PCA as an optimization problem

Find "direction of greatest variability" [toy graphic]

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 $\underline{u}$:

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

i.e. (proportional 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 $\underline{u}$:

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

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

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

and the “diagonal quadratic form” becomes

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

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

  [Toy Graphic]
- 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

Go to next part