CS4786
Lec 4
2/3/15

Announcements:

1. Office hour times and locations are nearly finalized, and will be updated on the course webpage tonight.
   → http://www.cs.cornell.edu/courses/cs4786

2. Piazza (the course question & answer forum) has been active, and you are encouraged to sign up.
   - suggestion: you can use Piazza to form study groups.

3. If we received an AØ from you, we entered you into CMS.
   → http://cms.csuglab.cornell.edu.
   This will allow you to submit subsequent assignments.
   ⊗ We'll discuss and return AØ @ the end of class today. (You need to pick up your AØ to receive credit for it.)

4. Corrections to the handout from last lecture have been posted to the course website. (Full lecture notes still to come.)

Topic outline: an introduction to Canonical Correlation Analysis (CCA)

• From covariance ($\boxed{\Sigma}$)	to	correlation (different matrix, gets rid of units)
• From orthogonality (eigenvectors of $\boxed{\Sigma}$)	to	non-correlatedness (different eigenvectors)

purpose: better representation of the data items

(side effect: covariance matrix tells you about ~~the~~ relationships btwn features)

purpose: better representation of relationship between 2 sets of features.

Pedagogical note: in today's lecture, will have a demo in a 3rd language: R

"Crib sheet": (reminders, mostly)

Data vectors are $d \times 1$, for d features. $\mathbf{x}_i : \left.\begin{bmatrix} \\ \\ \end{bmatrix}\right\} d$

Ex: each document has counts for each ~~word~~ possible word

Ex: a student has ~~score~~ grades in each of math, English, history

Data matrix:

$$X: \quad n \left\{ \overbrace{\begin{bmatrix} \text{---} \; \mathbf{x}_1^T \; \text{---} \\ \text{---} \; \mathbf{x}_2^T \; \text{---} \\ \vdots \\ \text{---} \; \mathbf{x}_n^T \; \text{---} \end{bmatrix}}^{d} \right.$$

Mean vector: $\text{---}\; \mu^T \;\text{---}$ (sum each of the data matrix's columns, divide each entry by $1/n$).

Covariance matrix $\boxed{\Sigma}$: $d \times d$, the i,j^{th} entry tells you ~~about how~~ something about how the i^{th} & j^{th} feature relate (covary).

$$= \frac{1}{n} \sum_{t=1}^{n} \underbrace{(\mathbf{x}_t - \mu)}_{d \times 1} \underbrace{(\mathbf{x}_t - \mu)^T}_{1 \times d}$$

. Let X_c be the mean-centered version of X.

$$= \frac{1}{n} X_c^T X_c \quad , \quad \text{~~so entries look like~~}$$

$$\begin{bmatrix} \text{— } 1^{st} \text{ column of } X_c = 1^{st} \text{ feature —} \\ \text{— } 2^{nd} \text{ column of } X_c \,(2^{nd} \text{ feature}) \text{ —} \\ \vdots \\ d^{th} \text{ column of } X_c \,(d^{th} \text{ feature}) \end{bmatrix} \begin{bmatrix} | & | & & | \\ 1^{st} f & 2^{nd} f & \cdots & d^{th} f \\ | & | & & | \end{bmatrix}$$

$\Rightarrow$ entries are the $1/n$-scaled inner products of ~~[scribbled]~~ the features. ~~features~~.

And remember, $\mathbf{v}_1 \cdot \mathbf{v}_2 = \|\mathbf{v}_1\|_2 \|\mathbf{v}_2\| \underbrace{\cos(\angle(\mathbf{v}_1, \mathbf{v}_2))}$

big when the vectors point in the same direction.

They're also $E\big((i^{th} \text{ feature} - \text{its mean})(j^{th} \text{ feature} - \text{its mean})\big)$

hence ~~covari~~ the name "covariance".

Canonical correlation analysis:

Assume the features can be broken into two sets.

Find: { a linear combination of the first set, and
{ a linear combination of the second set

such that the correlation of these two is maximized.

Subsequent pairs should be uncorrelated to the previous pairs

Asides (gotchas when checking other sources)

- sometimes $\boxed{\Sigma}$ is used for a related matrix, the ~~matrix~~ diagonal matrix of singular values for X.
- Some ~~packages use~~ software packages use ~~the~~ sample covariance matrix, which divides by $n-1$ instead of n, for the purposes of unbiased estimation
- sometimes X is assumed to already be mean-centered.