Advanced Methods Homework 1

Bohao Tang

September 5, 2017

1 Vector spaces and inner products

1. *Proof.* Denote $\{\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_k\}$ be a basis of W. And we can expand them to a basis of $\mathbb{R}^n$ as $\{\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_k, \mathbf{a}_{k+1}, \cdots, \mathbf{a}_n\}$. Then we do Gram-Schmidt orthogonalization to the basis and get $\{\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_k, \mathbf{b}_{k+1}, \cdots, \mathbf{b}_n\}$. It's easy to see that $\{\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_k\}$ become an othogonal basis of W (they are othogonal, hence linear independent, and k is the dimension of W). Now we assert that $V = \mathbf{span}\{\mathbf{b}_{k+1}, \cdots, \mathbf{b}_n\}$ will be $W^\perp$.

   First, for every vector $\mathbf{x}$ in $W^\perp$, it has a representation $\mathbf{x} = \sum_{i=1}^n \omega_i \mathbf{b}_i$ for some $\omega_i \in \mathbb{R}^n$. Since $\mathbf{x} \in W^\perp$, we have $\langle \mathbf{x}, \mathbf{b}_j \rangle = 0$, for all $j = 1, 2, \cdots, k$. Therefore $\omega_j = 0$ for $1 \le j \le k$, which means $\mathbf{x} \in V$ and $W^\perp \subset V$.

   In the other side, for every vector $\mathbf{y} = \sum_{t=k+1}^n \gamma_t \mathbf{b}_t \in V$, it's quite direct to see that $\langle \mathbf{y}, \mathbf{b}_j \rangle = 0$, for all $j = 1, 2, \cdots, k$, hence $\langle \mathbf{y}, \mathbf{w} \rangle = 0$, for all $\mathbf{w} \in W$. Therefore $\mathbf{y} \in W^\perp$ and $W^\perp \subset V$.

   Now we get $W^\perp = V$, hence $\dim(W^\perp) = n - k$. Also we can see that $W^\perp$ is unique. If not, we can merge the two different $W^\perp$ and get a higher dimensional subspace Z that is orthogonal to W. Then $Z \oplus W \subset \mathbb{R}^n$ but $\dim(Z \oplus W) \ge n + 1 > n$, a contradiction. □

2. (a) *Proof.* According to Cauchy inequality, $\left(\sum u_i v_i\right)^2 \le \sum u_i^2 \sum v_i^2$. It's equivalent here that $|\langle \mathbf{u}, \mathbf{v} \rangle| \le ||\mathbf{u}||||\mathbf{v}||$. □

   (b) *Proof.*

$$
\begin{aligned}
||\mathbf{u} + \mathbf{v}||^2 + ||\mathbf{u} - \mathbf{v}||^2 &= (\mathbf{u} + \mathbf{v})'(\mathbf{u} + \mathbf{v}) + (\mathbf{u} - \mathbf{v})'(\mathbf{u} - \mathbf{v}) && (1) \\
&= \mathbf{u}'\mathbf{u} + 2\mathbf{u}'\mathbf{v} + \mathbf{v}'\mathbf{v} + \mathbf{u}'\mathbf{u} - 2\mathbf{u}'\mathbf{v} + \mathbf{v}'\mathbf{v} && (2) \\
&= 2\mathbf{u}'\mathbf{u} + 2\mathbf{v}'\mathbf{v} = 2||\mathbf{u}||^2 + 2||\mathbf{v}||^2 && (3)
\end{aligned}
$$

   □

3. In the definition of projection in Lecture 2, **y - Π(y—x)** should be orthogonal to **x**. So there is nothing need to be proved. Maybe we should changed the definition here that (suppose $\mathbf{x} \neq \mathbf{0}$, or there is no meaningful projection)

$$
\Pi(\mathbf{y}\text{—}\mathbf{x}) \triangleq \underset{\mathbf{u} \in \mathrm{span}\{\mathbf{x}\}}{\mathrm{argmin}} \; ||\mathbf{y} \text{ - } \mathbf{u}||^2
$$

   Proof. (This proof contains the part to assert that projection above is well defined) Suppose $\mathbf{u} = b\mathbf{x}$, then $||\mathbf{y} - \mathbf{u}||^2 = \mathbf{y}'\mathbf{y} - 2b\mathbf{x}'\mathbf{y} + b^2\mathbf{x}'\mathbf{x}$. It's a quadratic function with highest coefficient $\mathbf{x}'\mathbf{x} > 0$, therefore have a unique minimizer. Differentiate it with b, let it be zero and we get the normal equation $\mathbf{x}'\mathbf{y} = b\mathbf{x}'\mathbf{x}$, which means (since $\Pi(\mathbf{y}—\mathbf{x}) = b\mathbf{x}$) : $\langle \Pi(\mathbf{y}—\mathbf{x}), x \rangle = \langle \mathbf{y}, x \rangle$ and therefore $\mathbf{y} - \Pi(\mathbf{y}—\mathbf{x}) \perp \mathbf{x}$. □

2 Regression

1. Slope $\hat{\beta}_{yx}$ of regressing $\mathbf{y}$ on $\mathbf{x}$ is $\hat{\rho}_{xy}\,\frac{\hat{\sigma}_y}{\hat{\sigma}_x}$, where $\hat{\rho}_{xy}$ is the sample correlation coefficient and $\hat{\sigma}_y, \hat{\sigma}_x$ are the sample standard deviation of y and x. So, slope $\hat{\beta}_{xy}$ of $\mathbf{x}$ on $\mathbf{y}$ is $\hat{\rho}_{yx}\,\frac{\hat{\sigma}_x}{\hat{\sigma}_y}$. We have that $\hat{\rho}_{xy} = \hat{\rho}_{yx}$, so $\hat{\beta}_{xy}\hat{\beta}_{yx} = \hat{\rho}^2_{xy}$.

2. *Proof.* In the setting of mean only regression of $\mathbf{y}$, we have $\hat{\mu} = \overline{\mathbf{y}} = \sum_{i=1}^n y_i/n$. And the residue is $\mathbf{r} = \mathbf{y} - \overline{\mathbf{y}}\mathbf{J}_n$, therefore sum of residue is $\sum r_i = \sum_1^n y_i - n \cdot \sum_1^n y_i/n = 0$. □

3. *Proof.* In the setting of mean only regression of $\mathbf{y}$ on $\mathbf{x}$, we have $\hat{\beta} = \frac{\langle \mathbf{y},\mathbf{x} \rangle}{\langle \mathbf{x},\mathbf{x} \rangle}$. Therefore residue is $\mathbf{r} = \mathbf{y} - \frac{\langle \mathbf{y},\mathbf{x} \rangle}{\langle \mathbf{x},\mathbf{x} \rangle}\mathbf{x}$, therefore $\langle \mathbf{r}, \mathbf{x} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle - \frac{\langle \mathbf{y},\mathbf{x} \rangle}{\langle \mathbf{x},\mathbf{x} \rangle} \cdot \langle \mathbf{x}, \mathbf{x} \rangle = 0$, $\mathbf{r} \perp \mathbf{x}$. □

4. Notation as above, when $\mathbf{y} \perp \mathbf{x}$, we find that the residue $\mathbf{r} = \mathbf{y}$, which of course need not sum to zero.

5. *Proof.* In this setting of regressing $\mathbf{y}$ on $\mathbf{x}$, we get the normal equation that:

$$\mathbf{J}_n'\mathbf{y} = \hat{\beta}_0\mathbf{J}_n'\mathbf{J}_n + \hat{\beta}_1\mathbf{J}_n'\mathbf{x} \tag{4}$$
$$\mathbf{x}'\mathbf{y} = \hat{\beta}_0\mathbf{x}'\mathbf{J}_n + \hat{\beta}_1\mathbf{x}'\mathbf{x} \tag{5}$$

and also the residue $\mathbf{r} = \mathbf{y} - \hat{\beta}_0\mathbf{J}_n - \hat{\beta}_1\mathbf{x}$. Then:

$$\langle \mathbf{r}, \mathbf{J}_n \rangle = \mathbf{J}_n'\mathbf{y} - \hat{\beta}_0\mathbf{J}_n'\mathbf{J}_n + \hat{\beta}_1\mathbf{J}_n'\mathbf{x} = 0$$

and

$$\langle \mathbf{r}, \mathbf{x} \rangle = \mathbf{x}'\mathbf{y} - \hat{\beta}_0\mathbf{x}'\mathbf{J}_n + \hat{\beta}_1\mathbf{x}'\mathbf{x} = 0$$

So we get the residue $\mathbf{r} \perp \mathbf{J}_n$ and $\mathbf{r} \perp \mathbf{x}$. □

3 Least squares

1. *Proof.* Given $\mathbf{H}^2 = \mathbf{H}$, then $(\mathbf{I - H})^2 = \mathbf{I}^2 - 2\mathbf{H} + \mathbf{H}^2 = \mathbf{I} - 2\mathbf{H} + \mathbf{H} = \mathbf{I - H}$. □

2.

3. *Proof.* Suppose the singular value decomposition of $\mathbf{X}$ is $\mathbf{X} = \mathbf{USV}$, where □

4.