STAT 151A: Lab 1

Stephanie DeGraaf

August 24, 2018

1 Logistics

Stephanie DeGraaf (sdegraaf@berkeley.edu)
Bryan Liu (runjing_liu@berkeley.edu)
Office Hours: 444 Evans Hall
    Stephanie: Tuesday 3-5pm, Wednesday 2-4pm
    Bryan: Monday 5-7pm, Wednesday 5-7pm
Lab Sections:
    Stephanie: Friday 9-11am, 2-4pm
    Bryan: Friday 11am-1pm, 4-6pm
Piazza: https://piazza.com/berkeley/fall2018/stat151a

2 Resources

The textbook for this class has a helpful appendix: http://socserv.socsci.mcmaster.ca/jfox/Books/Applied-Regression-3E/Appendices.pdf
Linear models with R, by Julian Faraway
Extending the linear model with R, by Julian Faraway
Plane Answers to Complex Questions, by Ronald Christensen:
https://link.springer.com/book/10.1007%2F978-1-4419-9816-3
Matrix Cookbook: everything you could ever want to know about matrices
https://www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf

3 Linear Algebra Review

"Any mathematical problem can be solved if it can be reduced to a problem in linear algebra."

In statistics and mathematics in general, linear algebra is at the core of (almost) everything. So if you understand linear algebra, you have the tools to be a great statistician. On the other hand, if your linear algebra knowledge is a little foggy, that's okay. This lab covers the basics, and there are plenty of resources available.

3.1 Vector Spaces

A **vector space** is a set of vectors that can be added together or multiplied by a scalar. The quintessential vector space is the $\mathbb{R}^n$, the set of all n-tuples of real numbers:

$$\{(x_1, ..., x_n) | x_i \in \mathbb{R}\}.$$

The two operations of a vector space, addition and multiplication, must also satisfy the following additional properties:

- additive identity (zero): $0 + v = v$
- additive inverse: $v + (-v) = 0$
- commutative: $v + w = w + v$
- associative: $u + (v + w) = (u + v) + w$
- multiplicative identity (unit): $1 \cdot v = v$
- compatibility: $a(bv) = (ab)v$
- vector distributive: $a(v + w) = av + aw$
- scalar distributive: $(a + b)v = av + bv$

A **subspace** is a subset U of a vector space V if U is itself a vector space.

3.1.1 Inner Products

An **inner product space** is a vector space with an additional structure, the **inner product**. Inner products generalize the concepts of distance and angles. The most familiar inner product space is Euclidean space, which is simply $\mathbb{R}^n$ equipped with an inner product: for vectors $x, y \in \mathbb{R}^n$, the inner product is

$$\langle x, y \rangle = x^T y = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

The **norm** of a vector is $||x|| = \sqrt{\langle x, x \rangle}$. This allows us to give a geometric definition of the inner product:

$$\langle x, y \rangle = ||x|| \, ||y|| \cos \theta,$$

where θ is the angle between x and y.
Two vectors are **orthogonal** if and only if $\langle x, y \rangle = 0$.
Related: Pythagorean theorem, triangle inequality, Cauchy-Schwartz inequality

3.2 Matrices

An $n \times p$ matrix A is a set of column vectors $x_1, ..., x_p \in \mathbb{R}^n$ written as

$$\left[x_1 \; x_2 \; \cdots \; x_p \right].$$

A **spanning set** is a set of vectors $\{v_1, v_2, ...\} \in V$ that allow us to write any vector $v \in V$ as a linear combination

$$v = a_1 v_1 + a_2 v_2 + \cdots$$

for unique scalars $a_1, a_2, ...$
Vectors are **linearly independent** if whenever

$$a_1 v_1 + a_2 v_2 + \cdots a_n v_n = 0,$$

then all a_i must equal 0.
A spanning set that is linearly independent is a **basis**. The **dimension** of a vector space is the number of elements in its basis. A basis is not unique, but the number of elements in any basis is the same.
The space spanned by the columns of A is the **column space** of A, denoted as

$$C(A) = \{v \mid v = a_1 x_1 + \cdots + a_p x_p; a_i \in \mathbb{R}\}.$$

Thus, the column space consists of all linear combinations of the column vectors of A.

3.3 Linear Transformations

Linear transformations are another way to formulate matrix operations. Any matrix defines a linear transformation, and every linear transformation can be represented by a matrix, although this form is NOT unique. A linear map simply takes a vector in one vector space to another vector in some other vector space, just a matrix does to a vector.
A **linear transformation** $T : V \to W$ is a function from a vector space V to a vector space W, such that for any real numbers a_1, a_2 and vectors $v_1, v_2 \in V$,

$$T(a_1 v_1 + a_2 v_2) = a_1 T(v_1) + a_2 T(v_2).$$

Examples: matrix multiplication from $\mathbb{R}^n$ to $\mathbb{R}^m$, differentiation, expectation
The **kernel** (or **nullspace**) of a linear transformation is the set of vectors that map to 0:

$$ker(T) = \{v \in V : T(v) = 0\}.$$

The **image** of a linear transformation is

$$Im(T) = \{w \in W : \text{there exists a } v \in V \text{ with } T(v) = w\}.$$

The kernel and image are subspaces of V and W, respectively (prove).

3.4 Projections

A special kind of linear transformation is a **projection**, which is a map T that has the idempotent property, that $T^2 = T$.
The projection of a vector v onto a subspace A is given by

$$Proj_A v = A(A^T A)^{-1} A^T.$$

Geometrically, $Proj_A v$ is the vector in A that is closest in distance to the original vector v.
A matrix M is an **orthogonal projection matrix** onto $C(X)$ if and only if
(i) $Mv = v$ for all $v \in C(X)$
(ii) $Mv = 0$ for all v where $v \perp C(X)$.
Important: $X(X^T X)^{-1} X^T$ is the perpendicular projection matrix onto $C(X)$.

3.5 Eigenvalues and Eigenvectors

An **eigenvector** of a matrix A is a non-zero vector v such that

$$Av = \lambda v.$$

The value λ is an **eigenvalue** of A. The intuition from a geometric point of view, is that applying A to the vector v only stretches or shrinks it by the scalar λ. The direction of v is not affected. The eigenvalue tells us how much A shrinks or stretches the eigenvector v.

3.6 The Invertible Matrix Theorem

The **Invertible Matrix Theorem** gives many equivalent ways of telling when a system of n linear equations and n unknowns has a solution. There are many equivalent conditions, and only a few are listed here.
For an $n \times n$ matrix A, the following are equivalent:

1. A is invertible.
2. $det(A) \neq 0$.
3. $ker(A) = 0$.
4. For a column vector $b \in \mathbb{R}^n$, there is a unique solution x to the equation $Ax = b$.
5. The columns (or rows) are linearly independent.

$\vdots$

3.6.1 How does this relate to statistics?

Linear regression deals with solving they system $Ax = b$. However, in statistics it is usually the case that the number of equations we want to satisfy is greater than the number of variables. That is, A is an $n \times k$ matrix where $n > k$. This is called an *overdetermined* or *inconsistent* system. Typically, there is no exact solution.
Instead of looking for an exact solution, we instead look for the vector x^* in the column space of A that is closest to the vector b. This is the *least-squares solution*.
In linear algebra terms, we want to find x^* such that $Ax^* = Proj_{C(A)}b$. Consider

$$Ax^* - b = Proj_{C(A)}b - b.$$

The right hand side lies in the orthogonal complement of $C(A)$. Thus we must have that

$$A^T(Proj_{C(A)}b - b) = 0;$$

that is, $(Proj_{C(A)}b - b)$ must be in the left null space, $N(A^T)$. This also means that $A^T(Ax^* - b) \in N(A^T)$. Solving $A^T(Ax^* - b) = 0$ yields the familiar least-squares solution

$$x^* = (A^TA)^{-1}A^Tb.$$