Linear Factor Model

黄正杰

Content

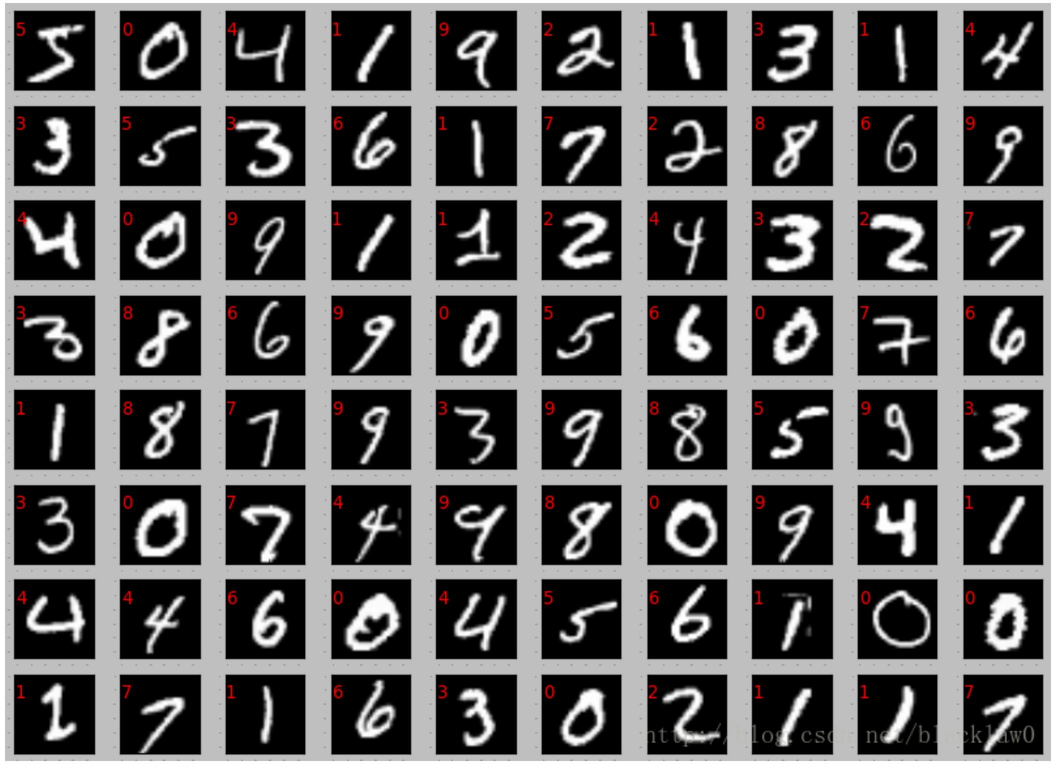
- Linear Factor Model General Form
- Specified Model
  - Probabilistic PCA
  - Independent Component Analysis
  - Sparse Coding
- Conclusion

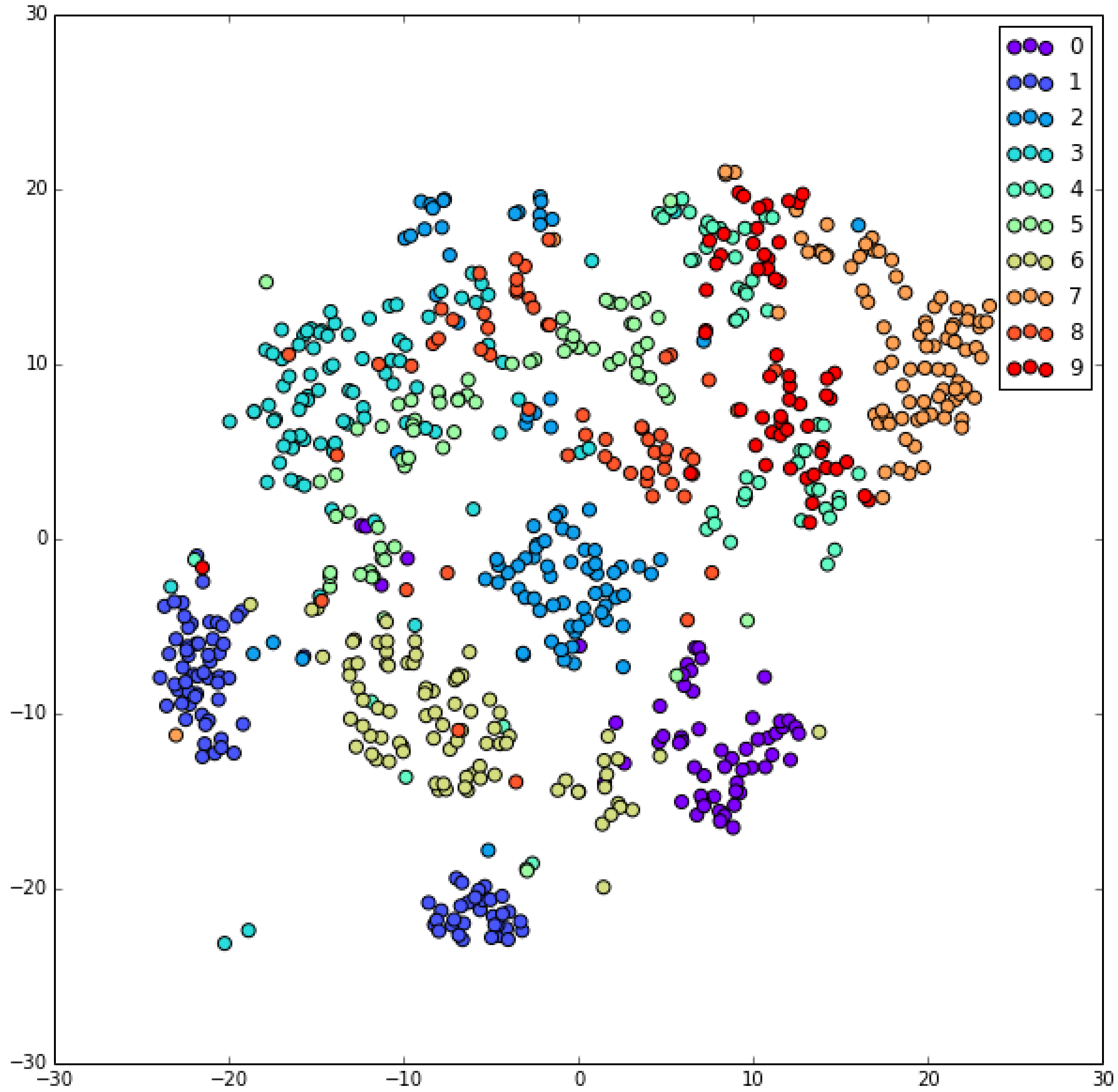
What do we learn?

- Inference
- 获取共有的特征

$$\boldsymbol{x} \xrightarrow{\textbf{encode}} \boldsymbol{h}$$

MNIST CLUSTER





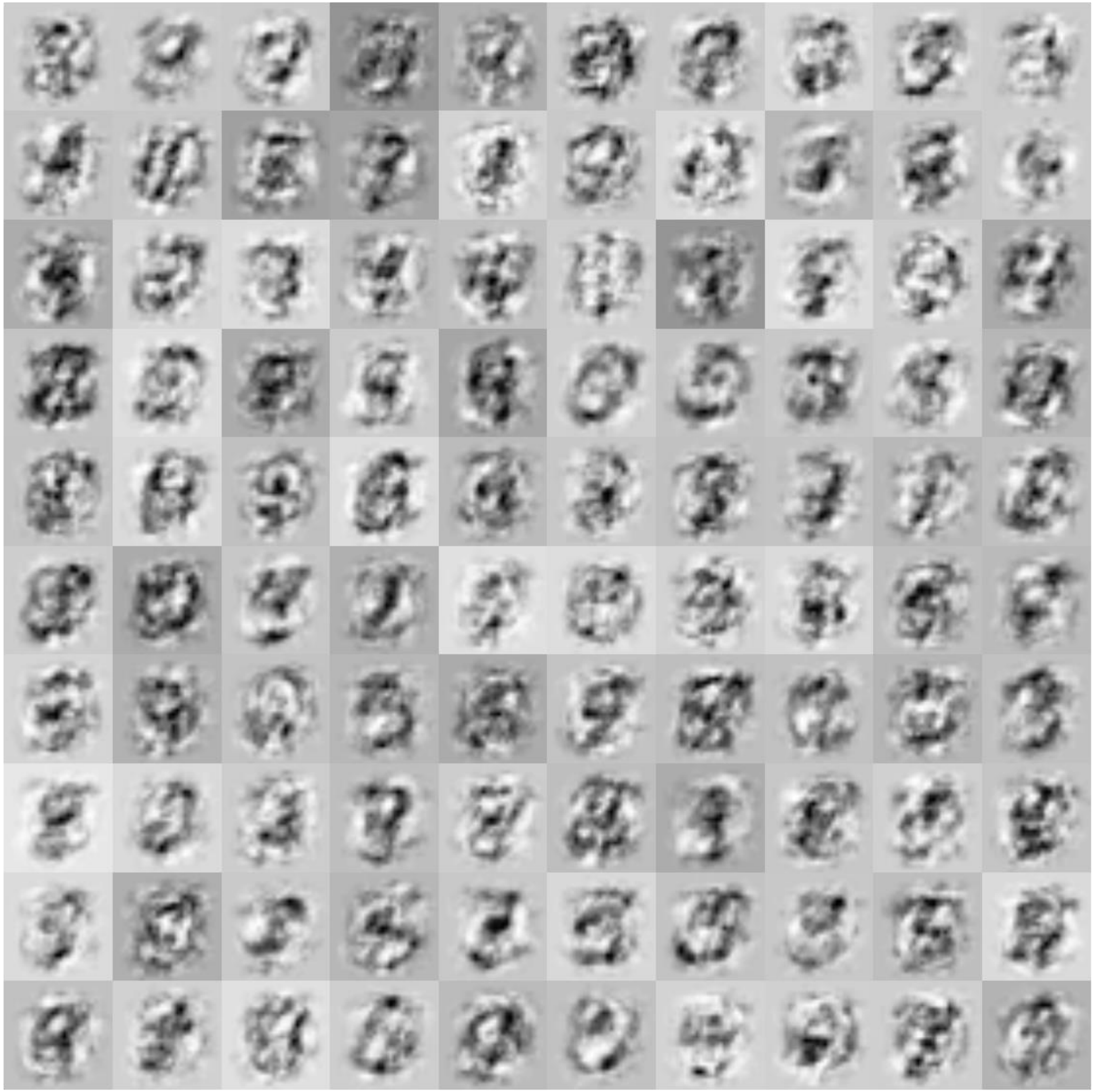
What do we learn?

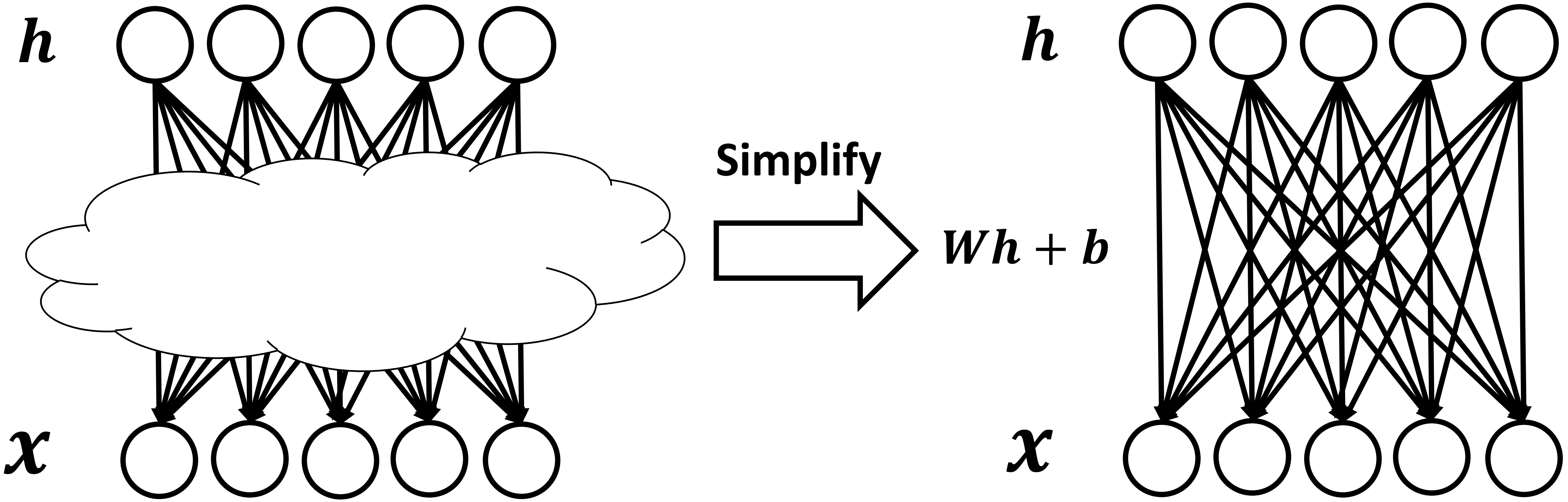
- Generalization

$$\boldsymbol{p(h)} \xrightarrow{\textbf{sampling}} \boldsymbol{h} \xrightarrow{\textbf{decode}} \boldsymbol{x}$$

Generalization

Gaussian vector




h
x
Simplify
$Wh + b$
h
x

Linear Factor Model

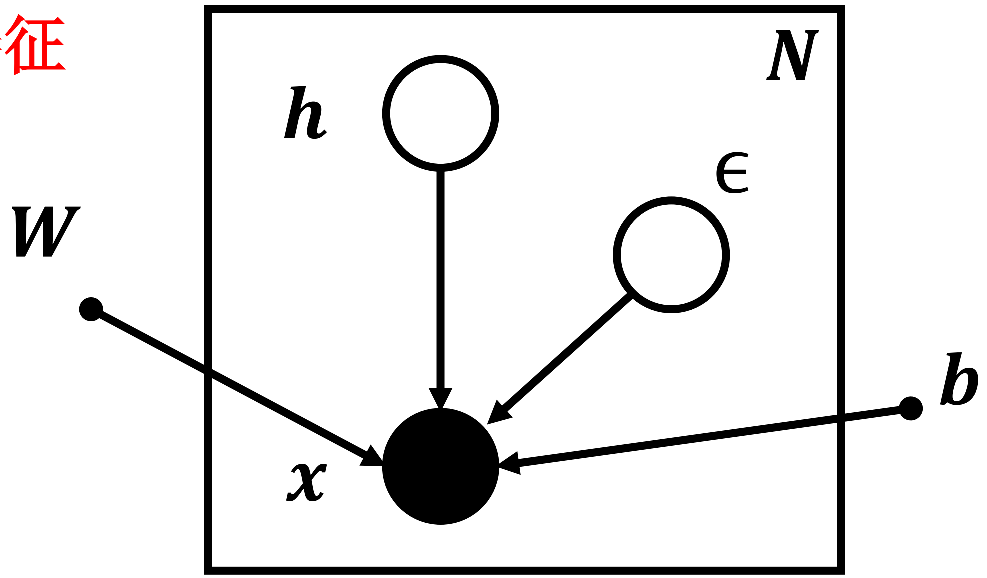
$$\underbrace{x}_{\text{观察值}} = \overbrace{\boxed{W\underbrace{h}_{\text{隐含特征}} + b}}^{\text{线性变换}} + \underbrace{\epsilon}_{\text{观察误差}}$$

- h, ϵ是随机变量，W和b是参数
- 引入h的先验分布

$$h \sim p(h)$$

- 引入误差的分布

$$\epsilon \sim N(0, \sigma^2 I)$$



Linear Factor Model

- Inference h

$$h^* = \underset{h}{\operatorname{argmax}}\, p(h|x)$$

- Generalize x

1. Sample $h \sim p(h)$
2. Sample $\epsilon \sim N(0, \sigma^2 I)$
3. Generalize $x = Wh + b + \epsilon$

	pPCA	ICA	Sparse Coding
h先验	$N(0,I)$	non-gaussian	$Laplace(\lambda)$
误差分布	$N(0,\sigma^2 I)$	无	$N(0,\frac{1}{\beta}I)$
参数估计方法	最大似然 EM算法	最大似然 梯度下降	最大后验 梯度下降
模型	$x = Wh + b + \epsilon$	$x = Wh$	$x = Wh + \epsilon$
参数	W, b, σ	W	W

Probabilistic PCA

Probabilistic PCA

小黑板
$x = Wh + b + \epsilon$

- 高斯先验假设

$$h \sim N(0, I)$$

- 误差分布

$$\epsilon \sim N(0, \sigma^2 I)$$

- 观察值分布

$$x \sim N(b, WW^T + \sigma^2 I), x|h \sim N(Wh + b, \sigma^2 I)$$

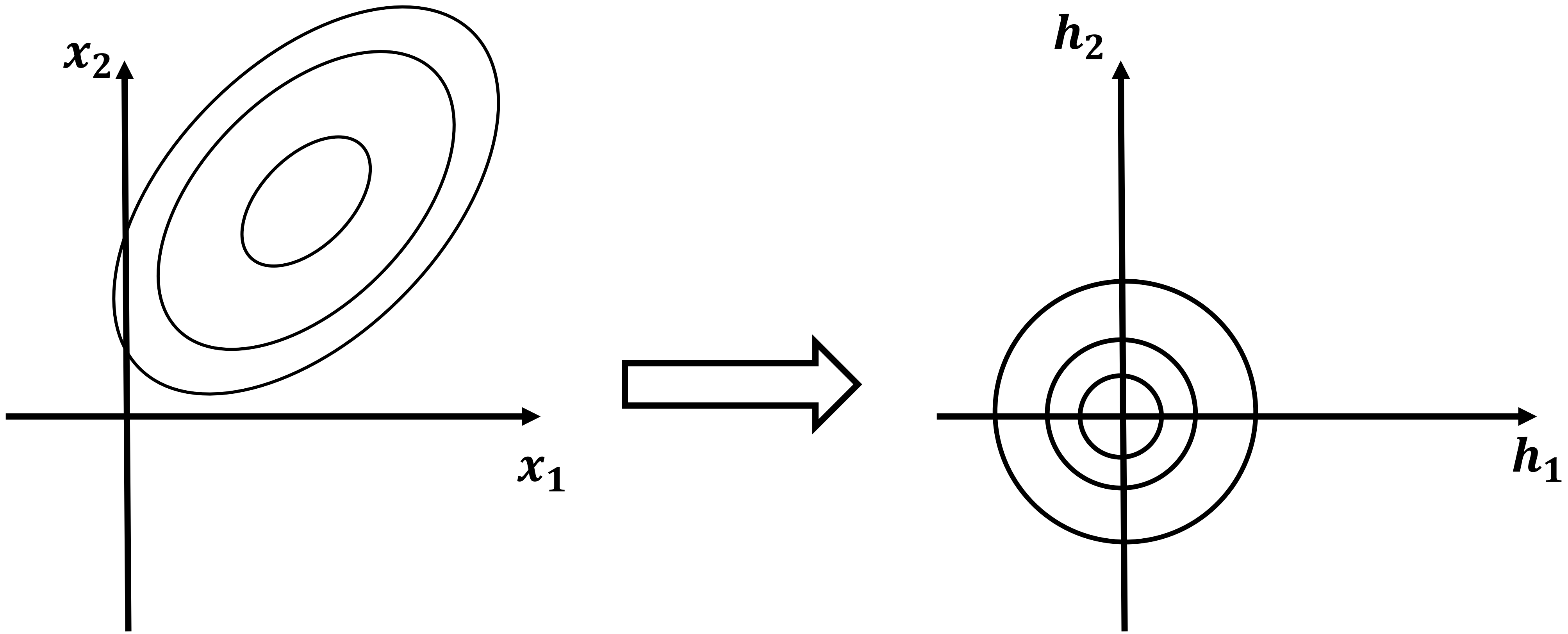
- 参数估计W, σ, b

$$L(W, \sigma, b) = \sum_i^N \log p(x_i)$$

难以直接优化

$$L(W, \sigma, b) = \sum_i^N \log \frac{1}{(2\pi)^{d/2} |WW^T + \sigma^2 I|^{1/2}} e^{-\frac{1}{2}(x_i - b)^T (WW^T + \sigma^2 I)^{-1} (x_i - b)}$$

Probabilistic PCA

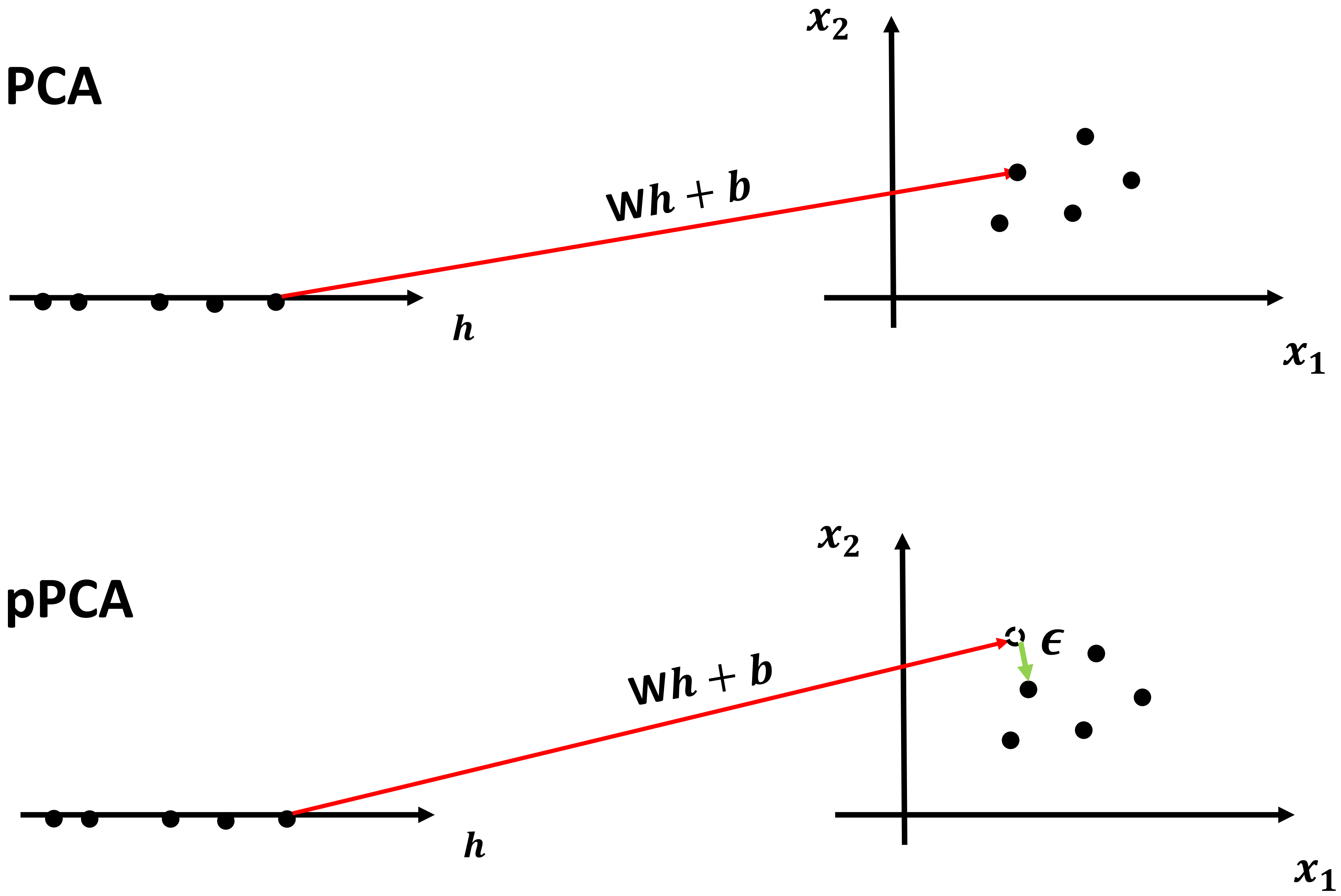


$$x \sim N\left(b, WW^T + \sigma^2 I\right)$$

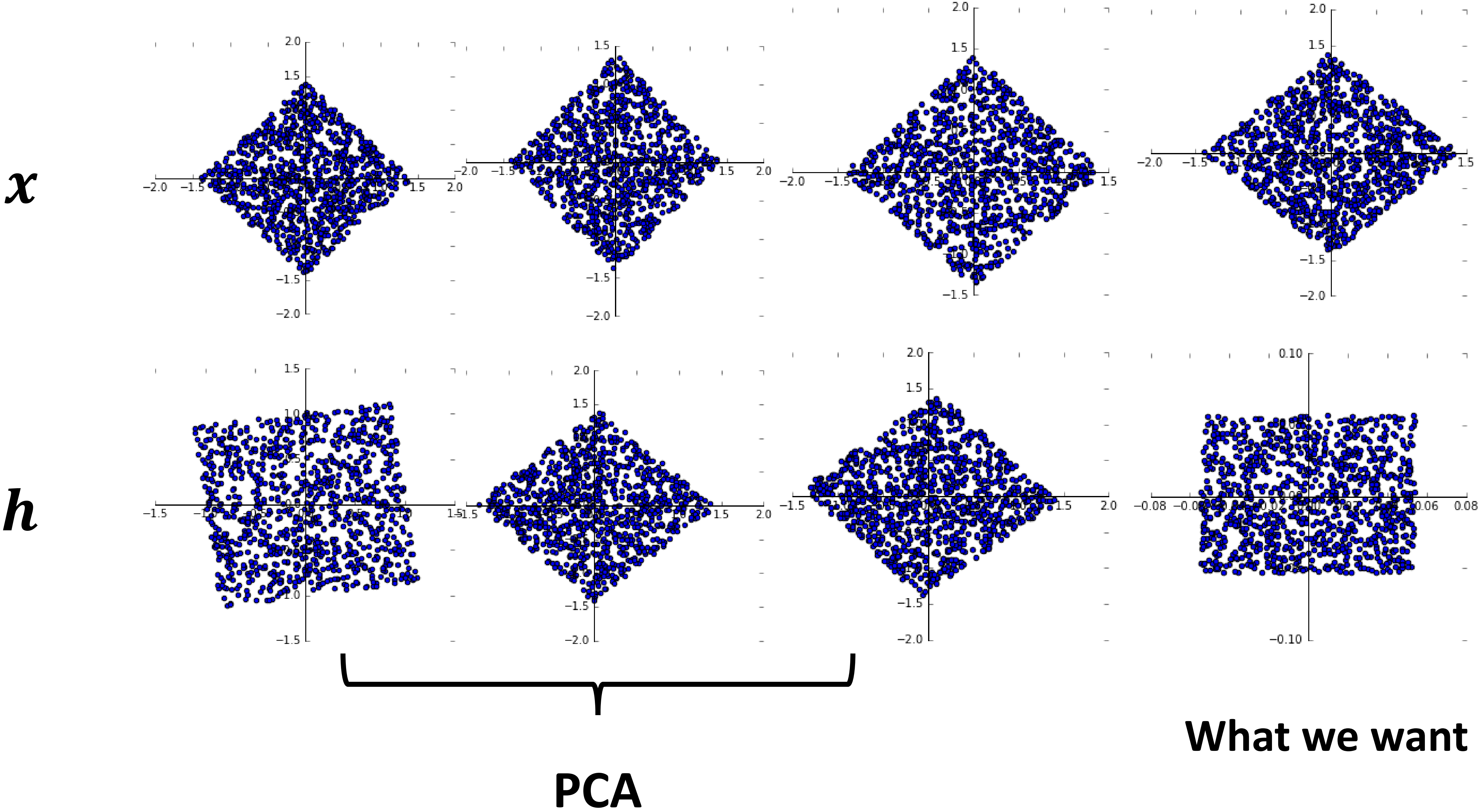
$$h \sim N(0, I)$$

pPCA vs PCA

- pPCA引入了观察误差，对噪声和数据缺失不敏感
  - 当$\sigma \to 0$，有$pPCA \to PCA$
- EM算法迭代计算，效率比PCA要高


PCA
x_2
$Wh + b$
h
x_1
pPCA
x_2
ϵ
$Wh + b$
h
x_1

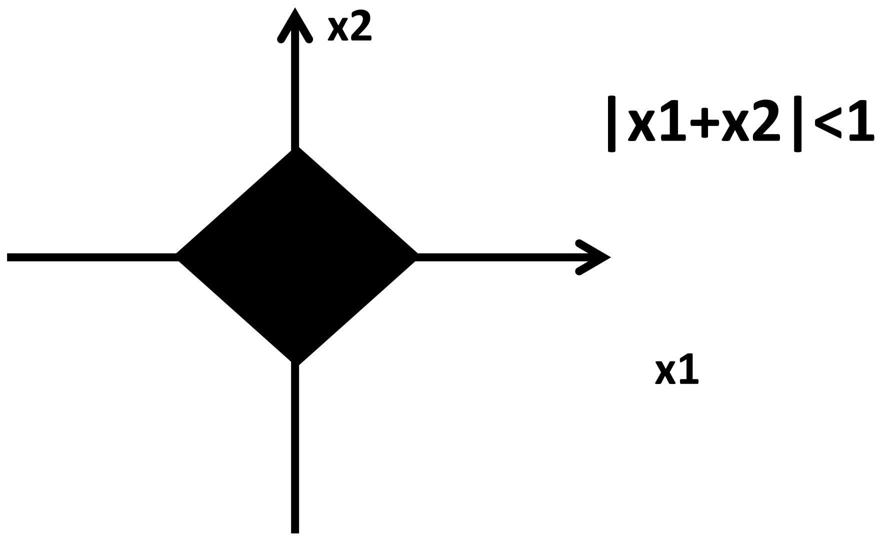
Independent Component Analysis



x
h
PCA
What we want

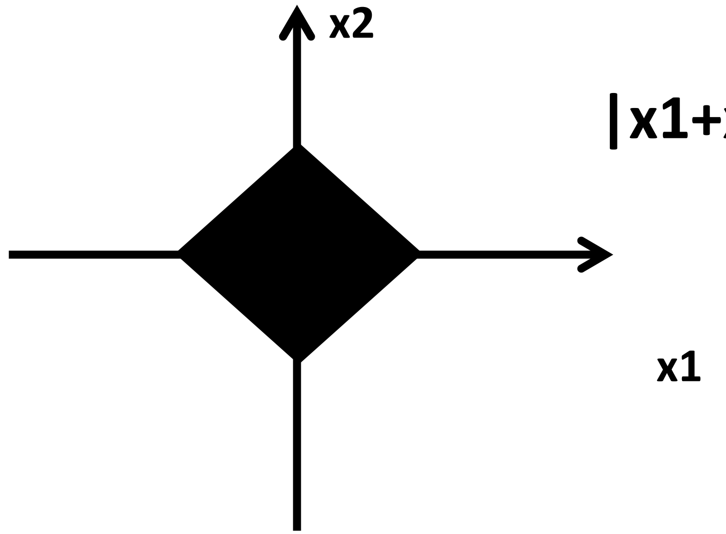
Motivation

- 原始数据不相关且不独立
- 不相关 协方差为0, $Cov(x, y) = E[xy] - E[x]E[y] = 0$
- 独立 $p(x, y) = p(x)p(y)$

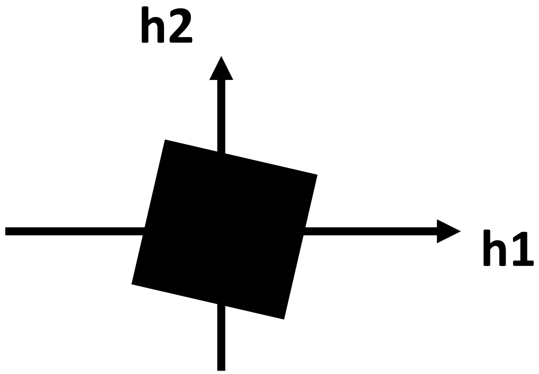


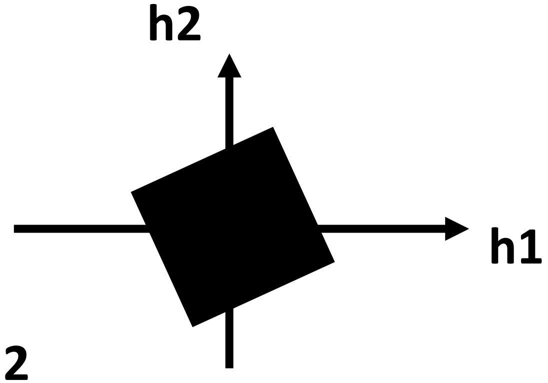
Motivation

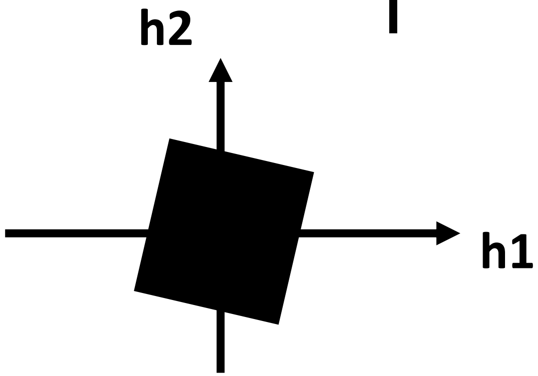
- 高斯分布旋转对称
- 只要是高斯分布解决不了这个问题



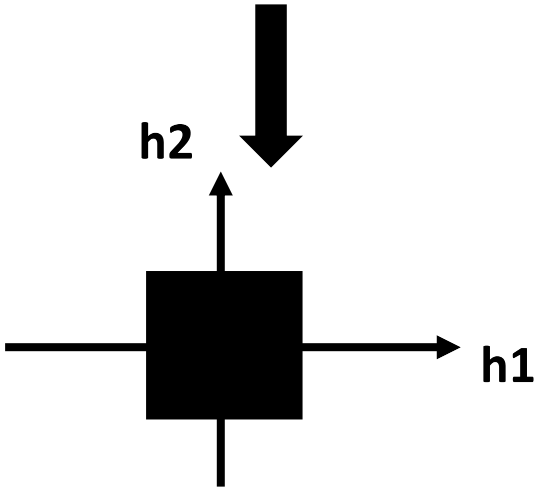
|x1+x2|<1







只有这个是独立的解



ICA

- 隐含变量相互独立

$$p(h) = \prod_i p(h_i)$$

- 引入非高斯分布先验

$$p(h_i) = \sigma'(h_i) = \big(1 - \sigma(h_i)\big)\sigma(h_i)$$

- 不对噪声建模

$$x = Wh \quad \Longrightarrow \quad h = Ax, \ \ A = W^{-1}$$

$$p(x) = p(Ax)|\det A| = \prod_i p(a_i^T x)\,|\det A\,|$$

- 最大似然求解,梯度上升

$$L(A) = \sum_l \log p(x_l) \ \ = \sum_l \sum_i (\log \sigma'\big(a_i^T x\big) + \log|\det(A)|)$$

$$A \coloneqq A + \alpha \frac{\partial L}{\partial A}$$

Experiment

原图片

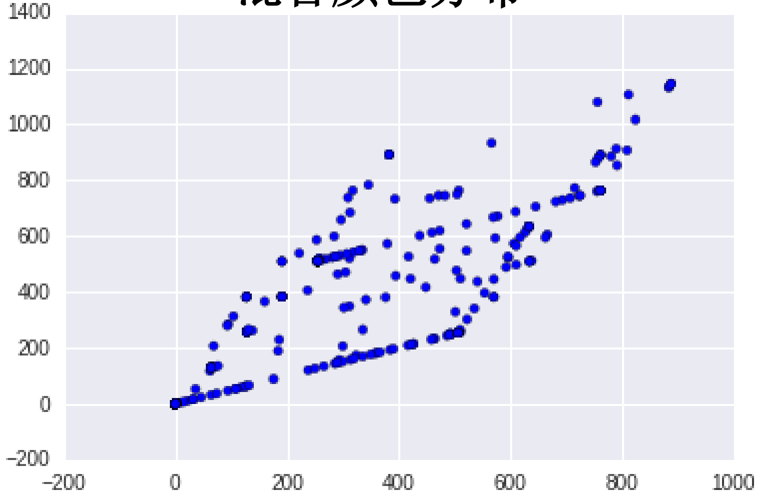
混合图片

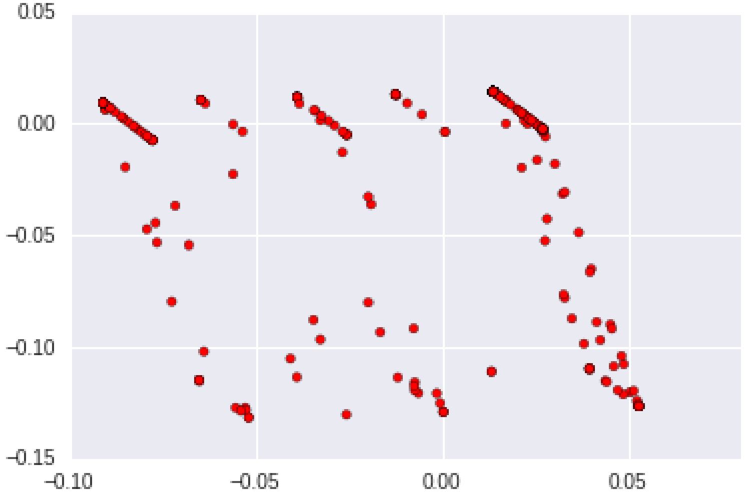
ICA

PCA

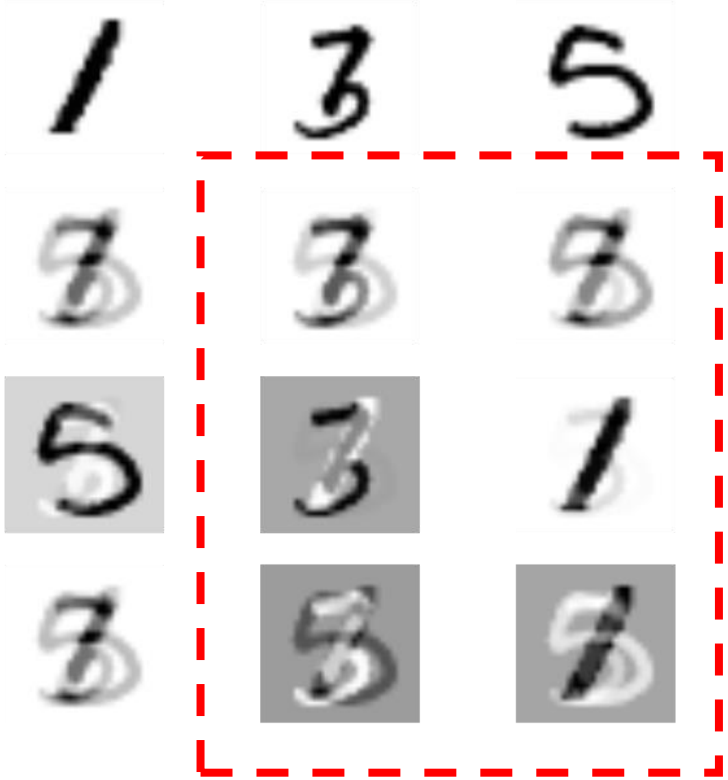
Experiment

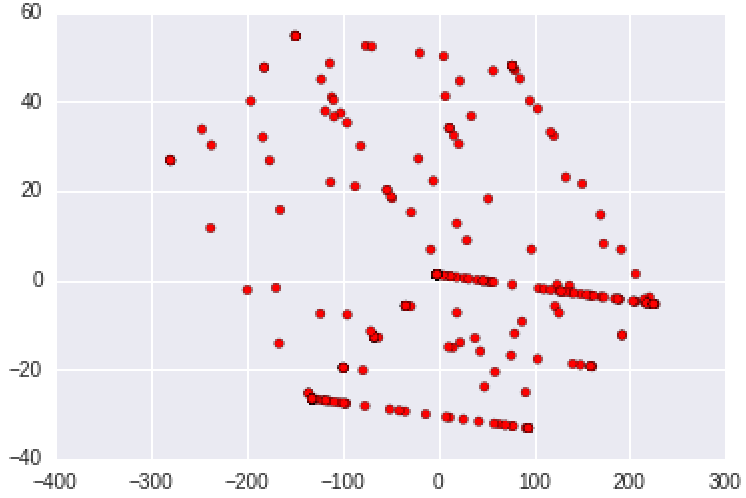
混合颜色分布





ICA分解分布

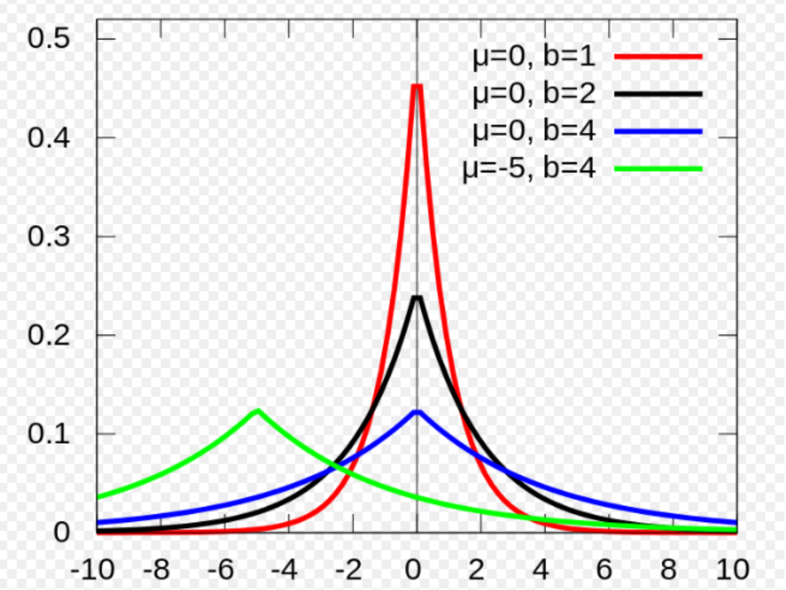




PCA分解分布

Sparse Coding

Sparse Coding



- 为了得到稀疏的特征

$$x = Wh + \epsilon$$

- 隐含特征相互独立

$$p(h) = \prod p(h_i)$$

$$h_i \sim Laplace\left(0, \frac{2}{\lambda}\right), p(h_i) = \frac{\lambda}{4} e^{-\frac{1}{2}|h_i|}$$

- 误差服从高斯分布

$$\epsilon \sim N(0, \frac{1}{\beta} I)$$

Sparse Coding

- 迭代优化encode和decode
- Encode过程

$$\boldsymbol{p}(\boldsymbol{h}|\boldsymbol{x}) = \frac{\boldsymbol{p}(\boldsymbol{x}|\boldsymbol{h})\boldsymbol{p}(\boldsymbol{h})}{\boldsymbol{p}(\boldsymbol{x})}$$

$$h^* = \underset{h}{\operatorname{argmax}}\, p(h|x) = \underset{h}{\operatorname{argmax}} \log p(h|x)$$

$$= \underset{h}{\operatorname{argmax}} \log p(x|h) + \log p(h)$$

$$= \underset{h}{\operatorname{argmax}} \log e^{-\frac{1}{2}\beta\|x-Wh\|_2} + \log e^{-\frac{1}{2}\lambda\|h\|_1}$$

$$= \underset{h}{\operatorname{argmin}}\, \beta\|x - Wh\|_2 + \lambda\|h\|_1$$

Sparse Coding

- Decode过程，给定$\boldsymbol{h}$求解$\boldsymbol{x}$

$$\text{Maximize} \log p(x|h) = \log e^{-\frac{1}{2}\beta\|x-Wh\|_2} = -\frac{1}{2}\beta\|x-Wh\|_2$$

- 等价于

$$\underset{W}{\operatorname{argmin}}\, \beta\|x-Wh\|_2$$

- 交替更新W, h

$$Loss = \|x-Wh\|_2 + \boxed{\lambda\|h\|_1}$$

L1正则

Conclusion

	pPCA	ICA	Sparse Coding
h先验	$N(0,I)$	non-gaussian	$Laplace(\lambda)$
误差分布	$N(0,\sigma^2 I)$	无	$N(0,\frac{1}{\beta}I)$
参数估计方法	最大似然 EM算法	最大似然 梯度下降	最大后验 梯度下降
模型	$x = Wh + b + \epsilon$	$x = Wh$	$x = Wh + \epsilon$
参数	W, b, σ	W	W