From last meetings

Class Web Page:
http://www.stat.unc.edu/faculty/marron/321FDAhome.html

Finding Structure in Populations of Complex Objects: PCA

Important duality:
Object Space ↔ Feature Space

Cornea Data: motivated robust PCA

Robust PCA: Toy Example I

E.g. previous “random parabolas”, with an outlier(?) added

Show CurvDat\Parabs1outRaw.ps

Notes:

- this is not a coordinate-wise outlier
- but recall many other directions in $\Re^d$
- very different in “shape” (nearly orthogonal?)
- random parabolas “live in a special part” of $\Re^d$

Robust PCA: Toy Example I (cont.)

Effects of outlier on PCA:

Show CurvDat\Parabs1outCurvDat.ps and again show CurvDat\ParabsCurvDat.ps (flip back and forth)

- Very minor effect on mean (since it feels outliers less)

- Not large effect on PC1 (large variab’ty in that dir’n “wins”)

Again show CorneaRobust\OutliersPCA.ps

- Major effect on PC2 (see both proj’ns and mean ± ext.)

- Still some effect on PC3

- Major redist’n of Sums of Squares (signal power)

- PC didn’t find “best” directions? (but 3d subspace is right)

Naïve Robust PCA:

Spearman Correlation:

Idea: base PCA on correlation matrix computed on ranks
from: "Rank based" nonparametric statistics

(for $X_1,....,X_n$, ranks are $(1),...,(n)$, where $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$)

Result: Small improvements, but outlier is still PC2

Show CurvDat\Parabs1outCurvDatSCorr.ps

Better Robust PCA

Recall Huber's L^1 M-estimate of "center"

show CorneaRobust\L1Center.ps

Corresponding PCA: work with data projected to sphere

Show CorneaRobust\OutliersPCA.ps and CorneaRobust\ SphericalPCA.ps

Toy Example 1:

Show CurvDat\Parabs1outCurvDatSph.ps and compare with CurvDat\ParabsCurvDat.ps

- PC1 and PC2 very similar to original (except scale)
- Outlier goes into PC3

Robust PCA: Toy Example II

E.g. previous “random parabolas”, with 2 outliers added

Show CurvDat\Parabs2outRaw.ps

Ordinary PCA:

Show CurvDat\Parabs2outCurvDat.ps

- PC1 & PC2: feels 1st outlier as before
- PC3 & PC4: “tilt” and 2nd outlier are confounded
- found right 4d subspace, but poor directions within
- can see outliers in jitter plots / smoothed histograms

Robust PCA: Toy Example II

Spearman PCA: fails again (essent’ly same as ordinary PCA)

Show CurvDat\Parabs2outCurvDatSCorr.ps

Spherical PCA:

Show CurvDat\Parabs2outCurvDatSph.ps

- PC1 & PC2: similar to no outlier case
- PC3 & PC4: outliers appear here
- Outliers are “mixed” between 3 & 4 (by previous dir’ns)
- i.e. didn’t find “nicest 4 directions”

Numerical Aside

In these toy examples, $n = 50$, $d = 10$, so could also have used:

1. Projection pursuit robust PCA

2. PCA based on standard robust covariance matrices

But for the cornea data, $n = 43$, $d = 66$, so these don’t work.

Thus Spherical PCA is only choice

Elliptical PCA

Result: Spherical PCA is good, not great, for the cornea data

Idea: problem is “Fourier type signal compression”

“high frequency terms” << “low frequency terms”

Solution: Replace “sphere” by “ellipse”,
Which reflects “proper scaling (of coordinate axes)”

Problem: simple and computable ellipse?

Elliptical PCA (cont.)

Three Step solution (keying on “parallel to coordinate axes”):

1. Rescale coordinate axes by Median Absolute Deviation:

$$MAD = \underset{i}{med} \left| X_i - \underset{i'}{med}\, X_{i'} \right|$$

2. Project onto circle

3. Return axes to original scale

Show CorneaRobust\EllipticalPCA.ps

Elliptical PCA for Cornea Data

Show CorneaRobust\NORMLWR.MPG

PC1:

Show CorneaRobust\NORM100.MPG and CorneaRobust\NORM122.MPG

- robust center slightly better
- same main lesson (about this direction)
- robust slightly better at edge

Elliptical PCA for Cornea Data

PC2:

Show CorneaRobust\NORM200.MPG and CorneaRobust\NORM222.MPG

- same main lesson
- outlier edge effect eliminated

PC3:

Show CorneaRobust\NORM300.MPG and CorneaRobust\NORM322.MPG

- outlier effect eliminated
- main effect slightly diminished (no free lunch)

Elliptical PCA

Problem: nonlinear analysis leads to systematic bias

Approach (N. Locantore): iterative improvement

Cornea Data: not a major issue

Cornea Data – Radius of analysis

Idea: can control radius of disk

General purpose: 4.2 mm, e.g. study PRK recovery

Show CorneaRobust\EgPRKmorph.mpg

Avoid edge effects (usually completely):   3 mm

Here: 4.0 mm   (accentuated impact of robustness!)

Big Picture

Goal 1: Understanding Population Structure

PCA: illustrated with Cornea Data

Goal 2: Discrimination (classification)

Corpora Callosa data

F. L. D. failed

Now derive “Orthogonal Subspace Projection”