Notes

STA314H1 - Fall 2020

Ziyue Yang

October 30, 2020

Contents

1 Week 5

1.1 The Singular Value Decomposition

Remark 1.1. We know that $\|X\|_2$ is the square root of the largest eigenvalue of $X^T X$, hence at some point , whenever we see one of the inner product matrices, we should recall the **PCA**.

The SVD is a good way to understand exactly how PCA is working in terms of the feature matrix X.

Theorem 1.1 (The Singular Value Decomposition)**.** Assume that $p < n$.

Let X be an $n \times p$ matrix. Then there exists an orthogonal $p \times p$ matrix V (i.e. $V^T V = VV^T = I$) and an orthogonal $n \times n$ matrix U such that

$$U^T X V = D, \tag{1.1}$$

where $D = \operatorname{diag}(\sigma_1, \ldots, \sigma_p)$, and $\sigma_1 \geq \cdots \geq \sigma_p \geq 0$.

Proof. Omitted for now. □

Remark 1.2. We can express the SVD of an $n \times p$ matrix X in several equivalent ways:

1. As a singular tuple (σ, u, v) that satisfies $Xv = \sigma u$ and $X^T u = \sigma v$.
2. As a matrix decomposition $X = UDV^T$, where V is a $p \times p$ orthogonal matrix, and U is a $n \times n$ orthogonal matrix.
3. As a way of representing the matrix as a sum

$$X = \sum_{j=1}^{p} \sigma_j u_j v_j^T. \tag{1.2}$$

Remark 1.3 (The SVD and Principal Components)**.** Recall that the factor loadings are the eigenvectors of $X^T X$.

If $X = UDV^T$, then $X^T X = V^T D U^T U D V^T = V D^2 V^T$.

- The V in the SVD is exactly the matrix of factor loadings.
- The eigenvalues of $X^T X$ are the squares of the singular values.

Note that the score vectors were defined as $t_j = Xv_j$, and We can use one of the representations of singular vectors to see that

$$t_j = Xv_j = \sigma_j u_j. \tag{1.3}$$

Remark 1.4. The SVD makes it easy to solve the normal equations.

Recall that

$$\hat{\beta} = (X^T X)^{-1} X^T y \tag{1.4}$$
$$= V D^{-2} V^T V D U^T y \tag{1.5}$$
$$= V D^{-1} U^T y \tag{1.6}$$
$$= \sum_{j=1}^{p} \frac{u_j^T y}{\sigma_j} v_j. \tag{1.7}$$

PCR just snipes off the small eigenvectors:

$$\hat{\beta}_{\text{pcr}} = V_k D_k^{-2} V_k^T V D U^T = \sum_{j=1}^{k} \frac{u_j^T y}{\sigma_j} v_j. \tag{1.8}$$

1.2 Ridge Regression

What does overfitting look like?

Skipped, **TODO**.

2 Week 7

2.1 Introduction to Lasso

Remark 2.1. Ridge regression stabilizes the least-squares estimates by shrinking low-variance directions, which makes it like a *softer* version of **principal component regression**.

> Can we use penalized regression to make a softer version of variable selection? Yes. But we need to use a different penalty.