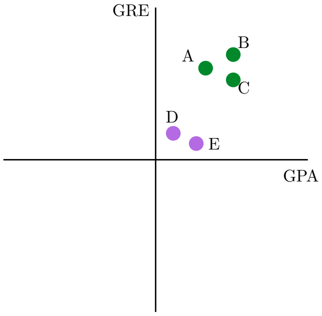
- The distance between two points, $\mathbf{x}$ and $\mathbf{y}$, is the norm of their difference.

$$\|\mathbf{x} - \mathbf{y}\|$$

- We can use this information to determine which points are more similar to each other.
- May create a **distance matrix**, $\mathbf{D}$, which contains pairwise distances between points (observations).

$$\mathbf{D}_{ij} = \|obs_i - obs_j\|$$

Distance Matrix



$$\mathbf{D} = \begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array} \begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{pmatrix} 0 & 1 & 1 & 3 & 4 \\ 1 & 0 & 1 & 5 & 5 \\ 1 & 1 & 0 & 4 & 3 \\ 3 & 5 & 4 & 0 & 1 \\ 4 & 5 & 3 & 1 & 0 \end{pmatrix} \end{array}$$

Other Norms

- 1-Norm (Manhattan/CityBlock/Taxicab distance)

$$\|\mathbf{x}\|_1 = |x_1| + |x_2| + \ldots + |x_n|$$

- **∞**-Norm (Max Distance)

$$\|\mathbf{x}\|_\infty = \max\{|x_1|, |x_2|, \ldots, |x_n|\}$$

Norms in Statistics

- Standard deviation:

$$s_x = \frac{1}{\sqrt{n-1}}\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

$$\frac{1}{\sqrt{n-1}}\|\mathbf{x}\|$$ ← vector of centered data

- Correlation Coefficient:

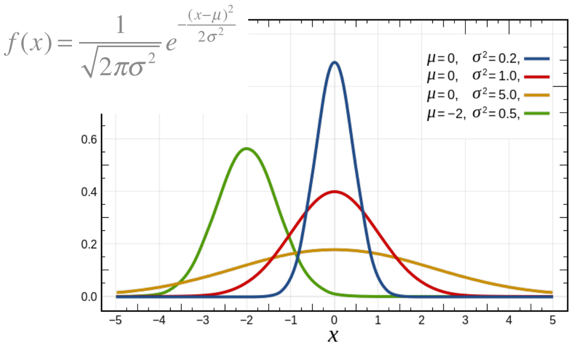
$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}}$$

$$\frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\|\,\|\mathbf{y}\|}$$ ← vectors of centered data

Covariance and Correlation

The Multivariate Normal Distribution

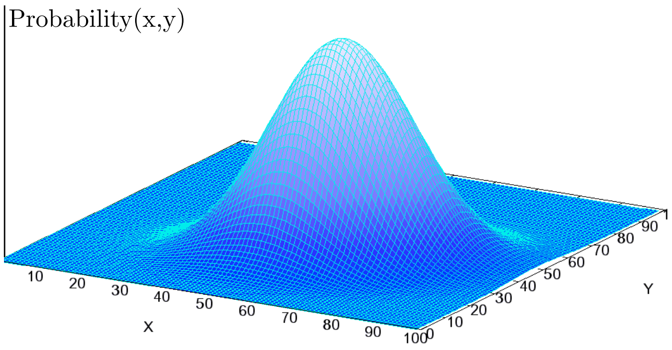
Normal (Gaussian) Density Function



Covariance

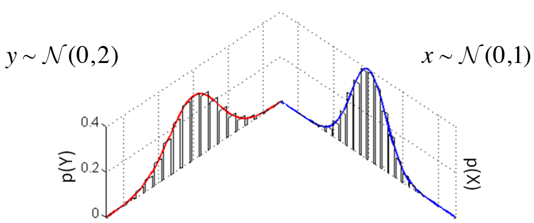
- Covariance is a number that describes how two variables change together.
- If x increases/decreases, does y tend to increase/ decrease? Covariance can be negative.
- Is a parameter of the **joint distribution** of x and y
  - joint distribution: how likely are we to see the pair (x,y) together?

Joint Distribution of (x,y)



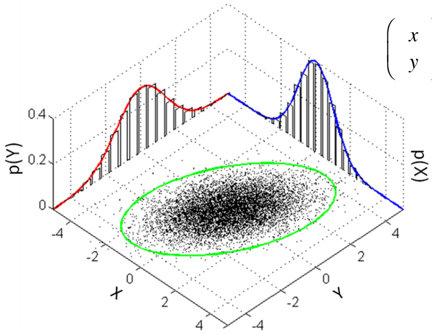
Multivariate Normal Distribution

Suppose x and y are normally distributed



Multivariate Normal Distribution

The vector (x,y) is multivariate normally distributed



$$\begin{pmatrix} x \\ y \end{pmatrix} \sim \mathcal{N}\left(\underbrace{\begin{pmatrix} 0 \\ 0 \end{pmatrix}}_{\substack{\mu \\ \text{mean} \\ \text{vector}}}, \underbrace{\begin{pmatrix} 1 & 3/5 \\ 3/5 & 2 \end{pmatrix}}_{\substack{\Sigma_{p\times p} \\ \text{covariance} \\ \text{matrix}}}\right)$$

Multivariate Normal Distribution

The vector (x,y) is multivariate normally distributed

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim \mathcal{N}\left(\underbrace{\begin{pmatrix} 0 \\ 0 \end{pmatrix}}_{\substack{\mu \\ \text{mean} \\ \text{vector}}}, \underbrace{\begin{pmatrix} 1 & 3/5 \\ 3/5 & 2 \end{pmatrix}}_{\substack{\Sigma_{p\times p} \\ \text{covariance} \\ \text{matrix}}}\right)$$

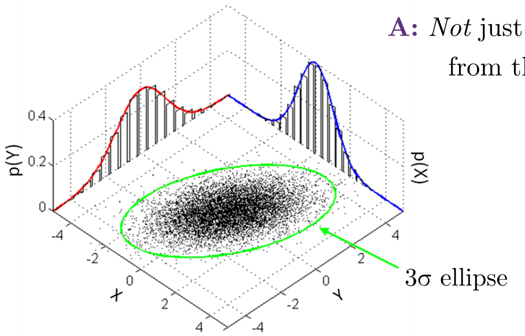
Covariance Matrix Fun Facts

- Variances of each variable on the main diagonal
- Covariances of each pair of variables on the off diagonal
- Always symmetric

Multivariate Normal Distribution

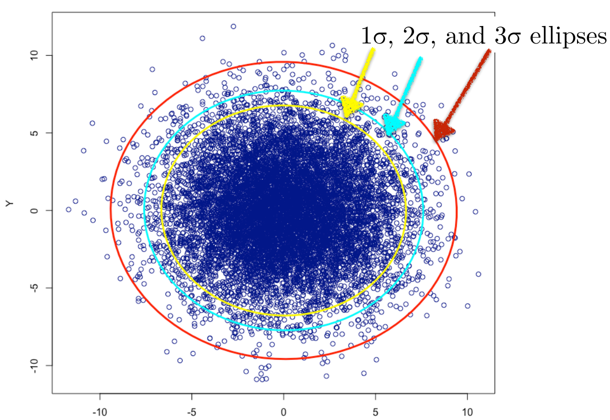
Q: How can we characterize a point as ***rare***?

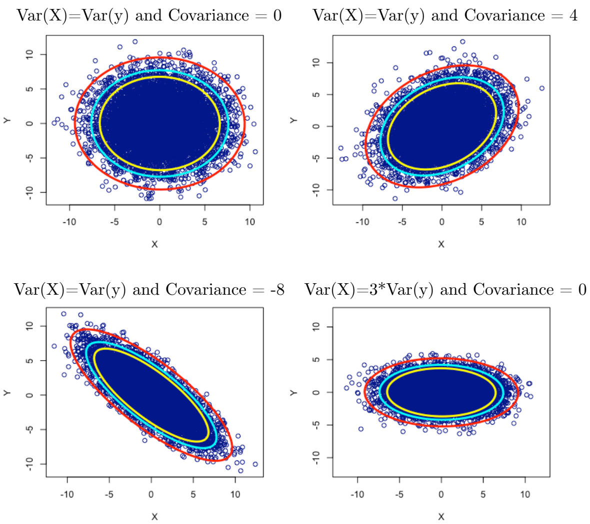
A: *Not* just distance from the mean!



Multivariate Normal Distribution

Var(X)=Var(y) and Covariance = 0





Covariance

Covariance is calculated from the data:

$$Cov(x,y) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = \frac{1}{n-1}\mathbf{x}^T\mathbf{y}$$

vectors of <u>centered</u> data

When covariance is positive:

x larger than mean, y tends to be larger than the mean
x smaller than the mean, y tends to be smaller than the mean

When covariance is negative:

x larger than mean, y tends to be smaller than the mean
x smaller than the mean, y tends to be larger than the mean

The units will have a strong effect on this number so we **cannot interpret magnitude**!!

Correlation

Correlation is the covariance of the standardized data:

$$Corr(x,y) = \frac{1}{n-1}\sum_{i=1}^{n}\frac{(x_i-\bar{x})}{s_x}\frac{(y_i-\bar{y})}{s_y} = \frac{1}{n-1}\mathbf{x}^T\mathbf{y}$$

vectors of standardized data

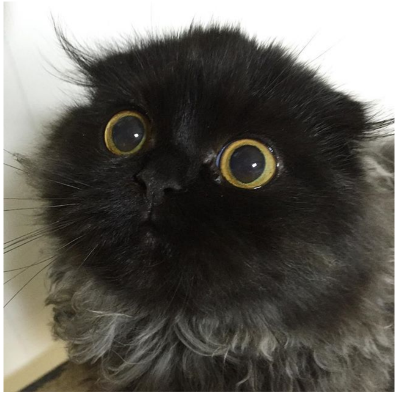
As we already know, correlation is between -1 and 1 and its magnitude measures the strength of a relationship.

NOT THE SLOPE

(Some) Conclusions

- Matrix arithmetic is a series of multiplications and additions done neatly in one operation
- Linear algebra will help us think about multivariate data geometrically
- Norms, Distances and Similarities between points in space will be the basis for many algorithms
- Covariance Matrix tells us about the shape of an elliptical probability distribution (think normal distribution) in space.
- Correlation is the covariance of the standardized data.

Feel like this in class today?



Primer Tutorials

(Prioritized)

http://www4.ncsu.edu/~slrace/LAprimer/index.html

- Tutorial 2 (Basic terminology) **12 minutes**
- Tutorials 3-4 (Matrix Arithmetic) **33 minutes**
- Tutorial 5 (Applications of Arithmethic) **17 minutes**
- Tutorial 13 (Basic Matrix Algebra) **12 minutes**
- Tutorial 15 (Norms&Distance Measures) **27 minutes**

101 minutes