04/25/07 1

Dimension Reduction III

PCA: preserve variability in the lower dimensional space.

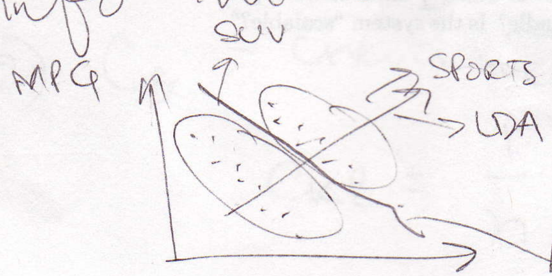
1) Rescale

2) Covariance Matrix C → ~~components~~ computes eigenvectors $u^{(1)} \ldots u^{(k)}$ corresponding to the largest eigenvalues.

3) $P = \left(u^{(1)} \ldots u^{(k)} \right) \quad y^{(i)} = P^T x^{(i)}$

Linear Discriminant Analysis

Goal: Given data w/ labels enhance discriminant info in data in lower dimensional space.



If we were to do PCA on this data, it would create an axis that separates the pts. the most → like this. But if you project the pts on this new line, the pts. will be meshed. So, in a way, you have to create an almost orthogonal axis.

Approach:

1) Minimize var. within classes. Squish pts. together

2) Maximize var. b/w classes (overall variance).

---

1) Rescale to zero mean & unit var. while disregarding labels.

class 1 class 2

$$X = \underbrace{\begin{bmatrix} x^{(1)} & \cdots & x^{(i)} & \cdots & x^{(n)} \end{bmatrix}}_{n \text{ pts.}} \Big\} p \text{ attrib.}$$

2) C_B - covariance matrix of all data pts.

$$C_{kl} = \frac{1}{n} \sum_{i=1}^{n} x_k^{(i)} x_l^{(i)}.$$

$$C_B = \frac{1}{n} X X^T.$$

$\rightarrow C_W^{(l)}$ — covariance matrix of datapts
>within.
with label l. You take X cut out all pts except those w/ label 'l' & calculate covariance matrix in exactly the same way as above.

Now, find a matrix combining all $C_W^{(l)}$ by doing a weighted avg. of $C_W^{(l)}$'s.

$$C_W = \frac{1}{n} \sum_{l=1}^{\#\text{labels}} n_l C_W^{(l)}$$ where n_l = # of pts. w/ label 'l'.

So now, we need to max. C_B & min. C_W.

For former, we apply PCA $\Rightarrow$ eig. vector of C_B maximizes variance. $\max_{\|u\|=1} u^T C_B u$. — (1)

For latter: $\text{minimize}_{\|u\|=1}\ u^T C_W u = \max_{\|u\|=1} \frac{1}{u^T C_W u}$ — (2)

can do that $\because$ C_W has +ive element.

Combining (1) & (2): $\max_{\|u\|=1} \frac{u^T C_B u}{u^T C_W u}$

$\Rightarrow$ eigenvectors of $C_W^{-1} C_B$ corresponding to k largest eigenvalues for k-dimensional solving the LDA problem

$\Rightarrow P = \left(u^{(1)} \dots u^{(k)}\right)$ $\boxed{y^{(i)} = P^T x^{(i)}}$

Latent Semantic Indexing

Focuses on NLP & IR. Given a corpus, we'd like to answer queries.

- vectors = documents $\approx 10^6$ } huge.
- basis = individual words $\approx 25\times 10^3$ }

$$A = \Big\{ \underbrace{a^{(1)} \;\cdots\cdots\; a^{(n)}}_{n} \Big\} \} p$$

$a_j^{(i)}$ = # of times j^{th} word appears in i^{th} do

q = query & "small document" $\in \mathbb{R}^p$

1) First attempt is to use inner prod. to measure similarity b/w query & documents.

$$\max_i \frac{q^T a^{(i)}}{\|a^{(i)}\|}$$