STAT 151A: Lab 1

Billy Fang

2 September 2017

1 Logistics

Billy Fang (blfang@berkeley.edu)
Office hours: Monday 9am-11am, Wednesday 10am-12pm, Evans 428 (room changes will be written on the chalkboard)
Piazza: piazza.com/berkeley/fall2017/stat151a

2 Reference

Whenever I refer to "Fox," I mean the third edition of the course textbook. If you have an earlier edition, some of the section numbering and page numbering may be different.

Some of the material below is from Christensen's *Plane Answers to Complex Questions*, which can be downloaded for free at Springer Link (https://link.springer.com/book/10.1007%2F978-1-4419-9816-3) provided you are connected to Berkeley internet.

3 Playing with R: graphics and `lm()`

See `STAT151A_lab01_demos.html` on bCourses.

4 Random vectors

The following material can also be found in Chapter 1 of Christensen.

Let $y_1, \dots, y_n$ be random variables. Then the vector

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

is a **random vector**.

We define the **expected value** of the random vector Y entry-wise, i.e.

$$\mathbb{E}[Y] = \mathbb{E}\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} := \begin{bmatrix} \mathbb{E}[y_1] \\ \vdots \\ \mathbb{E}[y_n]. \end{bmatrix}$$

So the expected value of a random vector is itself a random vector with the same dimension. We define the expected value of a random matrix similarly (entry-wise).

We define the **covariance matrix** of Y as the $n \times n$ matrix whose (i,j) entry is $\mathrm{Cov}(y_i, y_j)$.

$$\mathrm{Cov}(Y) := [\mathrm{Cov}(y_i, y_j)]_{i,j=1}^{n} = \begin{bmatrix} \mathrm{Cov}(y_1,y_1) & \mathrm{Cov}(y_1,y_2) & \cdots & \mathrm{Cov}(y_1,y_n) \\ \mathrm{Cov}(y_2,y_1) & \mathrm{Cov}(y_2,y_2) & \cdots & \mathrm{Cov}(y_2,y_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(y_n,y_1) & \mathrm{Cov}(y_n,y_2) & \cdots & \mathrm{Cov}(y_n,y_n) \end{bmatrix}.$$

Note that the diagonal entries of $\mathrm{Cov}(Y)$ are $\mathrm{Cov}(y_i, y_i) = \mathrm{Var}(y_i)$.

We also have the following neat expression for the covariance matrix.

$$\mathrm{Cov}(Y) = \mathbb{E}[(Y - \mathbb{E}[Y])(Y - \mathbb{E}[Y])^\top].$$

Why is this true? We just need to check that the right-hand side is an $n \times n$ matrix whose (i,j) entry is $\mathrm{Cov}(y_i, y_j)$. Remember that $\mathbb{E}[Y]$ is a vector, so $Y - \mathbb{E}[Y]$ is a vector as well. Then $(Y - \mathbb{E}[Y])(Y - \mathbb{E}[Y])^\top$ is an $n \times n$ matrix whose (i,j) entry is $(y_i - \mathbb{E}[y_i])(y_j - \mathbb{E}[y_j])$. Taking the expectation of this quantity yields the definition of $\mathrm{Cov}(y_i, y_j)$.

Recall that if y is a random variable (one dimension), then for any constants a and b, we have

$$\mathbb{E}[ay + b] = a\mathbb{E}[y] + b$$
$$\mathrm{Var}(ay + b) = a^2 \,\mathrm{Var}(y).$$

The generalization to random vectors is the following.

Proposition 4.1. Let Y be an n-dimensional random vector. Let A be a fixed $r \times n$ matrix, and b a fixed r-dimensional vector. Then

$$\mathbb{E}[AY + b] = A\mathbb{E}[Y] + b$$
$$\mathrm{Cov}(AY + b) = A\,\mathrm{Cov}(Y)A^\top.$$

Note the resemblance to the one-dimensional case Section 4.

Proof. Prove the first equality on your own. We now prove the second equality.

$$\begin{aligned} \mathrm{Cov}(AY + b) &= \mathbb{E}[(AY + b - \mathbb{E}[AY + b])(AY + b - \mathbb{E}[AY + b])^\top] \\ &= \mathbb{E}[(AY - A\mathbb{E}[Y])(AY - A\mathbb{E}[Y])^\top] \\ &= \mathbb{E}[A(Y - \mathbb{E}[Y])(Y - \mathbb{E}[Y])^\top A^\top] \\ &= A\mathbb{E}[(Y - \mathbb{E}[Y])(Y - \mathbb{E}[Y])^\top]A^\top \\ &= A\,\mathrm{Cov}(Y)A^\top. \end{aligned}$$

□

5 I forgot linear algebra

"You don't really grasp the material in class n until you take class $n + 1$."

Having a good grasp of linear algebra will be indispensable for this course. In lab we will try to do our best to review important concepts that are relevant to the class, but you should also make an effort to review on your own. I will assume you know some of the basics (e.g., how to multiply matrices), but if you need more review, consult your linear algebra textbook. The textbook for this class also has an appendix that reviews linear algebra basics that are relevant for the course, which you can download here: `http://socserv.socsci.mcmaster.ca/jfox/Books/Applied-Regression-3E/Appendices.pdf`.

5.1 Some preliminary facts

Review the axioms of a **vector space** on your own. For intuition, just think about $\mathbb{R}^n$, the space of n-dimensional vectors.

A **subspace** W of a vector space $\mathbb{R}^n$ is a subset of $\mathbb{R}^n$ that

- contains 0, and
- $v + cw \in W$, for any $v, w \in W$ and scalar c.

The **inner product** (a.k.a. dot product) between two vectors $v, w \in \mathbb{R}^n$ is

$$v^\top w = v_1 w_1 + v_2 w_2 + \cdots + v_n w_n.$$

In some contexts this is denoted by $\langle v, w \rangle$ instead. Recall also that

$$v^\top w = \|v\| \|w\| \cos(\theta), \tag{1}$$

where θ is the angle between v and w.

For an $m \times n$ matrix A, we let

$$C(A) := \{Av : v \in \mathbb{R}^n\}$$

denote the **column space** or **range** of A. Note that it is a subspace.

Let V be a vector space, e.g. the n-dimensional space $\mathbb{R}^n$.

If W is a subspace of V, we define the orthogonal complement as

$$W^\perp := \{v \in V : v^\top w = 0 \text{ for all } w \in W\}.$$

In other words, $W^\perp$ consists of all vectors of V that are orthogonal (perpendicular) to every vector in the subspace W.

Fact 5.1. Let $W \subseteq \mathbb{R}^n$ be a subspace. Every vector $v \in V$ can be written uniquely in the form

$$v = v_W + v_{W^\perp}, \tag{2}$$

where $v_W \in W$ and $v_{W^\perp} \in W^\perp$.

5.2 Projection matrices

Most of the material here is from Appendix B in Christensen.

There are many equivalent definitions of an orthogonal projection onto a subspace.

Definition 5.2. Let $W \subseteq \mathbb{R}^n$ be a subspace. Let M be an $n \times n$ matrix that satisfies

(i) $Mv = v$ for all $v \in W$, and

(ii) $Mv = 0$ for all $v \in W^\perp$.

We call M an **orthogonal projection onto** W.

To get an intuition for this particular definition, it is important to view it in conjunction with Fact 5.1. Take an arbitrary vector $v \in \mathbb{R}^n$ and a subspace W. Let M be an orthogonal projection onto W. By considering the unique decomposition (2) of v, we have

$$Mv = M(v_W + v_{W^\perp}) = Mv_W + Mv_{W^\perp} = v_W + 0 = v_W, \tag{3}$$

where we have used properties (i) and (ii) of M. Thus, this definition is saying that M maps a vector v to its "W-component" v_W in the decomposition in (2).

We have essentially shown that there is there is only one orthogonal projection onto a given subspace W. (So, we can say "*the* orthogonal projection onto W" instead of "*an* orthogonal projection onto W.")

Proposition 5.3. Let $W \subseteq \mathbb{R}^n$ be a subspace. There is only one matrix M that is an orthogonal projection onto W.

We can make one more easy observation.

Proposition 5.4. If M is the orthogonal projection onto W, then $W = C(M)$.

Proof. Condition (i) in Definition 5.2 shows that $W \subseteq C(M)$. Conversely, by (3), we have $Mv \in W$ for any v, so $C(M) \subseteq W$. □

Thus, if M is an orthogonal projection, the subspace onto which it is projecting is precisely its column space.

5.2.1 Alternate definition of orthogonal projection

We now state an equivalent definition for the orthogonal projection.

Theorem 5.5 (Alternate definition for orthogonal projection)**.** M is an orthogonal projection [onto some subspace] if and only if it satisfies

(a) $MM = M$ (idempotent), and

(b) $M^\top = M$ (symmetric).

Note that if M satisfies (a) and (b) in Theorem 5.5, the result does not explicitly say which subspace M is an orthogonal projection for. But Proposition 5.3 implies that it is $C(M)$, the column space of M.

In other words, one way to verify if a matrix M is the orthogonal projection onto a subspace W is to check

- $MM = M$,
- $M^\top = M$, and
- $C(M) = W$.

Proof (optional). We need to prove both directions: that Definition 5.2 implies (a) and (b), and vice versa.

Suppose Definition 5.2 holds for some subspace $W \subseteq \mathbb{R}^n$. Let u and v be arbitrary vectors. By looking at (2), we have $Mu = u_W \in W$. Moreover, we have $(I_n - M)v = v - v_W = v_{W^\perp} \in W^\perp$. Thus, Mu and $(I_n - M)v$ are orthogonal, i.e.

$$0 = (Mu)^\top((I_n - M)v) = u^\top M^\top (I_n - M)v.$$

Since this holds for any u and v, we must have $M^\top(I - M) = 0$. Rearranging yields $M^\top = M^\top M$. Since $M^\top M$ is symmetric, $M^\top$ must be symmetric, which yields (b). Then the previous equation can be rewritten as $M = MM$, yielding (a).

We now prove the reverse direction. Suppose M satisfies (a) and (b). We show the conditions in Definition 5.2 are satisfied with the subspace $W = C(M)$. If $v \in C(M)$, then $v = Mu$ for some vector u. Then by (a) we have $Mv = MMu = Mu = v$, yielding (i). On the other hand if $v \in C(M)^\perp$, then $Mv = M^\top v = 0$ because v is orthogonal to the columns of M (i.e. the rows of $M^\top$). □

5.2.2 Another alternate definition: explicit constructions of orthogonal projection matrices

So far we have not talked about how an orthogonal projection matrix actually looks like. You may remember from various courses (e.g., multivariable calculus, linear algebra, physics, etc.) the following interpretation of the inner product, based on the cosine formula (1).

Fact 5.6. Let u be a unit vector and v be a vector. Let θ be the angle between them. Then $u^\top v = \|u\|\|v\|\cos\theta = \|v\|\cos\theta$ is the length of the scalar projection onto the direction u. (See Figure 1.) Thus the orthogonal projection of v onto $\text{span}\{u\}$ is

$$u(u^\top v)$$

that is, the orthogonal projection matrix onto $\text{span}\{u\}$ is

$$uu^\top.$$

So, if W is one-dimensional, we know exactly what the orthogonal projection onto W is. For general subspaces W, the generalization is the following.

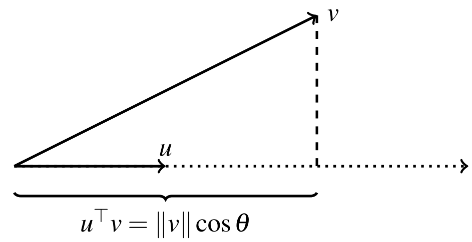

Figure 1: Scalar projection of v onto the span of u.

Theorem 5.7 (Explicit construction of orthogonal projection)**.** Let $W \subseteq \mathbb{R}^n$ be a subspace. Let $u_1, \dots, u_r$ be an orthonormal basis for W. Let U be the $n \times r$ matrix having columns $u_1, \dots, u_r$. Then the orthogonal projection onto $W = C(U)$ is

$$UU^\top = \sum_{i=1}^{r} u_i u_i^\top.$$

In other words, if $u_1, \dots, u_r$ is an orthonormal basis for W, then the orthogonal projection of v onto W is

$$(u_1^\top v)u_1 + (u_2^\top v)u_2 + \cdots + (u_r^\top v)u_n,$$

i.e. the component of v in direction u_i is simply $u_i^\top v$. Compare the expressions here and in Theorem 5.7 with those in Figure 1.

Exercise 5.8. *Prove Theorem 5.7. [Hint: show that $UU^\top$ satisfies Definition 5.2 with $W = C(X)$.]*

Let us return to the one-dimensional situation. What if we want to project a vector v onto the span of another vector x, but x is not a unit vector? Simply take the normalization $u = x/\|x\|$ and note that $\text{span}\{x\} = \text{span}\{u\}$, and apply Fact 5.1. Thus, the projection of v onto $\text{span}\{x\}$ is

$$(u^\top v)u = \left(\frac{x^\top v}{\|x\|}\right)\frac{x}{\|x\|} = \frac{xx^\top v}{x^\top x},$$

that is, the orthogonal projection matrix onto $\text{span}\{x\}$ is

$$\frac{xx^\top}{x^\top x}.$$

In general, we have the following.

Theorem 5.9. Let $W \subseteq \mathbb{R}^n$ be a subspace with basis $x_1, \dots, x_r$. Let X be the $n \times r$ matrix with columns $x_1, \dots, x_r$. Then the orthogonal projection onto $W = C(X)$ is

$$X(X^\top X)^{-1}X^\top.$$

This will appear a lot in our study of linear models. Note that Theorem 5.9 is in the "nice" case where X is full rank. This ensures that the $r \times r$ matrix $X^\top X$ is invertible, thanks to the following lemma.

Lemma 5.10. Let X be an $n \times r$ matrix with $r \leq n$.

$$\text{rank}(X^\top X) = \text{rank}(X).$$

Consequently, $X^\top X$ is invertible if X is full rank (i.e. $\text{rank}(X) = r$).

Exercise 5.11. *Prove Lemma 5.10. [Hint: by the rank-nullity theorem, it is equivalent to show that $X^\top X$ and X have the same kernel/nullspace.]*

Note also that if X has orthonormal columns, then $X^\top X = I_r$ and we recover the result of Theorem 5.7.

Exercise 5.12. *Prove Theorem 5.9. [Hint: Directly show that $X(X^\top X)^{-1}X^\top$ satisfies Definition 5.2 with $W = C(U)$.]*

5.2.3 What does this have to do with least squares?

Theorem 5.13 (Least squares characterization of the orthogonal projection)**.** Let M be the orthogonal projection onto a subspace $W \subseteq \mathbb{R}^n$. Choose any $v \in V$. Then Mv is the minimizer of the function

$$f(w) = \|w - v\|^2$$

over all $w \in W$.

This alternate definition of the orthogonal projection onto W has a clear geometric interpretation: Mv is the closest vector (in Euclidean distance) to v, among all vectors in W.

Proof. For any $w \in W$, we have

$$\begin{aligned}
\|w - v\|^2 &= \|w - Mv + Mv - v\|^2 \\
&= \|w - Mv\|^2 + 2\underbrace{(w - Mv)^\top (Mv - v)}_{=0} + \|Mv - v\|^2 \\
&= \|w - Mv\|^2 + \|Mv - v\|^2,
\end{aligned}$$

where the cross term is zero because the two vectors are orthogonal (explicitly, $Mv - v = v_W - (v_W + v_{W^\perp}) = v_{W^\perp} \in W^\perp$ and $w - Mv = w - v_W \in W$).

Thus if we want to minimize $\|w - v\|^2$ over all $w \in W$, it suffices to minimize $\|w - Mv\|^2$. It is then clear that $w = Mv$ is the minimizer. $\square$