

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 $\hat{y} = X\beta$.

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

$$\hat{y} = X\beta = X(VD^{-1}U^t)y \quad (5.1)$$

$$= (UDV^t)(VD^{-1}U^t)y \quad (5.2)$$

$$= UDV^tVD^{-1}U^ty \quad (5.3)$$

$$= UU^ty. \quad (5.4)$$

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:

$$V^tV_p = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \quad (5.5)$$

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

$$XV_p = 0. \quad (5.6)$$

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:

$$XV_p = UDV^tV_p = UD \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} = U \cdot \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \sigma_p \end{bmatrix}. \quad (5.7)$$

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 \hat{y} :

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

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:

$$X(\beta + a \cdot V_p) = X\beta + aXV_p \quad (5.9)$$

$$= X\beta + 0 = X\beta. \quad (5.10)$$

Therefore the predictions \hat{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:

$$\|w\|_2^2 = w^t w = \sum_k w_k^2 = 1. \quad (5.11)$$

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

$$w = \sum_k a_k \cdot V_k. \quad (5.12)$$

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 :

$$1 = \|w\|_2^2 = w^t w = \left(\sum_k a_k \cdot V_k \right)^t \left(\sum_k a_k \cdot V_k \right) \quad (5.13)$$

$$= \left(\sum_k a_k \cdot V_k^t \right) \left(\sum_k a_k \cdot V_k \right) \quad (5.14)$$

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

As a general rule, we can apply the following formula for the product of two sums:

$$\left(\sum_k c_k \right) \times \left(\sum_j d_j \right) = \sum_k \sum_j (c_k d_j). \quad (5.15)$$

Using this, from the equation in the question we have:

$$1 = \|w\|_2^2 = w^t w = \left(\sum_k a_k \cdot V_k \right)^t \left(\sum_k a_k \cdot V_k \right) \quad (5.16)$$

$$= \left(\sum_k a_k \cdot V_k^t \right) \left(\sum_k a_k \cdot V_k \right) \quad (5.17)$$

$$= \left(\sum_k a_k \cdot V_k^t \right) \left(\sum_j a_j \cdot V_j \right) \quad (5.18)$$

$$= \sum_j \sum_k a_k a_j V_k^t V_j \quad (5.19)$$

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

$$\sum_j \sum_k a_k a_j V_k^t V_j = \sum_k a_k a_k V_k^t V_k \quad (5.20)$$

$$= \sum_k a_k^2 \quad (5.21)$$

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:

$$\|Xw\|_2^2 = \|DV^t w\|_2^2 \quad (5.22)$$

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

By expanding the Euclidean norm, this follows quickly as:

$$\|Xw\|_2^2 = w^t X^t X w \quad (5.23)$$

$$= w^t (UDV^t)^t (UDV) w \quad (5.24)$$

$$= w^t V D^t U^t U D V^t w \quad (5.25)$$

$$= w^t V D D V^t w \quad (5.26)$$

$$= (DV^t w)^t (DV^t w) \quad (5.27)$$

$$= \|DV^t w\|_2^2. \quad (5.28)$$

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

$$w = \sum_k a_k \cdot V_k. \quad (5.29)$$

Show that:

$$V^t w = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}. \quad (5.30)$$

We can start by showing that:

$$V^t w = V^t \left(\sum_k a_k \cdot V_k \right) = \sum_k a_k V^t V_k \quad (5.31)$$

$$= \sum_k a_k \cdot \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}. \quad (5.32)$$

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

$$DV^t w = \begin{bmatrix} \sigma_1 \cdot a_1 \\ \sigma_2 \cdot a_2 \\ \vdots \\ \sigma_p \cdot a_p \end{bmatrix} \quad (5.33)$$

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

$$\|Xw\|_2^2 = \sum_k a_k^2 \sigma_k^2. \quad (5.34)$$

From the the previous two questions, we simply have:

$$\|Xw\|_2^2 = \|DV^t w\|_2^2 \quad (5.35)$$

$$= \left\| \begin{bmatrix} \sigma_1 \cdot a_1 \\ \sigma_2 \cdot a_2 \\ \vdots \\ \sigma_p \cdot a_p \end{bmatrix} \right\|_2^2 \quad (5.36)$$

$$= \sum_k a_k^2 \sigma_k^2. \quad (5.37)$$

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:

$$B_p = \{v \in \mathbb{R}^p, \quad \text{s.t. } \|v\|_2 = 1\}. \quad (5.38)$$

Argue that:

$$\min_{v \in B_p} \{\|Xv\|_2^2\} = \sigma_p^2 \quad (5.39)$$

And

$$\max_{v \in B_p} \{\|Xv\|_2^2\} = \sigma_1^2. \quad (5.40)$$

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:

$$\min_a \sum_k \sigma_k^2 \cdot a_k^2, \quad \text{s.t. } \sum_k a_k^2 = 1. \quad (5.41)$$

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`.