From Last Meeting

Studied Approximation of Corpora Callosa

- by Fourier (raw and centered)
- by PCA

Fisher Linear Discrimination

Recall Toy Problem:

Show HDLSS/HDLSSod1Raw.ps

Want to find "separating direction vector"

Recall PCA didn't work

Show HDLSS/HDLSSod1PCA.ps

Also "difference between means" doesn't work:

Show HDLSS/HDLSSod1Mdif.ps

Fisher Linear Discrimination

A view of Fisher Linear Discrimination:

Adjust to "make covariance structure right"

Show HDLSS/HDLSSod1FLD.ps

Mathematical Notation (vectors with dimension d):

Class 1: $\underrightarrow{X}_1^{(1)},...,\underrightarrow{X}_{n_1}^{(1)}$ Class 2: $\underrightarrow{X}_1^{(2)},...,\underrightarrow{X}_{n_2}^{(2)}$

Class Centerpoints: $\overline{\underrightarrow{X}}^{(1)} = \frac{1}{n_1}\sum_{i=1}^{n_1}\underrightarrow{X}_i^{(1)}$ and $\overline{\underrightarrow{X}}^{(2)} = \frac{1}{n_2}\sum_{i=1}^{n_2}\underrightarrow{X}_i^{(2)}$

Fisher Linear Discrimination (cont.)

Covariances: $\hat{\Sigma}^{(j)} = \tilde{X}^{(j)} \tilde{X}^{(j)t}$, for $j = 1,2$ (outer products)

Based on "normalized, centered data matrices":

$$\tilde{X}^{(j)} = \frac{1}{\sqrt{n_j}} \left(\underrightarrow{X}_1^{(j)} - \underrightarrow{\overline{X}}^{(j)}, \ldots, \underrightarrow{X}_{n_j}^{(j)} - \underrightarrow{\overline{X}}^{(j)} \right)$$

note: Use "MLE" version of normalization, for simpler notation

Terminology (useful later): $\hat{\Sigma}^{(j)}$ are "within class covariances"

Fisher Linear Discrimination (cont.)

Major assumption: Class covariances are same (or "similar")

Good estimate of "common within class covariance"?

Show HDLSS/HDLSSod1FLD.ps

Pooled (weighted average) within class covariance:

$$\hat{\Sigma}^{w} = \frac{n_1 \hat{\Sigma}^{(1)} + n_2 \hat{\Sigma}^{(2)}}{n_1 + n_2} = \tilde{X}\tilde{X}^{t}$$

for the "full data matrix":

$$\tilde{X} = \frac{1}{\sqrt{n}} \left(\sqrt{n_1}\, \tilde{X}^{(1)} \; \sqrt{n_2}\, \tilde{X}^{(2)} \right)$$

Fisher Linear Discrimination (cont.)

Note: $\hat{\Sigma}^w$ is similar to $\hat{\Sigma}$ from before

- i.e. "covariance matrix ignoring class labels"
- important difference is "class by class centering"

Again show HDLSS/HDLSSod1FLD.ps

Fisher Linear Discrimination (cont.)

Simple way to find "correct covariance adjustment":

Individ'ly transform subpop'ns so "spherical" about their means

Show HDLSS\HDLSSod1egFLD.ps

$$\underrightarrow{Y}_i^{(j)} = \left(\hat{\Sigma}^w\right)^{-1/2} \underrightarrow{X}_i^{(j)}$$

then:

"best separating hyperplane"

is

"perpendicular bisector of line between means"

Fisher Linear Discrimination (cont.)

So in transformed space, the separating hyperlane has:

Transformed normal vector:

$$\underset{\rightarrow}{n}_{TFLD} = \left(\hat{\Sigma}^{w}\right)^{-1/2} \underset{\rightarrow}{\overline{X}}^{(1)} - \left(\hat{\Sigma}^{w}\right)^{-1/2} \underset{\rightarrow}{\overline{X}}^{(2)} = \left(\hat{\Sigma}^{w}\right)^{-1/2} \left(\underset{\rightarrow}{\overline{X}}^{(1)} - \underset{\rightarrow}{\overline{X}}^{(2)}\right)$$

Transformed intercept:

$$\underset{\rightarrow}{\mu}_{TFLD} = \frac{1}{2}\left(\hat{\Sigma}^{w}\right)^{-1/2} \underset{\rightarrow}{\overline{X}}^{(1)} + \frac{1}{2}\left(\hat{\Sigma}^{w}\right)^{-1/2} \underset{\rightarrow}{\overline{X}}^{(2)} = \left(\hat{\Sigma}^{w}\right)^{-1/2} \left(\frac{1}{2}\underset{\rightarrow}{\overline{X}}^{(1)} + \frac{1}{2}\underset{\rightarrow}{\overline{X}}^{(2)}\right)$$

Equation:

$$\left\{\underset{\rightarrow}{y} : \left\langle \underset{\rightarrow}{y}, \underset{\rightarrow}{n}_{TFLD} \right\rangle = \left\langle \underset{\rightarrow}{\mu}_{TFLD}, \underset{\rightarrow}{n}_{TFLD} \right\rangle \right\}$$

Again show HDLSS\HDLSSod1egFLD.ps

Fisher Linear Discrimination (cont.)

Thus discrimination rule is:

Given a new data vector $\underrightarrow{X}^0$, Choose Class 1 when:

$$\left\langle \left(\hat{\Sigma}^w\right)^{-1/2} \underrightarrow{X}^0, \underrightarrow{n}_{TFLD} \right\rangle \geq \left\langle \underrightarrow{\mu}_{TFLD}, \underrightarrow{n}_{TFLD} \right\rangle$$

i.e. (transforming back to original space)

$$\left\langle \underrightarrow{X}^0, \left(\hat{\Sigma}^w\right)^{-1/2} \underrightarrow{n}_{TFLD} \right\rangle \geq \left\langle \left(\hat{\Sigma}^w\right)^{1/2} \underrightarrow{\mu}_{TFLD}, \left(\hat{\Sigma}^w\right)^{-1/2} \underrightarrow{n}_{TFLD} \right\rangle$$

$$\left\langle \underrightarrow{X}^0, \underrightarrow{n}_{FLD} \right\rangle \geq \left\langle \underrightarrow{\mu}_{FLD}, \underrightarrow{n}_{FLD} \right\rangle$$

where:

$$\underrightarrow{n}_{FLD} = \left(\hat{\Sigma}^w\right)^{-1/2} \underrightarrow{n}_{TFLD} = \left(\hat{\Sigma}^w\right)^{-1} \left(\overline{\underrightarrow{X}}^{(1)} - \overline{\underrightarrow{X}}^{(2)} \right)$$

$$\underrightarrow{\mu}_{FLD} = \left(\hat{\Sigma}^w\right)^{1/2} \underrightarrow{\mu}_{TFLD} = \left(\frac{1}{2} \overline{\underrightarrow{X}}^{(1)} + \frac{1}{2} \overline{\underrightarrow{X}}^{(2)} \right)$$

Fisher Linear Discrimination (cont.)

Thus (in original space) have separating hyperplane with:

Normal vector: $\underrightarrow{n}_{FLD}$

Intercept: $\underrightarrow{\mu}_{FLD}$

Again show HDLSS\HDLSSod1egFLD.ps

FLD Likelihood View

Assume: Class distributions are multivariate $N\left(\underset{\rightarrow}{\mu}^{(j)}, \Sigma^{w}\right)$

(strong distributional assumption + common cov.)

At a location $\underset{\rightarrow}{x}^{0}$, the likelihood ratio,

for choosing between Class 1 and Class 2, is:

$$LR\left(\underset{\rightarrow}{x}^{0}, \underset{\rightarrow}{\mu}^{(1)}, \underset{\rightarrow}{\mu}^{(2)}, \Sigma^{w}\right) = \varphi_{\Sigma^{w}}\left(\underset{\rightarrow}{x}^{0} - \underset{\rightarrow}{\mu}^{(1)}\right) / \varphi_{\Sigma^{w}}\left(\underset{\rightarrow}{x}^{0} - \underset{\rightarrow}{\mu}^{(2)}\right)$$

where $\varphi_{\Sigma^{w}}$ is the Gaussian density with covariance Σ^{w}

FLD Likelihood View (cont.)

Simplifying, using the form of the Gaussian density:

$$\varphi_{\Sigma^w}(\underset{\rightarrow}{x}) = \frac{1}{(2\pi)^{d/2}\left|\Sigma^w\right|} e^{-\left(\underset{\rightarrow}{x}^t \Sigma^{w^{-1}} \underset{\rightarrow}{x}\right)/2}$$

Gives (critically using the common covariance):

$$LR\left(\underset{\rightarrow}{x}^0, \underset{\rightarrow}{\mu}^{(1)}, \underset{\rightarrow}{\mu}^{(2)}, \Sigma^w\right) = e^{-\left[\left(\underset{\rightarrow}{x}^0 - \underset{\rightarrow}{\mu}^{(1)}\right)^t \Sigma^{w^{-1}}\left(\underset{\rightarrow}{x}^0 - \underset{\rightarrow}{\mu}^{(1)}\right) - \left(\underset{\rightarrow}{x}^0 - \underset{\rightarrow}{\mu}^{(2)}\right)^t \Sigma^{w^{-1}}\left(\underset{\rightarrow}{x}^0 - \underset{\rightarrow}{\mu}^{(2)}\right)\right]/2}$$

$$-2\log LR\left(\underset{\rightarrow}{x}^0, \underset{\rightarrow}{\mu}^{(1)}, \underset{\rightarrow}{\mu}^{(2)}, \Sigma^w\right) =$$

$$= \left(\underset{\rightarrow}{x}^0 - \underset{\rightarrow}{\mu}^{(1)}\right)^t \Sigma^{w^{-1}}\left(\underset{\rightarrow}{x}^0 - \underset{\rightarrow}{\mu}^{(1)}\right) - \left(\underset{\rightarrow}{x}^0 - \underset{\rightarrow}{\mu}^{(2)}\right)^t \Sigma^{w^{-1}}\left(\underset{\rightarrow}{x}^0 - \underset{\rightarrow}{\mu}^{(2)}\right)$$

FLD Likelihood View (cont.)

But:

$$\left(\underset{\rightarrow}{x}^0 - \underset{\rightarrow}{\mu}^{(j)}\right)^t \Sigma^{w^{-1}} \left(\underset{\rightarrow}{x}^0 - \underset{\rightarrow}{\mu}^{(j)}\right) = \underset{\rightarrow}{x}^{0^t} \Sigma^{w^{-1}} \underset{\rightarrow}{x}^0 - 2\underset{\rightarrow}{x}^{0^t} \Sigma^{w^{-1}} \underset{\rightarrow}{\mu}^{(j)} + \underset{\rightarrow}{\mu}^{(j)} \Sigma^{w^{-1}} \underset{\rightarrow}{\mu}^{(j)}$$

so:

$$-2\log LR\left(\underset{\rightarrow}{x}^0, \underset{\rightarrow}{\mu}^{(1)}, \underset{\rightarrow}{\mu}^{(2)}, \Sigma^w\right) =$$

$$= -2\underset{\rightarrow}{x}^{0^t} \Sigma^{w^{-1}} \left(\underset{\rightarrow}{\mu}^{(1)} - \underset{\rightarrow}{\mu}^{(2)}\right) + \left(\underset{\rightarrow}{\mu}^{(1)} + \underset{\rightarrow}{\mu}^{(2)}\right) \Sigma^{w^{-1}} \left(\underset{\rightarrow}{\mu}^{(1)} - \underset{\rightarrow}{\mu}^{(2)}\right)$$

Thus $LR\left(\underset{\rightarrow}{x}^0, \underset{\rightarrow}{\mu}^{(1)}, \underset{\rightarrow}{\mu}^{(2)}, \Sigma^w\right) \geq 1$ when

$$-2\log LR\left(\underset{\rightarrow}{x}^0, \underset{\rightarrow}{\mu}^{(1)}, \underset{\rightarrow}{\mu}^{(2)}, \Sigma^w\right) \leq 0$$

i.e.

$$\underset{\rightarrow}{x}^{0^t} \Sigma^{w^{-1}} \left(\underset{\rightarrow}{\mu}^{(1)} - \underset{\rightarrow}{\mu}^{(2)}\right) \geq \frac{1}{2} \left(\underset{\rightarrow}{\mu}^{(1)} + \underset{\rightarrow}{\mu}^{(2)}\right) \Sigma^{w^{-1}} \left(\underset{\rightarrow}{\mu}^{(1)} - \underset{\rightarrow}{\mu}^{(2)}\right)$$

FLD Likelihood View (cont.)

Replacing $\underrightarrow{\mu}^{(1)}$, $\underrightarrow{\mu}^{(2)}$ and Σ^{w} by maximum likelihood estimates:

$$\underrightarrow{\overline{X}}^{(1)}, \quad \underrightarrow{\overline{X}}^{(2)} \quad \text{and} \quad \hat{\Sigma}^{w}$$

gives the likelihood ratio discrimination rule:

Choose Class 1, when

$$\underrightarrow{x}^{0^{t}} \hat{\Sigma}^{w^{-1}} \left(\underrightarrow{\overline{X}}^{(1)} - \underrightarrow{\overline{X}}^{(2)} \right) \leq \frac{1}{2} \left(\underrightarrow{\overline{X}}^{(1)} + \underrightarrow{\overline{X}}^{(2)} \right) \hat{\Sigma}^{w^{-1}} \left(\underrightarrow{\overline{X}}^{(1)} - \underrightarrow{\overline{X}}^{(2)} \right)$$

same as above

FLD Generalization I

Gaussian Likelihood Ratio Discrimination

(a. k. a. “nonlinear discriminant analysis”)

Idea: Assume class distributions are $N\left(\underrightarrow{\mu}^{(j)}, \Sigma^{(j)}\right)$

Different covariances!

Likelihood Ratio rule is straightforward calculation

(thus can easily do discrimination)

FLD Generalization I (cont.)

But no longer have "separating hyperplane" representation

(instead "regions determined by quadratics")

(fairly complicated case-wise calculations)

Graphical display: for each point, color as:

Yellow if assigned to Class 1

Cyan if assigned to Class 2

("intensity" is "strength of assignment")

show PolyEmbed\PEod1FLDe1.ps

FLD Generalization I (cont.)

Toy Examples:

1. Standard Tilted Point clouds:

also show PolyEmbed\PEod1GLRe1.ps

   - Both FLD and LR work well.

2. Donut:

Show PolyEmbed\PEdonFLDe1.ps & PEdonGLRe1.ps

   - FLD poor (no separating plane can work)
   - LR much better

FLD Generalization I (cont.)

3. Split X:

Show PolyEmbed\PExd3FLDe1.ps & PExd3GLRe1.ps

- neither works well
- although ∃ good separating surfaces
- they are not “from Gaussian likelihoods”
- so this is not “general quadratic discrimination”

FLD Generalization II

Different prior probabilities

Main idea: Give different weights to 2 classes

I.e. assume *not* a priori equally likely

Development is "straightforward"

- modifed likelihood
- change intercept in FLD

Might explore with toy examples, but time is short

FLD Generalization III

Principal Discriminant Analysis

Idea: FLD-like approach to more than two classes

Assumption: Class covariance matrices are the *same* (similar)

(but not Gaussian, as for FLD)

Main idea: quantify "location of classes" by their means

$$\underset{\longrightarrow}{\mu}^{(1)}, \ \underset{\longrightarrow}{\mu}^{(2)}, \ \ldots, \ \underset{\longrightarrow}{\mu}^{(k)}$$

FLD Generalization III (cont.)

Simple way to find "interesting directions" among the means:

PCA on set of means

i.e. Eigen-analysis of "between class covariance matrix"

$$\Sigma^B = MM^t$$

where

$$M = \frac{1}{\sqrt{k}}\left(\underset{\rightarrow}{\mu}^{(1)} - \underset{\rightarrow}{\overline{\mu}} \quad \cdots \quad \underset{\rightarrow}{\mu}^{(k)} - \underset{\rightarrow}{\overline{\mu}} \right)$$

Aside: can show: overall $\sqrt{n}\Sigma = \sqrt{n}\Sigma^B + \sqrt{k}\Sigma^w$

FLD Generalization III (cont.)

But PCA only works like “mean difference”,

Expect can improve by “taking covariance into account”.

Again show HDLSS\HDLSSod1egFLD.ps

Blind application of above ideas suggests eigen-analysis of:

$$\Sigma^{w^{-1}} \Sigma^{B}$$

FLD Generalization III (cont.)

There are:

- smarter ways to compute (“ generalized eigenvalue”)
- other representations (this solves optimization prob’s)

Special case: 2 classes,   reduces to standard FLD

Good reference for more: Section 3.8 of:

Duda, R. O., Hart, P. E. and Stork, D. G. (2001) *Pattern Classification*, Wiley.