Course Overview

Finished heuristic look at:

1. Understanding population structure – PCA
   - Toy examples
   - Cornea data - robustness

2. Discrimination (Classification)
   - Fisher Linear Discrimination
   - Corpus Callosum Data – Orthogonal subspace projection

Now take careful look at mathematics - numerics

Linear Algebra Review

Vector Space:

- set of "vectors", $\underset{\rightarrow}{x}$,
- and "scalars" (coefficients), a

- "closed" under "linear combination" ($\sum_i a_i \underset{\rightarrow}{x}_i$ in space)

- e.g. $\Re^d = \left\{ \underset{\rightarrow}{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix} : x_1, \ldots, x_d \in \Re^d \right\}$, "$d$ dim Euclid'n space"

Linear Algebra Review, (cont.)

Subspace:

- subset that is again a vector space
- i.e. closed under linear combination
- e.g. lines through the origin
- e.g. planes through the origin
- e.g. subspace “generated by” a set of vectors
  (all linear combos of them = containing hyperplane)

Linear Algebra Review, (cont.)

Basis of subspace: set of vectors that:

- "span", i.e. everything is a linear combo of them
- are "linearly independent", i.e. linear combo is unique
- e.g. $\Re^d$ "unit vector basis" $\left\{ \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \right\}$
- e.g. $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{pmatrix} = x_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \cdots + x_d \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$

Linear Algebra Review, (cont.)

Basis Matrix, of subspace of $\Re^d$

Given a basis, $\underset{\rightarrow}{v_1},\ldots,\underset{\rightarrow}{v_n}$, create "matrix of columns":

$$B = \begin{pmatrix} \underset{\rightarrow}{v_1} & \cdots & \underset{\rightarrow}{v_n} \end{pmatrix} = \begin{pmatrix} v_{11} & & v_{n1} \\ \vdots & \cdots & \vdots \\ v_{1d} & & v_{nd} \end{pmatrix}_{d \times n}$$

Then "linear combo" is a matrix multiplication:

$$\sum_{i=1}^{n} a_i \underset{\rightarrow}{v_i} = B\underset{\rightarrow}{a} \quad \text{where} \quad \underset{\rightarrow}{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$$

Often useful to check sizes: $d \times 1 = (d \times n) \leftrightarrow (n \times 1)$

Linear Algebra Review, (cont.)

Dimension of subspace (a notion of “size”):

- number of elements in a basis (unique)
- $\dim\left(\Re^{d}\right) = d$ (use basis above)
- e.g.   dim of a line is 1
- e.g.   dim of a plane is 2
- dimension is “degrees of freedom”

Linear Algebra Review, (cont.)

Norm of a vector:

- in $\Re^d$, $\quad \|\underset{\rightarrow}{x}\| = \left(\sum_{j=1}^{d} x_j^2 \right)^{1/2} = \left(\underset{\rightarrow}{x}^t \underset{\rightarrow}{x} \right)^{1/2}$

- Idea: “length” of the vector

- But recall strange properties for high d,
  e.g. “length of diagonal of unit cube” = $\sqrt{d}$

- “length normalized vector”: $\dfrac{\underset{\rightarrow}{x}}{\|\underset{\rightarrow}{x}\|}$
  (has length one, this is on surface of unit sphere)

- get “distance” as: $d\left(\underset{\rightarrow}{x}, \underset{\rightarrow}{y}\right) = \|\underset{\rightarrow}{x} - \underset{\rightarrow}{y}\| = \sqrt{\left(\underset{\rightarrow}{x} - \underset{\rightarrow}{y}\right)^t \left(\underset{\rightarrow}{x} - \underset{\rightarrow}{y}\right)}$

Linear Algebra Review, (cont.)

Inner (dot, scalar) product:

- for vectors $\underset{\rightarrow}{x}$ and $\underset{\rightarrow}{y}$, $\left\langle \underset{\rightarrow}{x}, \underset{\rightarrow}{y} \right\rangle = \sum_{j=1}^{d} x_j y_j = \underset{\rightarrow}{x}^t \underset{\rightarrow}{y}$

- related to norm, via $\left\| \underset{\rightarrow}{x} \right\| = \sqrt{\left\langle \underset{\rightarrow}{x}, \underset{\rightarrow}{x} \right\rangle} = \sqrt{\underset{\rightarrow}{x}^t \underset{\rightarrow}{x}}$

- measures "angle between $\underset{\rightarrow}{x}$ and $\underset{\rightarrow}{y}$" as:

$$angle\left(\underset{\rightarrow}{x}, \underset{\rightarrow}{y}\right) = \cos^{-1}\left(\frac{\left\langle \underset{\rightarrow}{x}, \underset{\rightarrow}{y} \right\rangle}{\left\| \underset{\rightarrow}{x} \right\| \cdot \left\| \underset{\rightarrow}{y} \right\|} \right) = \cos^{-1}\left(\frac{\underset{\rightarrow}{x}^t \underset{\rightarrow}{y}}{\sqrt{\underset{\rightarrow}{x}^t \underset{\rightarrow}{x} \cdot \underset{\rightarrow}{y}^t \underset{\rightarrow}{y}}} \right)$$

- key to "orthogonality", i.e. "perpendicularity":

$\underset{\rightarrow}{x} \perp \underset{\rightarrow}{y}$ if and only if $\left\langle \underset{\rightarrow}{x}, \underset{\rightarrow}{y} \right\rangle = 0$

Linear Algebra Review, (cont.)

Orthonormal basis $\underrightarrow{v_1}, ..., \underrightarrow{v_n}$:

- All ortho to each other, i.e. $\langle \underrightarrow{v_i}, \underrightarrow{v_{i'}} \rangle = 0$, for $i \neq i'$

- All have length 1, i.e. $\langle \underrightarrow{v_i}, \underrightarrow{v_i} \rangle = 1$, for $i = 1, ..., n$

- "Spectral Representation": $\underrightarrow{x} = \sum_{i=1}^{n} a_i \underrightarrow{v_i}$ where $a_i = \langle \underrightarrow{x}, \underrightarrow{v_i} \rangle$

  check: $\langle \underrightarrow{x}, \underrightarrow{v_i} \rangle = \left\langle \sum_{i'=1}^{n} a_{i'} \underrightarrow{v_{i'}}, \underrightarrow{v_i} \right\rangle = \sum_{i'=1}^{n} a_{i'} \langle \underrightarrow{v_{i'}}, \underrightarrow{v_i} \rangle = a_i$

- Matrix notation: $\underrightarrow{x} = B\underrightarrow{a}$ where $\underrightarrow{a}^t = \underrightarrow{x}^t B$ i.e. $\underrightarrow{a} = B^t \underrightarrow{x}$

- $\underrightarrow{a}$ is called "transform (e.g. Fourier, wavelet) of $\underrightarrow{x}$"

Linear Algebra Review, (cont.)

Parseval identity, for $\underset{\rightarrow}{x}$ in subsp. gen'd by o. n. basis $\underset{\rightarrow}{v_1}, \ldots, \underset{\rightarrow}{v_n}$:

- $\left\|\underset{\rightarrow}{x}\right\|^2 = \sum_{i=1}^{n} \left\langle \underset{\rightarrow}{x}, \underset{\rightarrow}{v_i} \right\rangle^2 = \sum_{i=1}^{n} a_i^2 = \left\|\underset{\rightarrow}{a}\right\|^2$

- Pythagorean theorem

- "Decomposition of Energy"

- ANOVA - sums of squares

- Transform, $\underset{\rightarrow}{a}$, has same length as $\underset{\rightarrow}{x}$, i.e. "rotation in $\Re^d$"

Linear Algebra Review, (cont.)

Projection of a vector $\underset{\rightarrow}{x}$ onto a subspace V:

- Idea: member of V that is closest to $\underset{\rightarrow}{x}$ (i.e. “approx’n”)
- Find $P_V \underset{\rightarrow}{x} \in V$ that solves: $\min_{v \in V} \left\| \underset{\rightarrow}{x} - \underset{\rightarrow}{v} \right\|$ (“least squares”)
- For inner product (Hilbert) space: exists and is unique
- General solution in $\Re^d$: for basis matrix B_V

$$P_V \underset{\rightarrow}{x} = B_V \left(B_V^t B_V \right)^{-1} B_V^t \underset{\rightarrow}{x}$$

- So “proj’n operator” is “matrix mult’n”: $P_V = B_V \left(B_V^t B_V \right)^{-1} B_V^t$

  (thus projection is another linear operation)

  (note same operation underlies “least squares”)

Linear Algebra Review, (cont.)

Projection using orthonormal basis $\underset{\rightarrow}{v_1},\ldots,\underset{\rightarrow}{v_n}$:

- Basis matrix is "orthonormal": $B_V^t B_V = I_{n \times n}$

$$\begin{pmatrix} \underset{\rightarrow}{v_1}^t \\ \vdots \\ \underset{\rightarrow}{v_n}^t \end{pmatrix} \begin{pmatrix} \underset{\rightarrow}{v_1} & \cdots & \underset{\rightarrow}{v_n} \end{pmatrix} = \begin{pmatrix} \langle \underset{\rightarrow}{v_1}, \underset{\rightarrow}{v_1} \rangle & \cdots & \langle \underset{\rightarrow}{v_1}, \underset{\rightarrow}{v_n} \rangle \\ \vdots & \ddots & \vdots \\ \langle \underset{\rightarrow}{v_n}, \underset{\rightarrow}{v_1} \rangle & \cdots & \langle \underset{\rightarrow}{v_n}, \underset{\rightarrow}{v_n} \rangle \end{pmatrix} = \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{pmatrix}$$

- So $P_V \underset{\rightarrow}{x} = B_V \left(B_V^t \underset{\rightarrow}{x} \right)$ = Recon(Coeffs of $\underset{\rightarrow}{x}$ "in V dir'n")

- For "orthogonal complement", $V^{\perp}$,

$$\underset{\rightarrow}{x} = P_V \underset{\rightarrow}{x} + P_{V^{\perp}} \underset{\rightarrow}{x} \quad \text{and} \quad \left\| \underset{\rightarrow}{x} \right\|^2 = \left\| P_V \underset{\rightarrow}{x} \right\|^2 + \left\| P_{V^{\perp}} \underset{\rightarrow}{x} \right\|^2$$

- Parseval inequality: $\left\| \underset{\rightarrow}{x} \right\|^2 \leq \left\| P_V \underset{\rightarrow}{x} \right\|^2 = \sum_{i=1}^{n} \left\langle \underset{\rightarrow}{x}, \underset{\rightarrow}{v_i} \right\rangle^2 = \sum_{i=1}^{n} a_i^2 = \left\| \underset{\rightarrow}{a} \right\|^2$

Linear Algebra Review, (cont.)

Eigenvalue Decomposition:

For a (symmetric) square matrix $X_{d \times d}$

Find a diagonal matrix $D = \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_d \end{pmatrix}$

And an orthonormal matrix $B_{d \times d}$ (i.e. $B^t \cdot B = B \cdot B^t = I_{d \times d}$)

So that: $X \cdot B = B \cdot D$, i.e. $X = B \cdot D \cdot B^t$

Linear Algebra Review, (cont.)

Intuition behind Eigenvalue Decomposition:

For X a "linear transformation" (via matrix multiplication)

- $X \cdot \underset{\rightarrow}{v} = \left(B \cdot D \cdot B^t\right) \cdot \underset{\rightarrow}{v} = B \cdot \left(D \cdot \left(B^t \cdot \underset{\rightarrow}{v}\right)\right)$
- First "rotate"
- Second "rescale coordinate axes (by λs)
- Third "invert rotation"

For X a basis matrix of $\Re^d$, B gives "rotation to make parallel to coordinate axes"

Linear Algebra Review, (cont.)

Computation of Eigenvalue Decomposition:

- Details too complex to spend time here
- A “primitive” of good software packages
- Eigenvalues $\lambda_1, \ldots, \lambda_d$ are unique
- Columns of $B = \begin{pmatrix} \underrightarrow{v_1} & \cdots & v_d \end{pmatrix}$ are called “eigenvectors”
- Eigenvectors are “λ-stretched” by X:

$$X \cdot \underrightarrow{v_i} = \lambda_i \cdot \underrightarrow{v_i}$$

Linear Algebra Review, (cont.)

Eigenvalue Decomposition solves matrix problems:

- Inversion: $$X^{-1} = B \cdot \begin{pmatrix} \lambda_1^{-1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_d^{-1} \end{pmatrix} \cdot B^t$$

- Square Root: $$X^{1/2} = B \cdot \begin{pmatrix} \lambda_1^{1/2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_d^{1/2} \end{pmatrix} \cdot B^t$$

- $rank(X) = \#\{\lambda_i : \lambda_i \neq 0\}$

- X is positive (semi) definite $\Leftrightarrow$ all $\lambda_i > (\geq) 0$

Multivariate Probability Review

Given a "random vector", $\underset{\rightarrow}{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$,

A "center" of the dist'n is the mean vector, $\underset{\rightarrow}{\mu} = E\underset{\rightarrow}{X} = \begin{pmatrix} EX_1 \\ \vdots \\ EX_n \end{pmatrix}$

A "measure of spread" is the covariance matrix:

$$\Sigma = \operatorname{cov}(X) = \begin{pmatrix} \operatorname{var}(X_1) & \cdots & \operatorname{cov}(X_1, X_n) \\ \vdots & \ddots & \vdots \\ \operatorname{cov}(X_n, X_1) & \cdots & \operatorname{var}(X_n) \end{pmatrix}$$

Multivariate Probability Review, (cont.)

Covariance matrix:

- Nonegative Definite (since all variances are ≥ 0)
- Provides "elliptical summary of distribution"
- Calculated via "outer product":

$$\Sigma = \mathrm{cov}(X) = E\begin{pmatrix} (X_1-\mu_1)(X_1-\mu_1) & \cdots & (X_1-\mu_1)(X_n-\mu_n) \\ \vdots & \ddots & \vdots \\ (X_n-\mu_n)(X_1-\mu_1) & \cdots & (X_n-\mu_n)(X_n-\mu_n) \end{pmatrix} =$$

$$\Sigma = E\left(\underset{\rightarrow}{X} - \underset{\rightarrow}{\mu}\right)\left(\underset{\rightarrow}{X} - \underset{\rightarrow}{\mu}\right)^t$$

Multivariate Probability Review, (cont.)

Empirical versions:

Given a "random sample" $\underset{\rightarrow}{X_1},\ldots,\underset{\rightarrow}{X_n}$,

Estimate the "theoretical mean" $\underset{\rightarrow}{\mu}$, with the "sample mean":

$$\underset{\rightarrow}{\hat{\mu}} = \underset{\rightarrow}{\overline{X}} = \begin{pmatrix} \overline{X}_1 \\ \vdots \\ \overline{X}_d \end{pmatrix} = \frac{1}{n}\sum_{i=1}^{n} \underset{\rightarrow}{X_i}$$