Probabilistic models for neural data: From single neurons to population dynamics

NEUROBIO 316QC

Jan Drugowitsch

jan_drugowitsch@hms.harvard.edu

Session 5: Dimensionality reduction

Today

Q&A about previous session

Discussing assignment (15-25min)

Paper discussion (~1h)

Dimensionality reduction (remaining time)

Overview

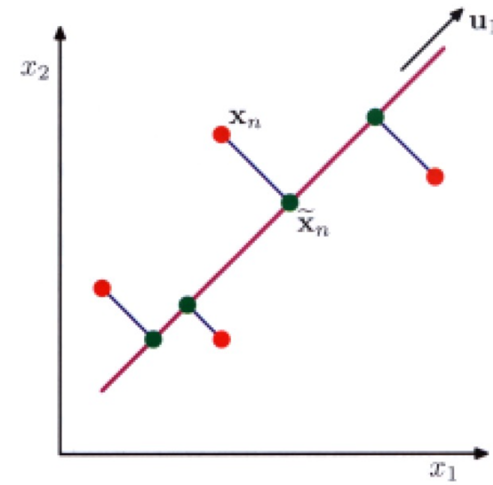
(Linear) Dimensionality reduction

- Linear dimensionality reduction in general
- Two perspectives on PCA
- Probabilistic PCA
- Factor analysis, and relationship to probabilistic PCA

Dimensionality reduction

Assumes Data $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$ lies on lower-dimensional manifold $\boldsymbol{u}_1$

Deviations $\boldsymbol{x}_n - \tilde{\boldsymbol{x}}_n$ from this manifold: noise



PRLM Fig. 12.2

Example high-d population activity might linearly encode small number of latent variables

Population (linear) dimensionality reduction methods

Principal component analysis

Linear mapping from latent space + isotropic noise

Factor analysis

Linear mapping from latent space + independent (but not isotropic) noise

Projections

Project $\boldsymbol{x}$ onto unit vector $\boldsymbol{u}$ ($\|\boldsymbol{u}\| = 1$)

1. Angle α between $\boldsymbol{x}$ and u

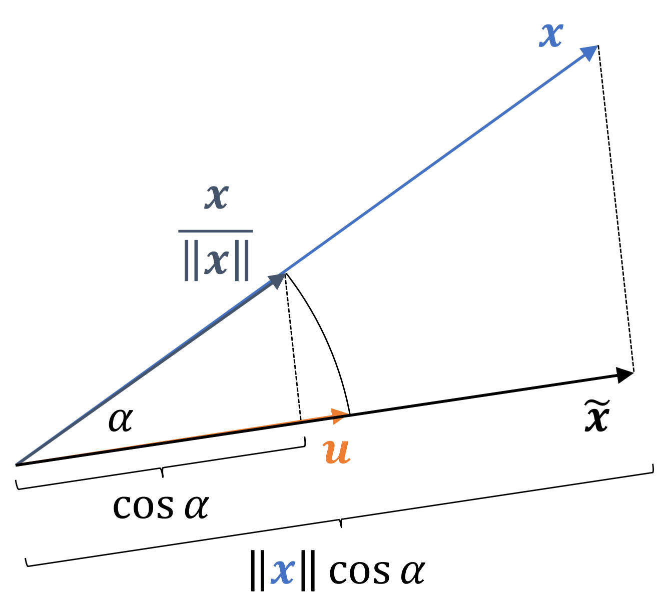
$$\cos\alpha = \frac{\boldsymbol{u}^T\boldsymbol{x}}{\|\boldsymbol{x}\|} = \boldsymbol{u}^T\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}$$

2. Length $\|\tilde{\boldsymbol{x}}\|$ of projected $\boldsymbol{x}$

$$\|\tilde{\boldsymbol{x}}\| = \|\boldsymbol{x}\|\cos\alpha = \boldsymbol{u}^T\boldsymbol{x}$$

3. Scaling $\boldsymbol{u}$ by this length

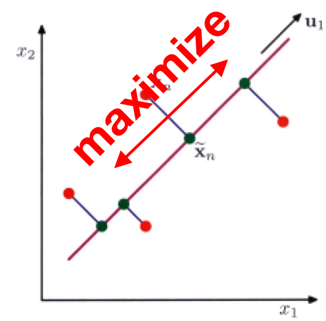
$$\tilde{\boldsymbol{x}} = \|\tilde{\boldsymbol{x}}\|\boldsymbol{u} = (\boldsymbol{x}^T\boldsymbol{u})\boldsymbol{u}$$



Project $\boldsymbol{x}$ into orthonormal basis $\boldsymbol{u}_{1:M}$

$$\tilde{\boldsymbol{x}} = \sum_{j=1}^{M}\underbrace{\left(\boldsymbol{x}^T\boldsymbol{u}_j\right)}_{\text{length of projection onto } \boldsymbol{u}_j}\boldsymbol{u}_j$$

Principal component analysis (PCA)



Variance maximization perspective

Find projection $\boldsymbol{u}_1^T \boldsymbol{x}_n$ (with $\|\boldsymbol{u}_1\| = 1$) for all $\boldsymbol{x}_{1:N}$ that captures most variance

Leads to

$$\boldsymbol{S}\boldsymbol{u}_1 = \lambda_1 \boldsymbol{u}_1 \qquad \text{with variance of projection, } \lambda_1 = \boldsymbol{u}_1^T \boldsymbol{S} \boldsymbol{u}_1$$

data covariance

Thus: variance-maximizing projection is eigenvector $\boldsymbol{u}_1$ of $\boldsymbol{S}$ for largest eigenvalue λ_1

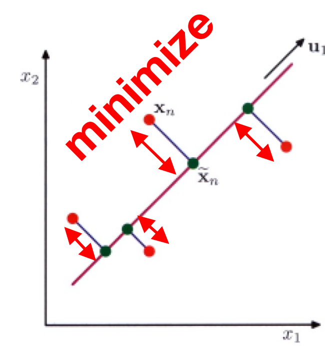
Remove $\boldsymbol{u}_1^T \boldsymbol{x}_n$ and repeat: $\boldsymbol{u}_j$ is eigenvector associated with jth largest eigenvalue λ_j

Recall spectral covariance decomposition

columns = $\boldsymbol{\Sigma}_x$ eigenvectors — diagonal = (positive) $\boldsymbol{\Sigma}_x$ eigenvalues

$$\boldsymbol{\Sigma}_x = \boldsymbol{R}\boldsymbol{D}^2\boldsymbol{R}^T$$

principal axes $\boldsymbol{u}_j$ (rotation) — variance λ_j along principal axes (scaling)



Minimum error perspective

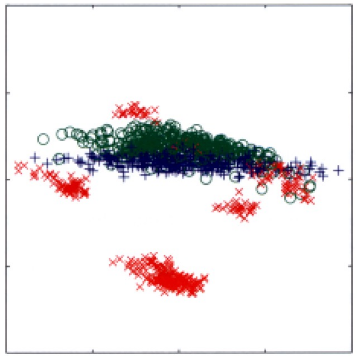
Assume orthonormal basis $\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_M$ ($\mathrm{M} \le D$, where M is dimensionality of $\boldsymbol{x}_n$)

Find basis that minimizes error $\sum_n \|\boldsymbol{x}_n - \tilde{\boldsymbol{x}}_n\|^2$, where $\tilde{\boldsymbol{x}}_n$ is $\boldsymbol{x}_n$ projected into this basis

Leads to the same eigenvalue problem as above

Some applications of PCA

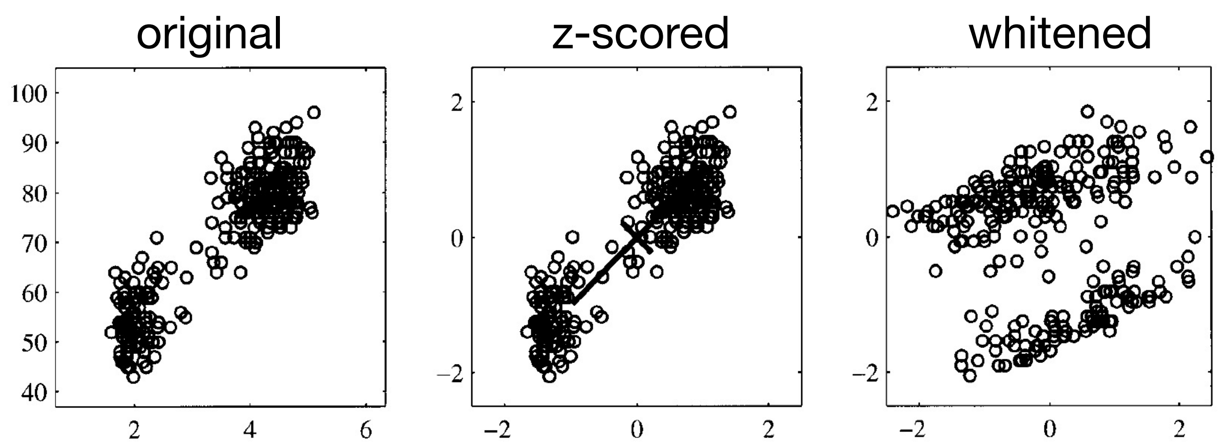
Visualization Plot high-d data along first two/three principal components



oil flow data
colors = flow type

PRML, Fig. 12.8

Pre-processing Whitening/sphering the data to achieve zero mean unit *co*variance data (rather than just zero mean unit variance along each dimension)



PRML, Fig. 12.6

Compression Maintain only the first $M < D$ data dimensions

Variance 'explained' by first M dimensions: $\sum_{m=1}^{M} \lambda_m$

Probabilistic PCA (PPCA)

Generative model: mapping from low-d latent space $\boldsymbol{z}$ to high-d space of observables $\boldsymbol{x}$

latent space $\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots$

space of observables, $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$

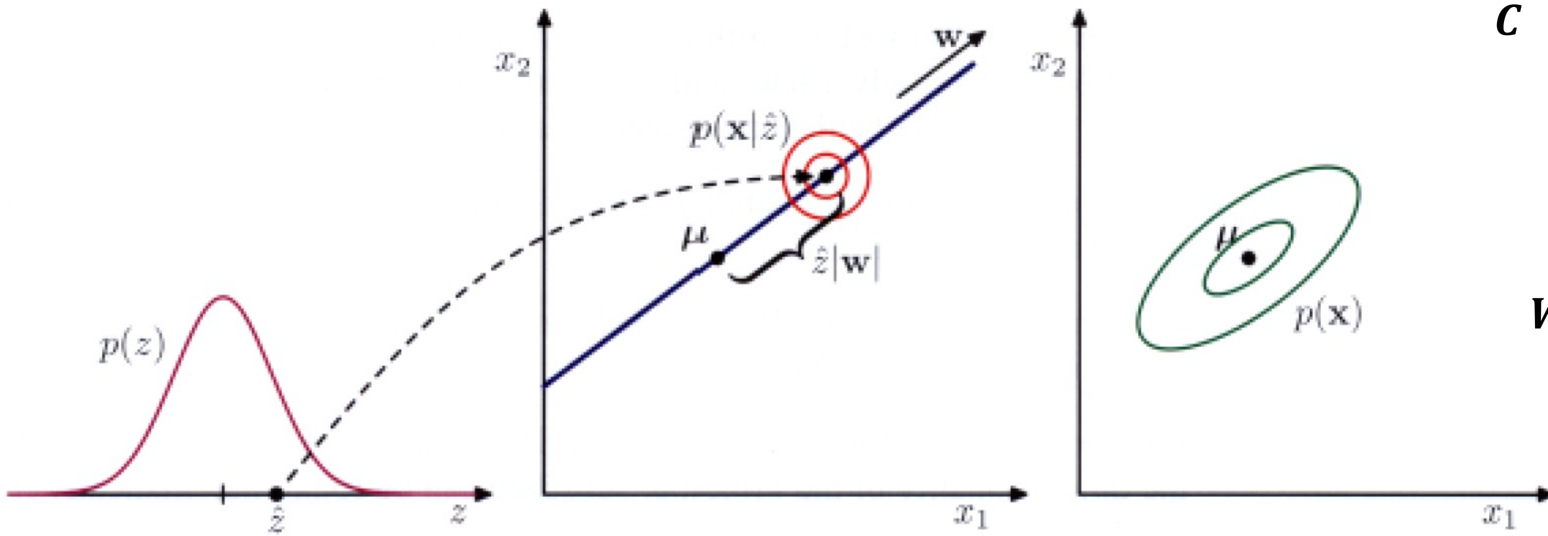
$p(\boldsymbol{z}) = \mathrm{N}(\boldsymbol{z}|\boldsymbol{0}, \boldsymbol{I})$

M-dimensional

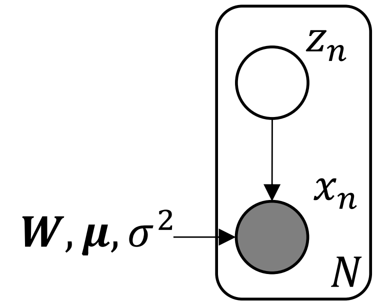
$p(\boldsymbol{x}|\boldsymbol{z}) = \mathrm{N}(\boldsymbol{x}|\boldsymbol{W}\boldsymbol{z} + \boldsymbol{\mu}, \sigma^2 \boldsymbol{I})$

D-dimensional

$$p(\boldsymbol{x}) = \int p(\boldsymbol{x}|\boldsymbol{z}) p(\boldsymbol{z}) d\boldsymbol{z} = \mathrm{N}(\boldsymbol{x}|\boldsymbol{\mu}, \underbrace{\sigma^2 \boldsymbol{I} + \boldsymbol{W}\boldsymbol{W}^T}_{\boldsymbol{C}})$$



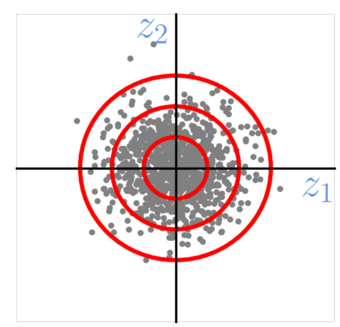
PRML, Fig. 12.9



Latent space rotation invariance ($\boldsymbol{R}$ = rotation/orthonormal matrix)

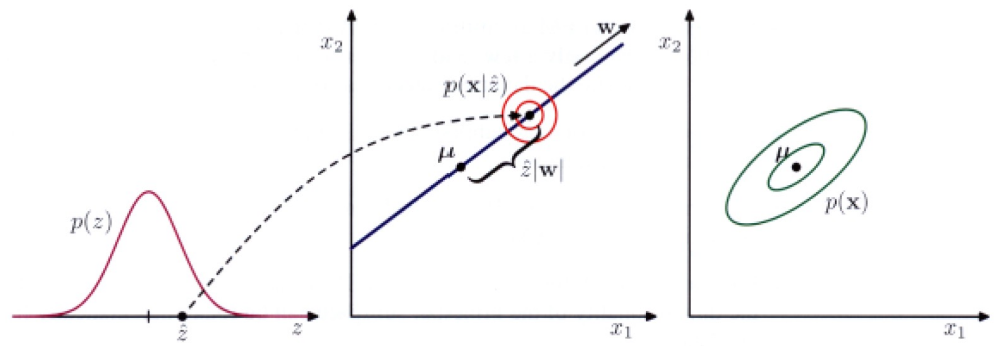
$$p(\boldsymbol{R}\boldsymbol{z}) = \mathrm{N}(\boldsymbol{z}|\boldsymbol{0}, \boldsymbol{R}\boldsymbol{R}^T) = \mathrm{N}(\boldsymbol{z}|\boldsymbol{0}, \boldsymbol{I}) = p(\boldsymbol{z})$$

$$p(\boldsymbol{x}|\boldsymbol{R}\boldsymbol{z}) = \mathrm{N}(\boldsymbol{x}|\boldsymbol{W}\boldsymbol{R}\boldsymbol{z} + \boldsymbol{\mu}, \sigma^2 \boldsymbol{I}) = \mathrm{N}(\boldsymbol{x}|\underbrace{\widetilde{\boldsymbol{W}}\boldsymbol{z}}_{} + \boldsymbol{\mu}, \sigma^2 \boldsymbol{I})$$



$\boldsymbol{W}$ can only be determined up to some rotation $\boldsymbol{R}$,
rotation does not impact $\boldsymbol{C}$ in $p(\boldsymbol{x})$

Inference in PPCA



Maximum likelihood solution

$$\widehat{\boldsymbol{\mu}}_{ML} = \overline{\boldsymbol{x}} = \frac{1}{N}\sum\nolimits_n \boldsymbol{x}_n$$

$$\widehat{\boldsymbol{W}}_{ML} = \boldsymbol{U}_M(\boldsymbol{\Lambda}_M - \sigma^2\boldsymbol{I})^{1/2}\boldsymbol{R}$$

eigenvectors $\boldsymbol{u}_{1:M}$

arbitrary rotation

largest eigenvalues $\lambda_{1:M}$ of $\boldsymbol{S}$
along diagonal

$$\hat{\sigma}^2_{ML} = \frac{1}{D-M}\sum_{i=M+1}^{D}\lambda_i$$

captures the deviation from principal axes
by average variance of remaining dimensions

Alternative: EM-algorithm (more efficient for larger datasets)

Mapping from observable x_n to latent z_n

$$p(\boldsymbol{z}_n|\boldsymbol{x}_n,\boldsymbol{\theta}) = \mathrm{N}(\boldsymbol{z}_n|\boldsymbol{M}^{-1}\boldsymbol{W}(\boldsymbol{x}-\boldsymbol{\mu}),\sigma^{-2}\boldsymbol{M})$$

$$\boldsymbol{M} = \boldsymbol{W}^T\boldsymbol{W} + \sigma^2\boldsymbol{I}$$

In the limit $\sigma^2 \to 0$: $E[\boldsymbol{z}_n|\boldsymbol{x}_n] \to \left(\widehat{\boldsymbol{W}}^T_{ML}\widehat{\boldsymbol{W}}_{ML}\right)^{-1}\widehat{\boldsymbol{W}}_{ML}(\boldsymbol{x}-\overline{\boldsymbol{x}})$ (standard PCA)

Variants: missing data, Bayesian inference with ARD

Why PPCA?

Reveals the generative model assumptions underlying PCA

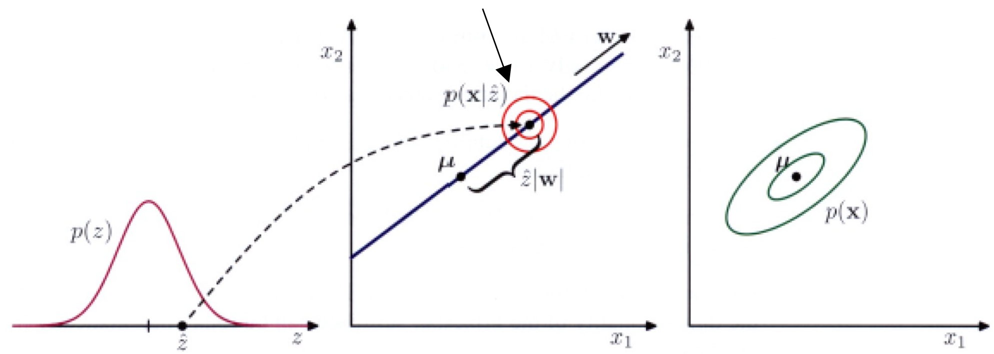
Supports efficient EM-style inference algorithm

Supports Bayesian treatment; use of ARD to determine latent space dimensionality M

Can be used to generate samples from $p(\boldsymbol{x})$, $p(\boldsymbol{z})$, $p(\boldsymbol{x}|\boldsymbol{z})$, $p(\boldsymbol{z}|\boldsymbol{x})$

Factor analysis (FA)

PPCA model



latent space prior — generative model — marginal likelihood

	Probabilistic PCA	Factor analysis
latent space prior	$p(\boldsymbol{z}) = \mathrm{N}(\boldsymbol{z}\|\boldsymbol{0}, \boldsymbol{I})$	$p(\boldsymbol{z}) = \mathrm{N}(\boldsymbol{z}\|\boldsymbol{0}, \boldsymbol{I})$ diagonal
generative model	$p(\boldsymbol{x}\|\boldsymbol{z}) = \mathrm{N}(\boldsymbol{x}\|\boldsymbol{W}\boldsymbol{z} + \boldsymbol{\mu}, \sigma^2 \boldsymbol{I})$	$p(\boldsymbol{x}\|\boldsymbol{z}) = \mathrm{N}(\boldsymbol{x}\|\boldsymbol{W}\boldsymbol{z} + \boldsymbol{\mu}, \boldsymbol{\Psi})$
marginal likelihood	$p(\boldsymbol{x}) = \mathrm{N}(\boldsymbol{x}\|\boldsymbol{\mu}, \sigma^2 \boldsymbol{I} + \boldsymbol{W}\boldsymbol{W}^T)$	$p(\boldsymbol{x}) = \mathrm{N}(\boldsymbol{x}\|\boldsymbol{\mu}, \boldsymbol{\Psi} + \boldsymbol{W}\boldsymbol{W}^T)$

Remains invariant to rotations $\boldsymbol{R}$ in latent space; $\boldsymbol{R}\boldsymbol{z} \rightarrow \boldsymbol{W}\boldsymbol{R}\boldsymbol{z} = \widetilde{\boldsymbol{W}}\boldsymbol{z}$

Separate roles of W and Ψ

- $\boldsymbol{W}$ contributes to marginal likelihood through $\boldsymbol{W}\boldsymbol{W}^T$;
  captures correlations in $\boldsymbol{x}$ components
- $\boldsymbol{\Psi}$ contributes to marginal likelihood as diagonal matrix;
  captures (different) independent variances in $\boldsymbol{x}$ components

Maximum likelihood inference: $\widehat{\boldsymbol{W}}_{ML}$ no longer close-form, requires EM algorithm

Overview

(Linear) Dimensionality reduction

Linear dimensionality reduction in general & subspace projections

Two perspectives on PCA

Probabilistic PCA

Factor analysis, and relationship to probabilistic PCA

Summary

PCA as variance maximizing or error minimizing linear projection

Covariance eigenvectors = principal axes; eigenvalues = variance along axes

Probabilistic PCA reveals assumptions underlying PCA & supports extensions

PCA & FA assumptions: observations = linear mapping from latent space (induce correlations)

PCA: isotropic noise; FA: independent (but not isotropic noise)

Until next week

Read paper and prepare presentation (see notes for Session 6)

Read statistical methods section (see notes for Session 6)

Next session

Q&A about previous session

Paper discussions (~1h)

Laplace approximations and intro to state space models (~40min)