Lab Solutions 05

1. I showed you how to get a nice equation for β in the ordinary least squares equation. Using the SVD of X, compute a compact formula for the values $\widehat{y} = X\beta$.

This comes from just plugging in the SVD decomposition into the equation:

$$\begin{aligned}
\widehat{y} = X\beta &= X(VD^{-1}U^t)y && (5.1)\\
&= (UDV^t)(VD^{-1}U^t)y && (5.2)\\
&= UDV^tVD^{-1}U^ty && (5.3)\\
&= UU^ty. && (5.4)
\end{aligned}$$

Note that the term UU^t does **not** cancel.

2. We glossed over the case where one or more of the singular values is equal to zero. In this question I will show you why we cannot deal with this case in the construction of β. Let V_p denote the last column of V (these columns are called the *right singular vectors*). Argue that:

$$V^tV_p = \begin{bmatrix} 0 \\ 0 \\ \cdots \\ 0 \\ 1 \end{bmatrix} \tag{5.5}$$

Now, assume that $\sigma_p = 0$. Show that (Hint: expand X with the SVD):

$$XV_p = 0. \tag{5.6}$$

The first assertion comes because $V_k^tV_p$ is zero if $k \neq p$ and 1 otherwise. That's why the last term is one ($k = p$) and the others are zero. Expanding X with the SVD, we have:

$$XV_p = UDV^tV_p = UD \begin{bmatrix} 0 \\ 0 \\ \cdots \\ 0 \\ 1 \end{bmatrix} = U \cdot \begin{bmatrix} 0 \\ 0 \\ \cdots \\ 0 \\ \sigma_p \end{bmatrix}. \tag{5.7}$$

Because $\sigma_p = 0$, this give U times a vector of zeros, from which we get that $XV_p = 0$.

3. Assume that we have a potential candidate β for the regression vector. Show that the fitted values $\widehat{y}$:

$$\widehat{y} = X\beta = X(\beta + a \cdot V_p), \quad \forall a \in \mathbb{R}. \tag{5.8}$$

Explain why this implies that we cannot uniquely determine a value for β according the minimization of the loss function on the training data when $\sigma_1 = 0$.

The equation follows with almost no work the last question because:

$$
\begin{aligned}
X(\beta + a \cdot V_p) &= X\beta + aXV_p && (5.9)\\
&= X\beta + 0 = X\beta. && (5.10)
\end{aligned}
$$

Therefore the predictions $\widehat{y}$ are not changed if we add a multiple of the last singular vector V_p to β. Therefore, there is no unique best β under the loss function (we can always add aV_p and get the same results).

4. Let X be a matrix with SVD equal to UDV^t and w be a p-dimensional vector with Euclidean norm equal to one:

$$||w||_2^2 = w^t w = \sum_k w_k^2 = 1. \tag{5.11}$$

It is generally true that we can write the vector w as a weighted sum of the columns of V:

$$w = \sum_k a_k \cdot V_k. \tag{5.12}$$

I want you to show that $\sum_k a_k^2 = 1$. This is straightforward assuming that you approach the problem in a particular way. Start by writing out $||w||_2^2$ as an inner product and expanding in the basis of V:

$$
\begin{aligned}
1 = ||w||_2^2 = w^t w &= \left(\sum_k a_k \cdot V_k\right)^t \left(\sum_k a_k \cdot V_k\right) && (5.13)\\
&= \left(\sum_k a_k \cdot V_k^t\right) \left(\sum_k a_k \cdot V_k\right) && (5.14)
\end{aligned}
$$

Then, take the cross terms to write this as a double sum and simplify the result.

From the equation in the question we have:

$$
\begin{aligned}
1 = ||w||_2^2 = w^t w &= \left(\sum_k a_k \cdot V_k\right)^t \left(\sum_k a_k \cdot V_k\right) && (5.15)\\
&= \left(\sum_k a_k \cdot V_k^t\right) \left(\sum_k a_k \cdot V_k\right) && (5.16)\\
&= \sum_j \sum_k a_k a_j V_k^t V_j && (5.17)
\end{aligned}
$$

But this sum is zero if $k \neq j$ and is 1 is $k = j$, so:

$$\sum_j \sum_k a_k a_j V_k^t V_j = \sum_k a_k a_k V_k^t V_k \tag{5.18}$$

$$= \sum_k a_k^2 \tag{5.19}$$

And that is exactly what we wanted to show.

5. Let X be a matrix with SVD equal to UDV^t and w be a p-dimensional vector. Show that:

$$||Xw||_2^2 = ||DV^t w||_2^2 \tag{5.20}$$

In other words, the matrix U does not effect the size of the product Xw.

By expanding the Euclidean norm, this follows quickly as:

$$\begin{aligned} ||Xw||_2^2 &= w^t X^t X w &(5.21)\\ &= w^t (UDV^t)^t (UDV) w &(5.22)\\ &= w^t V D^t U^t U D V^t w &(5.23)\\ &= w^t V D D V^t w &(5.24)\\ &= (DV^t w)^t (DV^t w) &(5.25)\\ &= ||DV^t w||_2^2. &(5.26) \end{aligned}$$

6. Let X be a matrix with SVD equal to UDV^t and w be a p-dimensional vector that we will write as:

$$w = \sum_k a_k \cdot V_k. \tag{5.27}$$

Show that:

$$V^t w = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}. \tag{5.28}$$

We can start by showing that:

$$V^t w = V^t(\sum_k a_k \cdot V_k) = \sum_k a_k V V_k \tag{5.29}$$

$$= \sum_k a_k \cdot \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}. \tag{5.30}$$

Then, the second value directly follows by the definition of the matrix product:

$$DV^t w = \begin{bmatrix} \sigma_1 \cdot a_1 \\ \sigma_2 \cdot a_2 \\ \vdots \\ \sigma_p \cdot a_p \end{bmatrix} \tag{5.31}$$

7. Using the set-up from the previous question, show that:

$$||Xw||_2^2 = \sum_k a_k^2 \sigma_k^2. \tag{5.32}$$

From the the previous two questions, we simply have:

$$||Xw||_2^2 = ||DV^t w||_2^2 \tag{5.33}$$

$$= || \begin{bmatrix} \sigma_1 \cdot a_1 \\ \sigma_2 \cdot a_2 \\ \cdots \\ \sigma_p \cdot a_p \end{bmatrix} ||_2^2 \tag{5.34}$$

$$= \sum_k a_k^2 \sigma_k^2. \tag{5.35}$$

8. Let X be a matrix with SVD equal to UDV^t. Consider the ℓ_2-ball given by all vectors with a Euclidean norm of 1:

$$B_p = \{v \in \mathbb{R}^p, \quad \text{s.t.} \, ||v||_2 = 1\}. \tag{5.36}$$

Argue that:

$$\min_{v \in B_p} \{||Xv||_2^2\} = \sigma_p^2 \tag{5.37}$$

And

$$\max_{v \in B_p} \left\{ ||Xv||_2^2 \right\} = \sigma_1^2. \tag{5.38}$$

Putting together the previous questions, we can write any $v \in \mathbb{R}^p$ as $\sum_k a_k V_k$ with $\sum_k a_k^2 = 1$. Also, then, $||Xv||_2^2 = \sum a_k^2 \sigma_k^2$. This gives:

$$\min_a \sum_k \sigma_k^2 \cdot a_k^2, \quad \text{s.t.} \sum_k a_k^2 = 1. \tag{5.39}$$

If the sum of the squared values is constrained to be one and we want to minimize a weighted sum of the singular values, this done by putting all of the weight on the smallest singular value. In other words, $v = V_p$. The exact same argument with the maximum yields the second result.

9. Finally, complete the questions in the file `class05.Rmd`.