PCA as a probabilistic model

Probabilistic model for PCA

- $\mathbf{x} \sim N(\mathbf{0},\mathbf{I})$ latent variables (dim k)
- $\mathbf{y} = \mathbf{xC} + \varepsilon$ observed variables (dim $p > k$)
- $\varepsilon \sim N(\mathbf{0},\delta\mathbf{I})$ noise (maybe small)

C is *k*p*

How does this compare to a GMM?

- **Thus**
  - $\mathbf{y} \sim N(\mathbf{0}, \mathbf{C}^T\mathbf{C} + \delta I)$
- **Take the limit where the noise is goes to zero**
  - Then $p(\mathbf{y}|\mathbf{x})$ collapses with all probability mass on a single point $\mathbf{y} = \mathbf{xC}$
  - And $p(\mathbf{x}|\mathbf{y})$ collapses with all probability mass on a single point $\mathbf{x} = \mathbf{yC}^T(\mathbf{CC}^T)^{-1}$

Penn
University of Pennsylvania

EM Algorithm for PCA

- **E-step** find "hidden" **X** given parameters **C**
  - $\mathbf{X} = \mathbf{Y}\,\mathbf{C}^T (\mathbf{C}\mathbf{C}^T)^{-1}$
- **M-step** find parameters **C** given "hidden" **X**)
  - $\mathbf{C} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\,\mathbf{Y}$ (the standard MLE weight estimates)

  for $\mathbf{y} = \mathbf{x}\mathbf{C} + \varepsilon$

What are the PCs?

A. X

B. Y

C. C

What are the loadings?

A. X

B. Y

C. C

How might this model minimize L_2 reconstruction error?

Probabilistic model for CCA

- $\mathbf{x} \sim N(\mathbf{0},\mathbf{I})$ latent variables (dim *k*)
- $\mathbf{y} = \mathbf{xB} + \varepsilon$ observed variables (dim *p* > *k*)
- $\mathbf{z} = \mathbf{xC} + \varepsilon$ observed variables (dim *q* > *k*)

  B is *k*p* B is *k*q*

- $\varepsilon \sim N(\mathbf{0},\delta\mathbf{I})$ noise (maybe small)

- **Thus**
  - $\mathbf{y} \sim N(\mathbf{0}, \mathbf{B}^T\mathbf{B} + \delta\mathbf{I})$
  - $\mathbf{z} \sim N(\mathbf{0}, \mathbf{C}^T\mathbf{C} + \delta\mathbf{I})$

- **Take the limit where the noise is goes to zero**
  - …