Linear Factor Models

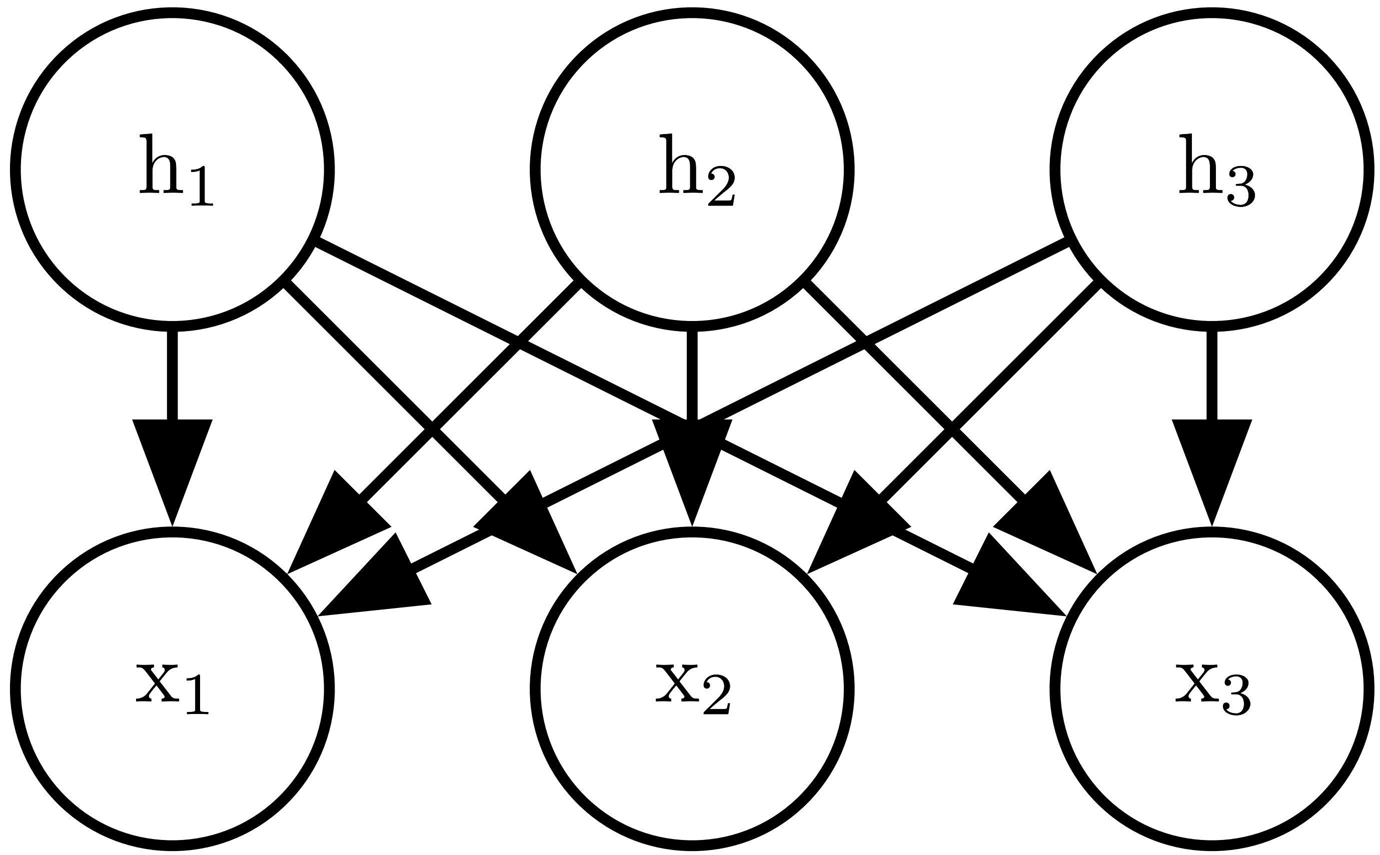
Lecture slides for Chapter 13 of *Deep Learning*

www.deeplearningbook.org

Ian Goodfellow

2016-09-27

Linear Factor Models



$$\mathbf{x} = \boldsymbol{W}\mathbf{h} + \boldsymbol{b} + \mathbf{noise}$$

Figure 13.1

Probabilistic PCA and Factor Analysis

- Linear factor model
- Gaussian prior
- Extends PCA
  - Given an input, yields a distribution over codes, rather than a single code
  - Estimates a probability density function
  - Can generate samples

Independent Components Analysis

- Factorial but non-Gaussian prior
- Learns components that are closer to statistically independent than the raw features
- Can be used to separate voices of n speakers recorded by n microphones, or to separate multiple EEG signals
- Many variants, some more probabilistic than others

Slow Feature Analysis

- Learn features that change gradually over time
- SFA algorithm does so in closed form for a linear model
- Deep SFA by composing many models with fixed feature expansions, like quadratic feature expansion

Sparse Coding

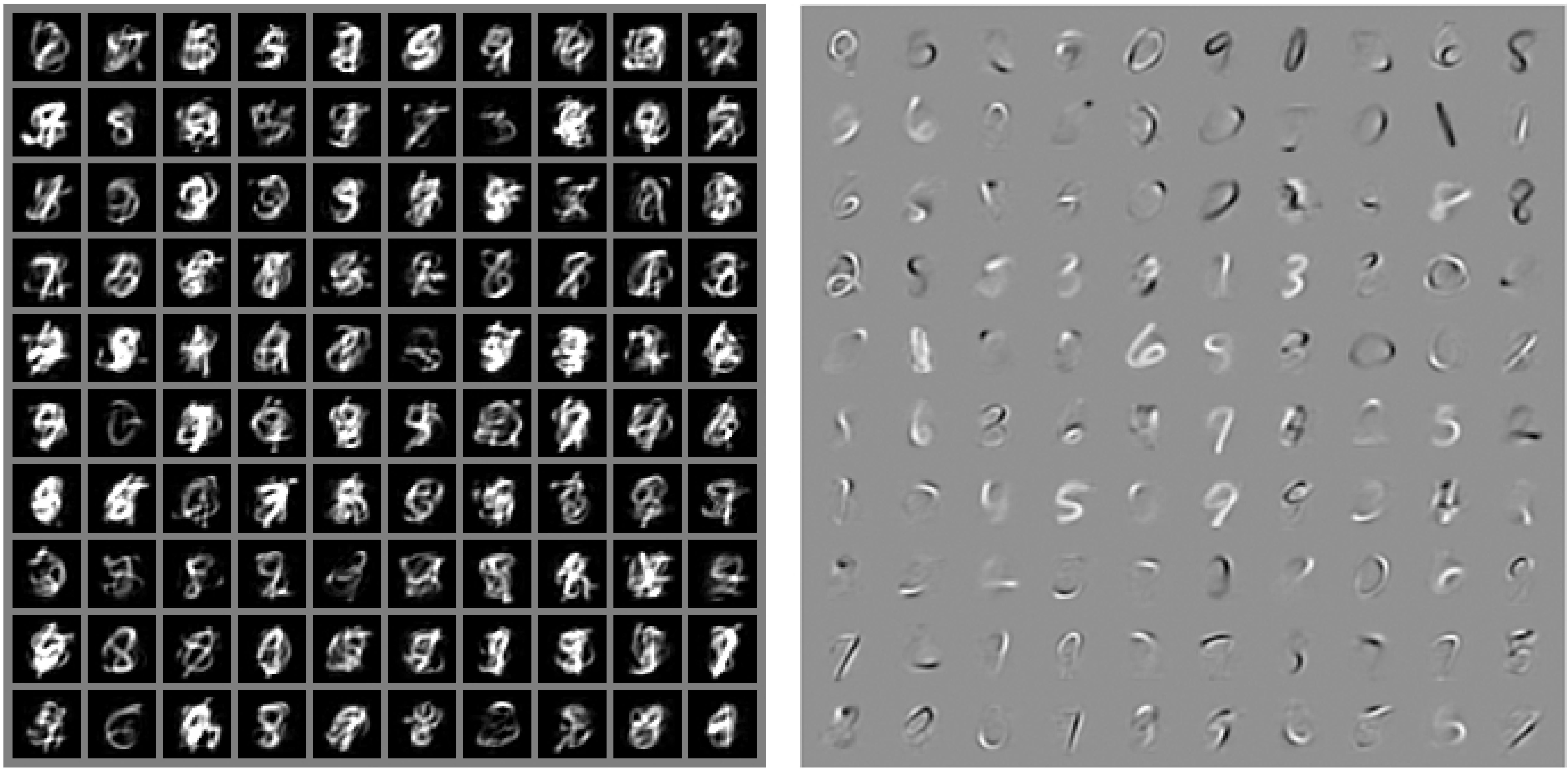
$$p(\boldsymbol{x} \mid \boldsymbol{h}) = \mathcal{N}(\boldsymbol{x}; \boldsymbol{W}\boldsymbol{h} + \boldsymbol{b}, \frac{1}{\beta}\boldsymbol{I}). \tag{13.12}$$

$$p(h_i) = \text{Laplace}(h_i; 0, \frac{2}{\lambda}) = \frac{\lambda}{4} e^{-\frac{1}{2}\lambda |h_i|} \tag{13.13}$$

$$\arg\min_{\boldsymbol{h}} \lambda ||\boldsymbol{h}||_1 + \beta ||\boldsymbol{x} - \boldsymbol{W}\boldsymbol{h}||_2^2, \tag{13.18}$$

(Goodfellow 2016)

Sparse Coding



Samples Weights

Figure 13.2

(Goodfellow 2016)

Manifold Interpretation of PCA

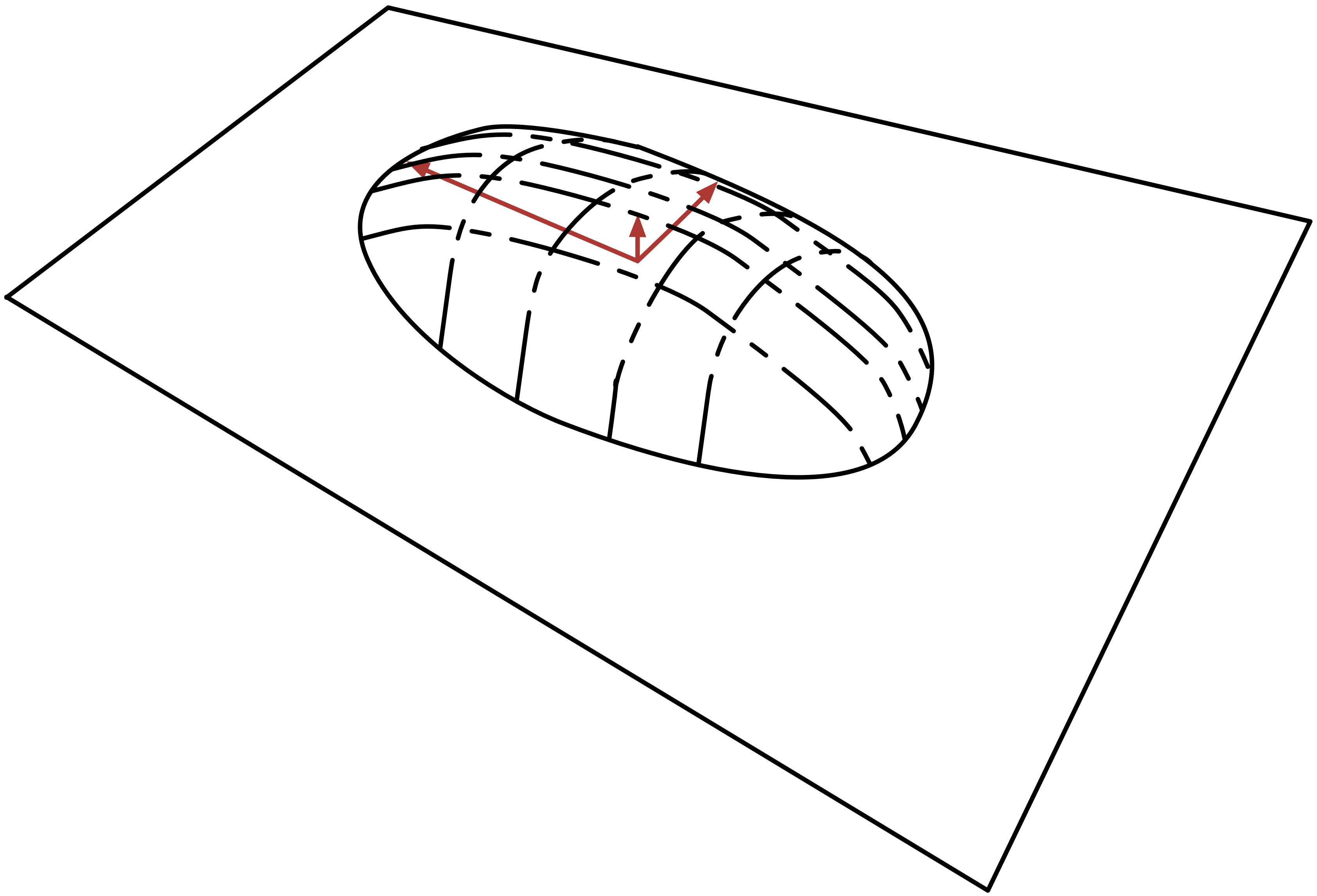


Figure 13.3

(Goodfellow 2016)