Discussion 5C

The following notes are useful for this discussion: Note 14.

1. Orthonormality and Least Squares

Recall that, if $U \in \mathbb{R}^{m\times n}$ is a tall matrix (i.e. $m \geq n$) with orthonormal columns, then

$$U^\top U = I_{n\times n} \tag{1}$$

However, it is not necessarily true that $UU^\top = I_{m\times m}$. In this discussion, we will deal with "orthonormal" matrices, where the term "orthonormal" refers to a matrix that is square with orthonormal columns and rows. Furthermore, for an orthonormal matrix U,

$$U^\top U = UU^\top = I_{n\times n} \implies U^{-1} = U^\top \tag{2}$$

This discussion will cover some useful properties that make orthonormal matrices favorable, and we will see a "nice" matrix factorization that leverages orthonormal matrices and helps us speed up least squares.

(a) Suppose you have a real, square, $n \times n$ orthonormal matrix U. You also have real vectors $\vec{x}_1$, $\vec{x}_2$, $\vec{y}_1$, $\vec{y}_2$ such that

$$\vec{y}_1 = U\vec{x}_1 \tag{3}$$
$$\vec{y}_2 = U\vec{x}_2 \tag{4}$$

This is analogous to a change of basis. Show that, in this new basis, the inner products are preserved. **Calculate $\langle \vec{y}_1, \vec{y}_2 \rangle = \vec{y}_2^\top \vec{y}_1 = \vec{y}_1^\top \vec{y}_2$ in terms of $\langle \vec{x}_1, \vec{x}_2 \rangle = \vec{x}_2^\top \vec{x}_1 = \vec{x}_1^\top \vec{x}_2$.**

(b) Using the change of basis defined in part 1.a, show that, in the new basis, the norms are preserved. **Express $\|\vec{y}_1\|^2$ and $\|\vec{y}_2\|^2$ in terms of $\|\vec{x}_1\|^2$ and $\|\vec{x}_2\|^2$.**

(c) Suppose you observe data coming from the model $y_i = \vec{a}^\top \vec{x}_i$, and you want to find the linear scale-parameters (each a_i). We are trying to learn the model $\vec{a}$. You have m data points $(\vec{x}_i, y_i)$, with each $\vec{x}_i \in \mathbb{R}^n$. Each $\vec{x}_i$ is a different input vector that you take the inner product of with $\vec{a}$, giving a scalar y_i.

Set up a matrix-vector equation of the form $X\vec{a} = \vec{y}$ for some X and $\vec{y}$, and propose a way to estimate $\vec{a}$.

(d) Let's suppose that we can write our X matrix from part 1.c as

$$X = MV^\top \tag{5}$$

for some matrix $M \in \mathbb{R}^{m\times n}$ and some orthonormal matrix $V \in \mathbb{R}^{n\times n}$. **Find an expression for $\widehat{\vec{a}}$ from the previous part, in terms of M and $V^\top$.**

Note: take this form as a given. We will go over how to find such a V and M later.

(e) Now suppose that we have the matrix

$$\begin{bmatrix} \vec{x}_1^\top \\ \vec{x}_2^\top \\ \vdots \\ \vec{x}_m^\top \end{bmatrix} := X = U\Sigma V^\top. \tag{6}$$

where $U \in \mathbb{R}^{m\times m}$ is an orthonormal matrix, and $V \in \mathbb{R}^{n\times n}$ is an orthonormal matrix. Here, $\Sigma = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}$. Here we assume that we have more data points than the dimension of our space (that is, $m > n$). Also, the transformation V in part e) is the same V in this factorized representation.

Set up a least squares formulation for estimating $\vec{a}$ and find the solution to the least squares. Why might this factorization help us compute $\widehat{\vec{a}}$ faster?

Note: again, take this factorization as a given. We will go over how to find U, Σ, and V later.

Contributors:

- Neelesh Ramachandran.
- Kuan-Yun Lee.
- Anant Sahai.
- Kumar Krishna Agrawal.