From Last Meeting

Developing Mathematics behind PCA:

Linear Algebra Review: vector spaces, subspaces, bases, inner products, norms, orthonormal bases, transforms, projections, eigenvalue decompositions

Multivariate Probability Review: random vectors, theoretical (distributional) mean vector (center) and covariance matrix (spread), empirical (sample) mean vector

Multivariate Probability Review, (cont.)

Empirical versions (cont.)

And estimate the "theoretical cov.", with the "sample cov.":

$$\hat{\Sigma} = \frac{1}{n-1} \begin{pmatrix} \sum_{i=1}^{n} (X_{i1} - \overline{X}_1)^2 & \cdots & \sum_{i=1}^{n} (X_{i1} - \overline{X}_1)(X_{id} - \overline{X}_d) \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{n} (X_{id} - \overline{X}_d)(X_{i1} - \overline{X}_1) & \cdots & \sum_{i=1}^{n} (X_{id} - \overline{X}_d)^2 \end{pmatrix}$$

Normalizations:

- $\frac{1}{n-1}$ gives unbiasedness
- $\frac{1}{n}$ gives MLE in Gaussian case

Multivariate Probability Review, (cont.)

Outer product representation:

$$\hat{\Sigma} = \frac{1}{n-1}\sum_{i=1}^{n}\begin{pmatrix} (X_{i1}-\bar{X}_1)^2 & \cdots & (X_{i1}-\bar{X}_1)(X_{id}-\bar{X}_d) \\ \vdots & \ddots & \vdots \\ (X_{id}-\bar{X}_d)(X_{i1}-\bar{X}_1) & \cdots & (X_{id}-\bar{X}_d)^2 \end{pmatrix}$$

$$\hat{\Sigma} = \frac{1}{n-1}\sum_{i=1}^{n}\left(\underrightarrow{X_i} - \underrightarrow{\bar{X}}\right)\left(\underrightarrow{X_i} - \underrightarrow{\bar{X}}\right)^t = \tilde{X}\tilde{X}^t,$$

where: $$\tilde{X} = \frac{1}{\sqrt{n-1}}\left(\underline{X_1} - \underline{\bar{X}} \quad \cdots \quad \underline{X_n} - \underline{\bar{X}}\right)_{d\times n}$$

PCA as an optimization problem

Find "direction of greatest variability"

Show CorneaRobust\SimplePCAeg.ps

Given a "direction vector", $\underline{u} \in \Re^d$ (i.e. $\|\underline{u}\| = 1$)

Projection of $\underline{X_i} - \underline{\overline{X}}$ in the direction $\underline{u}$: $P_{\underline{u}}\left(\underline{X_i} - \underline{\overline{X}}\right) = \left\langle \underline{X_i} - \underline{\overline{X}}, \underline{u} \right\rangle \underline{u}$

Variability in the direction $\underline{u}$:

$$\sum_{i=1}^{n} \left\| P_{\underline{v}}\left(\underline{X_i} - \underline{\overline{X}}\right) \right\|^2 = \sum_{i=1}^{n} \left\| \left\langle \underline{X_i} - \underline{\overline{X}}, \underline{u} \right\rangle \underline{u} \right\|^2 = \sum_{i=1}^{n} \left\langle \underline{X_i} - \underline{\overline{X}}, \underline{u} \right\rangle^2 \left\| \underline{u} \right\|^2 =$$

$$= \sum_{i=1}^{n} \left\langle \underline{X_i} - \underline{\overline{X}}, \underline{u} \right\rangle^2 = \sum_{i=1}^{n} \left(\left(\underline{X_i} - \underline{\overline{X}} \right)^t \underline{u} \right)^2 =$$

$$= \sum_{i=1}^{n} \underline{u}^t \left(\underline{X_i} - \underline{\overline{X}} \right) \left(\underline{X_i} - \underline{\overline{X}} \right)^t \underline{u} =$$

PCA as an optimization problem (cont.)

Variability in the direction $\underline{u}$:

$$\sum_{i=1}^{n} \left\| P_{\underline{v}} \left(\underline{X}_i - \overline{\underline{X}} \right) \right\|^2 = \underline{u}^t \left(\sum_{i=1}^{n} \left(\underline{X}_i - \overline{\underline{X}} \right) \left(\underline{X}_i - \overline{\underline{X}} \right)^t \right) \underline{u} = (n-1) \underline{u}^t \hat{\Sigma} \underline{u}$$

i.e. (prop'l to) a "quadratic form in the covariance matrix"

Simple solution comes from eigenvalue representation of $\hat{\Sigma}$:

$$\hat{\Sigma} = BDB^t$$

where $B = \left(\underline{v}_1, \ldots, \underline{v}_d \right)$ is orthonormal, and $D = \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_d \end{pmatrix}$

PCA as an optimization problem (cont.)

Variability in the direction $\underline{u}$:

$$\sum_{i=1}^{n} \left\| P_{\underline{u}} \left(\underline{X}_i - \underline{\overline{X}} \right) \right\|^2 = (n-1)\underline{u}^t \left(BDB^t \right) \underline{u} = (n-1)\left(\underline{u}^t B \right) D \left(B^t \underline{u} \right)$$

But $\quad B^t \underline{u} = \begin{pmatrix} \underline{v_1}^t \\ \vdots \\ \underline{v_d}^t \end{pmatrix} \underline{u} = \begin{pmatrix} \langle \underline{v_1}, \underline{u} \rangle \\ \vdots \\ \langle \underline{v_d}, \underline{u} \rangle \end{pmatrix}$ = “B transform of u”

= “$\underline{u}$ rotated into B coordinates”,

and the “diagonal quadratic form” becomes

$$\sum_{i=1}^{n} \left\| P_{\underline{u}} \left(\underline{X}_i - \underline{\overline{X}} \right) \right\|^2 = (n-1) \sum_{j=1}^{d} \lambda_j \left\langle \underline{v_j}, \underline{u} \right\rangle^2$$

PCA as an optimization problem (cont.)

Now since B is an orthonormal basis matrix,

$$\underline{u} = \sum_{j=1}^{d} \left\langle \underline{v_j}, \underline{u} \right\rangle \underline{v_j} \quad \text{and} \quad 1 = \left\| \underline{u} \right\|^2 = \sum_{j=1}^{d} \left\langle \underline{v_j}, \underline{u} \right\rangle^2$$

So the rotation $B^t \underline{u} = \begin{pmatrix} \left\langle \underline{v_1}, \underline{u} \right\rangle \\ \vdots \\ \left\langle \underline{v_d}, \underline{u} \right\rangle \end{pmatrix}$ gives a "distribution of the (unit)

energy of $\underline{u}$ over the $\hat{\Sigma}$ eigen-directions"

And $\sum_{i=1}^{n} \left\| P_{\underline{u}} \left(\underline{X_i} - \underline{\overline{X}} \right) \right\|^2 = (n-1) \sum_{j=1}^{d} \lambda_j \left\langle \underline{v_j}, \underline{u} \right\rangle^2$ is maximized (over $\underline{u}$),

by putting all energy in the "largest direction", i.e. $\underline{u} = \underline{v_1}$,

where "eigenvalues are ordered", $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$

PCA as an optimization problem (cont.)

Notes:

- Solution is unique when $\lambda_1 > \lambda_2$
- Otherwise have solutions in "subspace gen'd by 1st $\underline{v}$s"
- Projecting onto subsp. $\perp$ to $\underline{v_1}$, gives $\underline{v_2}$ as "next dir'n"

Again show CorneaRobust\SimplePCAeg.ps

- Continue through $\underline{v_3}, \ldots, \underline{v_d}$
- Replace $\hat{\Sigma}$ by Σ to get "theoretical PCA"
- Which is "estimated" by the empirical version

PCA redistribution of "energy"

Recall "amount of structure" is quantified as:

"sum of squares about the mean" = $\sum_{i=1}^{n} \left\| \underline{X}_i - \overline{\underline{X}} \right\|^2$

And insight comes from "decomposition" (ANOVA)

Show CurvDat\ParabsCurvDat.ps

PCA redistribution of "energy" (cont.)

$$\sum_{i=1}^{n} \left\| \underline{X}_i - \underline{\bar{X}} \right\|^2 = \sum_{i=1}^{n} \left(\underline{X}_i - \underline{\bar{X}} \right)^t \left(\underline{X}_i - \underline{\bar{X}} \right) = (n-1) tr\left(\tilde{X}^t \tilde{X} \right)$$

where: $$\tilde{X} = \frac{1}{\sqrt{n-1}} \left(\underline{X}_1 - \underline{\bar{X}} \quad \cdots \quad \underline{X}_n - \underline{\bar{X}} \right)_{d \times n}$$

$$\sum_{i=1}^{n} \left\| \underline{X}_i - \underline{\bar{X}} \right\|^2 = (n-1) tr\left(\tilde{X}\tilde{X}^t \right) = (n-1) tr\left(\hat{\Sigma} \right)$$

- "Energy is expressed in trace of covariance matrix"

PCA redistribution of "energy" (cont.)

$$\frac{1}{n-1}\sum_{i=1}^{n}\left\|\underline{X}_i - \overline{\underline{X}}\right\|^2 = tr\left(BDB^t\right) = tr\left(BB^tD\right) = tr(D) = \sum_{j=1}^{d}\lambda_j$$

- Eigenvalues provide "atoms of SS decomposition"
- Useful Plots are:

"Power Spectrum": λ_j vs. j

"log Power Spectrum": $\log(\lambda_j)$ vs. j

"Cumulative Power Spectrum": $\sum_{j'=1}^{j}\lambda_{j'}$ vs. j

Show CurvDat\ParabsCurvDat.ps

PCA dual problem

Idea: Recall for HDLSS settings:

Sample size = n < d = dimension

So $rank\left(\hat{\Sigma}\right) \leq n$, and $\lambda_{n+1} = \lambda_d = 0$

Thus have "really only n dimensional eigen problem"

Can exploit this to boost computation speed

PCA dual problem (cont.)

Again use notation: $\tilde{X} = \frac{1}{\sqrt{n-1}} \left(\underline{X}_1 - \underline{\overline{X}} \quad \cdots \quad \underline{X}_n - \underline{\overline{X}} \right)_{d \times n}$

Recall: $\hat{\Sigma}_{d \times d} = \tilde{X}\tilde{X}^t$ has the eigenvalue decomp. $\hat{\Sigma} = BDB^t$

The "dual eigen problem" replaces rows by columns in $\tilde{X}$:

Let $\Sigma^*_{n \times n} = \tilde{X}^t \tilde{X}$, and find B^*, D^*, so that $\Sigma^* = B^* D^* B^{*t}$

(now only $n < d$ dimensional)

PCA dual problem (cont.)

Now suppose know sol'n to dual problem, i.e. know B^* and D^*

How do we find B and D?

A heuristic approach:

i. want B so that

$$D = B^t \hat{\Sigma} B = B^t \tilde{X} \tilde{X}^t B$$

PCA dual problem (cont.)

ii. choose B to introduce form $\tilde{X}^t \tilde{X} = \Sigma^*$,
i.e. $B = \tilde{X} C$ (for some C), then

$$D = C^t \tilde{X}^t \left(\tilde{X} \tilde{X}^t \right) \tilde{X} C = C^t \left(\tilde{X}^t \tilde{X} \right) \left(\tilde{X}^t \tilde{X} \right) C = C^t \Sigma^* \Sigma^* C$$

iii. choose C to relate to $\Sigma^* = B^* D^* B^{*t}$, i.e. $B^{*t} \Sigma^* B^* = D^*$
i.e. $C = B^* R$ (for some R), then

$$D = C^t \Sigma^* \left(B^* B^{*t} \right) \Sigma^* C = \left(R^t B^{*t} \right) \Sigma^* B^* B^{*t} \Sigma^* \left(B^* R \right)$$

$$D = R^t \left(B^{*t} \Sigma^* B^* \right) \left(B^{*t} \Sigma^* B^* \right) R = R^t D^* D^* R$$

PCA dual problem (cont.)

iii. Choose R to "preserve energy",

i.e. "make B orthonormal",

i.e. "make B a rotation matrix",

i.e. choose $R = \left(D^{*}\right)^{-1/2}$, then

$D = D^{*}$, i.e. same (nonzero) eigenvalues!

PCA dual problem (cont.)

Heuristic summary: Want $B = \tilde{X}C = \tilde{X}\left(B^{*}R\right) = \tilde{X}B^{*}\left(D^{*}\right)^{-1/2}$

Technical problems:

- dimensions wrong: $B_{d\times d}$, $\tilde{X}_{d\times n}$, $B^{*}_{n\times n}$, $D^{*}_{n\times n}$
- $D^{*}_{n\times n}$ not full rank?, thus:
  - root inverse doesn't exist?
  - B is not a basis matrix

PCA dual problem (cont.)

Solution 1: Assume $D^* = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$ is of full rank,

i.e. $\lambda_1 \geq \cdots \geq \lambda_n > 0$

Then let $\breve{B}_{d\times n} = \tilde{X}B^*\left(D^*\right)^{-1/2}$,

Where

$$\left(D^*\right)^{-1/2} = \begin{pmatrix} \lambda_1^{1/2} & & 0 \\ & \ddots & \\ 0 & & \lambda_n^{1/2} \end{pmatrix}$$

PCA dual problem (cont.)

And "fill out the rest of B" with "columns in null space",

I.e. let $G_{d \times (d-n)}$ be $d-n$ orthonormal column vectors,

that are orthogonal to $\tilde{X}$ (compute by Gram-Schmidt process)

Thus "pad $\breve{B}$ out to a basis matrix", by defining:

$$B = \begin{pmatrix} \breve{B} & G \end{pmatrix}$$

PCA dual problem (cont.)

Check orthonormality:

$$B^t B = \begin{pmatrix} \breve{B}^t \\ G^t \end{pmatrix} \begin{pmatrix} \breve{B} & G \end{pmatrix} = \begin{pmatrix} \breve{B}^t \breve{B} & 0 \\ 0 & I \end{pmatrix}$$

but

$$\breve{B}^t \breve{B} = \left(\tilde{X} B^* \left(D^*\right)^{-1/2}\right)^t \left(\tilde{X} B^* \left(D^*\right)^{-1/2}\right) = \left(\left(D^*\right)^{-1/2} B^{*t} \tilde{X}^t\right)\left(\tilde{X} B^* \left(D^*\right)^{-1/2}\right)$$

$$\breve{B}^t \breve{B} = \left(D^*\right)^{-1/2} B^{*t} \left(\tilde{X}^t \tilde{X}\right) B^* \left(D^*\right)^{-1/2} = \left(\left(D^*\right)^{-1/2} B^{*t}\right) \Sigma^* \left(B^* \left(D^*\right)^{-1/2}\right)$$

$$\breve{B}^t \breve{B} = \left(D^*\right)^{-1/2} \left(B^{*t} \Sigma^* B^*\right) \left(D^*\right)^{-1/2} = \left(D^*\right)^{-1/2} D^* \left(D^*\right)^{-1/2} = I$$

so B is orthonormal.

PCA dual problem (cont.)

Check diagonalization:

$$B^t \hat{\Sigma} B = \begin{pmatrix} \breve{B}^t \\ G^t \end{pmatrix} \hat{\Sigma} \begin{pmatrix} \breve{B} & G \end{pmatrix} = \begin{pmatrix} \breve{B}^t \\ G^t \end{pmatrix} \begin{pmatrix} \hat{\Sigma} \breve{B} & \hat{\Sigma} G \end{pmatrix} = \begin{pmatrix} \breve{B}^t \hat{\Sigma} \breve{B} & \breve{B}^t \hat{\Sigma} G \\ G^t \hat{\Sigma} \breve{B} & G^t \hat{\Sigma} G \end{pmatrix}$$

but

$$\breve{B}^t \hat{\Sigma} \breve{B} = \left(D^*\right)^{-1/2} B^{*t} \tilde{X}^t \hat{\Sigma} \tilde{X} B^* \left(D^*\right)^{-1/2} =$$
$$= \left(D^*\right)^{-1/2} B^{*t} \tilde{X}^t \tilde{X} \tilde{X}^t \tilde{X} B^* \left(D^*\right)^{-1/2} =$$
$$= \left(D^*\right)^{-1/2} B^{*t} \Sigma^* \Sigma^* B^* \left(D^*\right)^{-1/2} = \left(D^*\right)^{-1/2} B^{*t} \Sigma^* \left(B^* B^{*t}\right) \Sigma^* B^* \left(D^*\right)^{-1/2} =$$
$$= \left(D^*\right)^{-1/2} \left(B^{*t} \Sigma^* B^*\right) \left(B^{*t} \Sigma^* B^*\right) \left(D^*\right)^{-1/2} = \left(D^*\right)^{-1/2} D^* D^* \left(D^*\right)^{-1/2} = D^*$$

PCA dual problem (cont.)

And using the orthogonality of the columns of $\tilde{X}$ and G

$$\breve{B}^t \hat{\Sigma} G = \breve{B}^t \left(\tilde{X} \tilde{X}^t \right) G = \left(\breve{B}^t \tilde{X} \right) \left(\tilde{X}^t G \right) = \left(\breve{B}^t \tilde{X} \right) 0 = 0_{n \times (d-n)}$$

$$G^t \hat{\Sigma} \breve{B} = G^t \left(\tilde{X} \tilde{X}^t \right) \breve{B} = \left(G^t \tilde{X} \right) \left(\tilde{X}^t \breve{B} \right) = 0 \left(\tilde{X}^t \breve{B} \right) = 0_{(d-n) \times n}$$

$$G^t \hat{\Sigma} G = G^t \left(\tilde{X} \tilde{X}^t \right) G = \left(G^t \tilde{X} \right) \left(\tilde{X}^t G \right) = 0 \cdot 0 = 0_{(d-n) \times (d-n)}$$

PCA dual problem (cont.)

Thus:

$$B^t \hat{\Sigma} B = \begin{pmatrix} D^* & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 & & \cdots & & 0 \\ 0 & \ddots & & & & \\ & & \lambda_n & \ddots & & \vdots \\ \vdots & & \ddots & 0 & & \\ & & & & \ddots & 0 \\ 0 & & \cdots & & 0 & 0 \end{pmatrix} = D$$

PCA dual problem (cont.)

Aside about orthogonal component G:

Usually don't need to compute,

Since only want "eigenvectors for non-zero eigenvalues"