LECTURE 24

Review PCA, examples next week

Q. Suppose we sample a normally distributed random variable

$$\vec{x} = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{bmatrix}$$

at ten different times to get observations $\vec{x}_1, \dots, \vec{x}_{10}$. Set $X = [\vec{x}_1 \dots \vec{x}_{10}]$.

(a) What is the sample covariance matrix S, in terms of X?
(b) What are its dimensions?
(c) What does S_{ij} measure?

—

(a) $S = \frac{1}{9} XX^T$

(b) $X \in \mathbb{R}^{4\times10} \Longrightarrow XX^T = \boxed{4\times10}\ \boxed{10\times4} = \boxed{4\times4}$

(c) $S_{ij}\ (=S_{ji})$ measures the covariance b/w p_i and p_j

---

Last time: T = temp, P = pressure, H = humidity. (p_1, p_2, p_3)

$$X = \underbrace{\begin{bmatrix} x_{11} & \cdots \\ x_{12} & \\ x_{13} & \end{bmatrix}}_{n} \Big\} d \qquad S = \frac{1}{n} XX^T = \Phi \Lambda \Phi$$

↑ SVD of symmetric matrix
singular values = eig. values

(Aside $|\det(M)| = \prod \sigma_i$
$\det(M) = \prod \lambda_i$ in general)

$\Phi = [\varphi_1\ \varphi_2\ \varphi_3]$

If $\varphi_1 \sim T+H$, ie φ_1 is unit vector in direction of $T+H$, then T and H are most correlated, and we really ought to consider a different basis (eg. "mugginess")

Suppose instead $T+2H \sim \varphi_1$. Then there's some scaling we need to do on H! (These all depend on our units!)

We could have also seen $\varphi_1 \sim T-P+H$, etc.
$\Longrightarrow$ indicates that this mixture of properties should be what we study; ie there is some property that $T, -P, H$ are all contributing to.

Here we only looked at the 1st principle component! Lets see what happens when the dimension of the data (d) is large.

eg. Semantics? (also not a linguist)

Imagine Words = { top 10000 words in English, maybe combined into buckets based on plurals, conjugation, etc } — most common words (the, and, etc)

eg. apple ~ apples, butter ~ butters ~ buttered ~ buttery

$\Longrightarrow$ d is huge. n = # documents you look at

Words = { words in English, up to plurals, conjugation, etc }

eg. Oxford dictionary has ~170,000 words, where "apple" and "apples" are under the same entry.

The dimension d is huge.

① Look at n online recipes from a particular recipe website.

Form $X = [\vec{x}_1 \; \cdots \; \vec{x}_n]$ each $\vec{x}_i \in \mathbb{R}^d$.

(This is a term-document matrix!)

- Dimension reduction: Maybe $\lambda_1, \dots, \lambda_{200}$ are large enough to be interesting; treat the rest as noise.

- If w_1 = "apple" and w_2 = "butter" (or equivalent)

$$S_{12} = \frac{1}{n} \sum \Big((\# w_1 \text{ in } D_j) - (\text{avg } \# w_1 \text{ in all docs}) \Big) \Big((\# w_2 \text{ in } D_j) - (\text{avg } \# w_1 \text{ in all Docs}) \Big)$$

- Maybe you'll see $\varphi_{150} = \frac{1}{\sqrt{5}} (w_1 + 2w_2)$ ie butter shows up more, so you need to actually weight the "sugar" # of instances

$$= \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 2 \\ 0 \\ \vdots \\ \vdots \end{bmatrix}$$

This is the 150th principal component out of d columns of Φ

- If w_3, w_4, w_5 = ketchup, mustard, relish, maybe you'll see

$$\varphi_{100} = \frac{1}{\sqrt{3}} (w_3 + w_4 + w_5)$$

$\Rightarrow$ these show up together a lot: high correlation, covariance ie. they vary together: if a recipe has one, it probably has the others too, and maybe they don't show up independently very much.

- Note that $\varphi_7 \perp \varphi_{1528}$; they are measuring uncorrelated variables

② Compare with Y = database of articles on technology

then here maybe $\varphi_{20} = \frac{1}{\sqrt{2}}$(apple + computer)

while S_{ij} for w_i = apple, w_j = butter may be rather small!

Summary of matrix methods:

- vector, matrix norms to measure proximity, distance, errors from approximations, etc
- linear regression: fit data to a predetermined model
  - minimize residual vector
  - condition # of system
- Matrix decompositions
  - $PA = LU$ LU decomposition : Gauss Elim (w/ partial pivoting)
  - QR decomp : CGS, MGS, Householder Δ'n
  - SVD : singular values + their meaning, the bases vectors u_i and v_i
- Other properties / aspects of matrices:
  - Pseudoinverse
  - Orthogonality

Next 5 classes: applications of matrix methods in various contexts

Now: final project info