STAT 151A: Lab 2

Billy Fang

8 September 2017

Feedback form is at the same place: https://goo.gl/forms/fKjLeKItix2Djg512. Please leave comments and suggestions for lab, office hours, etc.

1 Dummy variables

See STAT151A_lab02_demos.html on bCourses.

2 More linear algebra

2.1 Hints for Homework 1, Question 3

Note that the problem is not asking you to prove (a) and to prove (b). [One can come up with an example where $M_1M_2 \neq M_2M_1$ and $M_1M_2 \neq M_0$.] Rather, you are asked to show that (a) and (b) are equivalent. That is, you must show that (a) implies (b), and that (b) implies (a).

Use the properties from Lab 1. Proposition 2.2 may help too.

2.2 Intersections, sums, and complements, oh my!

Last time we defined the **orthogonal complement** of a subspace W.

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

That is, $W^\perp$ consists of all vectors that are orthogonal to every vector in W.

Exercise 2.1. *Show* $(W^\perp)^\perp = W$. ■

If W_1 and W_2 are subspaces, we define

$$W_1 + W_2 = \{w_1 + w_2 : w_1 \in W_1, w_2 \in W_2\}.$$

That is, $W_1 + W_2$ is the subspace that contains all vectors that can be written as the sum of a vector from W_1 with a vector from W_2.

You may remember **De Morgan's laws** from set theory. If A and B are sets, and A^c denotes the complement of A, then

$$(A \cap B)^c = A^c \cup B^c,$$
$$(A \cup B)^c = A^c \cap B^c.$$

It turns out that something similar holds for subspaces.

Proposition 2.2. If W_1 and W_2 are subspaces, then

$$(W_1 \cap W_2)^\perp = W_1^\perp + W_2^\perp \tag{1a}$$
$$(W_1 + W_2)^\perp = W_1^\perp \cap W_2^\perp. \tag{1b}$$

Proof. We first prove (1b). Suppose $v \in (W_1 + W_2)^\perp$. If $u \in W_1$, then $u = u + 0 \in W_1 + W_2$, so $v^\top u = 0$. Thus, $v \in W_1^\perp$. A similar argument yields $v \in W_2^\perp$. Thus $(W_1 + W_2)^\perp \subseteq W_1^\perp + W_2^\perp$.

We now prove the reverse inclusion $(W_1 + W_2)^\perp \supseteq W_1^\perp + W_2^\perp$. Suppose $v \in W_1^\perp \cap W_2^\perp$. Let $w_1 \in W_1$ and $w_2 \in W_2$ so that $w_1 + w_2 \in W_1 + W_2$. Then $v^\top (w_1 + w_2) = v^\top w_1 + v^\top w_2 = 0 + 0 = 0$ so $v \in (W_1^\perp + W_2^\perp)^\perp$.

We now show that (1a) is a consequence of (1b). By taking orthogonal complements on both sides of the (1a) and using Exercise 2.1, we see that it suffices to show

$$W_1 \cap W_2 = (W_1^\perp + W_2^\perp)^\perp.$$

This is simply (1b) applied to $W_1^\perp$ and $W_2^\perp$.

□

Side note: while set theory results sometimes have similar analogues for subspaces (like the above) that can help with memorization and intuition, one should still be careful. See this interesting example where the inclusion-exclusion principle for sets seems to generalize to an analogue for subspace dimensions, but fails as soon as you consider more than two subspaces!

2.3 Gram matrix fun

The **column space** (a.k.a. range space, image) of an $n \times p$ matrix X, denoted $C(X)$, is the subspace consisting of vectors in $\mathbb{R}^n$ of the form Xv for some $v \in \mathbb{R}^p$. Equivalently, it is the subspace spanned by the columns of X. The **rank** of X is defined to be $\operatorname{rank}(X) := \dim C(X)$.

Exercise 2.3. *For an $n \times p$ matrix X, prove that*

$$\operatorname{rank}(X) \leq \min\{n, p\}.$$

■

We say X has **full rank** if $\operatorname{rank}(X) = \min\{n, p\}$.

The **null space** (a.k.a. kernel) of an $n \times p$ matrix X, denoted $N(X)$, is the subspace of $\mathbb{R}^p$ of all $v \in \mathbb{R}^p$ such that $Xv = 0$. The **nullity** of X is defined to be $\operatorname{null}(X) := \dim N(X)$.

The column space and null space of a matrix are subspaces of different spaces! The column space is a subspace of $\mathbb{R}^n$ (the space that X maps *to*) while the null space is a subspace of $\mathbb{R}^p$ (the space that X maps *from*).

Exercise 2.4. *Prove $N(X^\top)^\perp = C(X)$.* ■

The **rank-nullity theorem** states that

$$\operatorname{rank}(X) + \operatorname{null}(X) = p,$$

where p is the number of columns in X.

The matrix $X^\top X$ is called the **Gram matrix** of X.

Lemma 2.5. For any $n \times p$ matrix X,

$$C(X^\top X) = C(X^\top), \tag{2a}$$

$$N(X^\top X) = N(X). \tag{2b}$$

Consequently,

$$\operatorname{rank}(X^\top) \overset{(i)}{=} \operatorname{rank}(X^\top X) \overset{(ii)}{=} \operatorname{rank}(X).$$

Proof. We first prove (2b). First note that if $v \in N(X)$ then $v \in N(X^\top X)$ because $X^\top X v = X^\top 0 = 0$. This proves $N(X^\top X) \supseteq N(X)$. For the reverse inclusion $N(X^\top X) \subseteq N(X)$, suppose $v \in N(X^\top X)$. Then $X^\top X v = 0$, and thus $0 = v^\top X^\top X v = \|Xv\|^2$ which implies $Xv = 0$.

Having proven (2b), we immediately have

$$\operatorname{null}(X^\top X) = \operatorname{null}(X).$$

Combining this with the rank-nullity theorem, we have which implies

$$\mathrm{rank}(X^\top X) + \mathrm{null}(X^\top X) = p = \mathrm{rank}(X) + \mathrm{null}(X),$$

which implies (ii).

We now prove (2a). In the previous version of these notes, I proved this from scratch. (See gray text below.) However, as Andrew pointed out, (2a) and (2b) are equivalent due to Exercise 2.4. Specifically, taking the orthogonal complement of both sides of (2b) yields (2a), and vice versa.

The inclusion $C(X^\top X) \subseteq C(X^\top)$ is immediate from the definition of column space. For the reverse inclusion $C(X^\top) \subseteq C(X^\top X)$, suppose $v \in C(X^\top)$. Then by definition, $v = X^\top u$ for some u. Now recall that u has a unique decomposition into $u = u' + u''$ where $u' \in N(X^\top)$ and $u'' \in N(X^\top)^\perp$. Note that Exercise 2.4 implies $u'' \in C(X)$. Thus $u'' = Xw$ for some w. Then we have

$$v = X^\top u = X^\top u' + X^\top u'' = 0 + X^\top u'' = X^\top X w,$$

which proves $v \in C(X^\top X)$.

Having proved (2a), we immediately have equality (i). □

Proposition 2.6. Let X be an $n \times p$ matrix. There always exists a solution β to the normal equation

$$X^\top X \beta = X^\top y.$$

The solution is unique if and only if $\mathrm{rank}(X) = p$.

Proof. The existence of a solution is an immediate consequence of (2a). The uniqueness of a solution is equivalent to the $p \times p$ matrix $X^\top X$ being invertible, which is equivalent to $\mathrm{rank}(X^\top X) = p$, which by Lemma 2.5 is equivalent to $\mathrm{rank}(X) = p$. □

Exercise 2.7. *What situations can cause* $\mathrm{rank}(X) < p$*?* ■