

# Lecture 3: 9-11-18

①

Last time: PCA, i.e. computing an orthogonal basis  $u_1, \dots, u_d \in \mathbb{R}^d$  such that projecting data onto  $u_1$  gives the variance maximizing 1-dimensional projection, projecting onto  $u_1, u_2$  " " 2-dimensional projection, etc.

• We only proved ~~the~~ the  $u_1$  case, by considering

$$F(u) = \underbrace{u^T \Sigma u}_{\text{to maximize}} - \underbrace{\lambda(u^T u - 1)}_{\text{constraint that } u^T u = 1}$$

differentiating, setting = 0, choosing best critical point.

Q: How to go from  $u_1$  to  $\{u_1, u_2\}$   
or more generally

$$\{u_i\}_{i=1}^{r-1} \text{ to } \{u_i\}_{i=1}^r ?$$

• Let's proceed inductively. Suppose that  $\{u_i\}_{i=1}^{r-1}$  give the variance-maximizing  $r-1$  orthogonal vectors, and that these are the eigenvectors of  $\Sigma$  corresponding to the largest  $r-1$  eigenvalues.

• So, my new vector  $u$  should 1.) be orthogonal to  $u_1, u_2, \dots, u_{r-1}$   
2.) be variance-maximizing among all ~~other~~

unit norm vectors satisfying 1.)

So, want to maximize  $u^T \sum u$  as before, but the constraints are more serious: 1)  $u^T u = 1$  ( $u$  should be unit-norm)

- 2.)  $u^T u_1 = 0$
  - 3.)  $u^T u_2 = 0$
  - ...
  - r.)  $u^T u_{r-1} = 0$
- ( $u$  should be orthogonal to  $u_i, i=1, \dots, r-1$ .)

So, the Lagrange multiplier formulation seeks to maximize

$$F(u) = u^T \sum u - \alpha_1 (u^T u - 1) - \alpha_2 (u^T u_1 - 0) - \dots - \alpha_r (u^T u_{r-1} - 0)$$

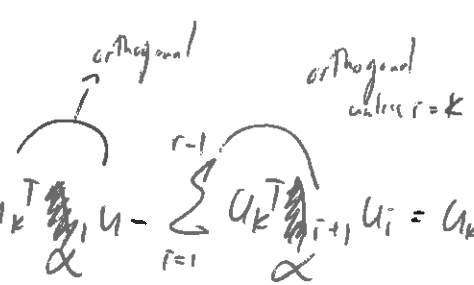
$$= u^T \sum u - \alpha_1 (u^T u - 1) - \sum_{i=1}^{r-1} \alpha_{i+1} u^T u_i$$

Nasty! But, lets proceed as before and differentiate  $F$ :

$$\frac{\partial F}{\partial u} = 2 \sum u - 2 \alpha_1 u - \sum_{i=1}^{r-1} \alpha_{i+1} u_i$$

Set = 0 to find critical points:

$$2 \sum u - 2 \alpha_1 u - \sum_{i=1}^{r-1} \alpha_{i+1} u_i = 0$$



Trick: hit both sides with  $u_k^T$ :  $2 u_k^T \sum u - 2 u_k^T \alpha_1 u - \sum_{i=1}^{r-1} u_k^T \alpha_{i+1} u_i = u_k^T \cdot 0$

$\Leftrightarrow 2 u_k^T \sum u - \alpha_{k+1} = 0$

$$\begin{aligned}
 \text{But, } u_r^T \sum u & \\
 &= u^T \sum u_k \\
 &= u^T (\lambda_k u_k), \text{ since } u_k \text{ is an eigenvector of } \sum \\
 &= \lambda_k u^T u_k \\
 &= \lambda_k \cdot 0 \quad \text{since } u \text{ is orthogonal to } u_k. \\
 &= 0
 \end{aligned}$$

So,  $\alpha_{k+1} = 0$  for  $k=1, \dots, r-1$

$$\begin{aligned}
 \Rightarrow \text{our equation is } 2\sum u - 2\alpha_1 u &= 0 \\
 \Leftrightarrow \sum u &= \alpha_1 u
 \end{aligned}$$

$\Leftrightarrow u$  is an eigenvector of  $\sum$  with eigenvalue  $\alpha_1$ .

It is not hard to see that to maximize variance and remain orthogonal to  $u_1, \dots, u_{r-1}$ ,  $\alpha_1 = \lambda_r$  the next biggest eigenvalue, so  $u = u_r$ , as desired.

Computational complexity: PCA has several steps:

1) Centering data: Compute  $\mu = \frac{1}{n} \sum_{i=1}^n x_i \Rightarrow O(n \cdot d)$

2) Computing  $\sum$ :  $\sum = \sum_{i=1}^n u_i^T u_i \Rightarrow O(n \cdot d^2) = O(nd^2)$

3) Eigendecomposition of  $\sum \in \mathbb{R}^{d \times d}$ :  $O(d^3)$

$\Rightarrow$  Overall  
 $O(nd + nd^2 + d^3)$   
 $= O(nd^2 + d^3)$   
 So, big  $n$  ok, big  $d$  not really...