ECE 544NA
11/5/2013

TODAY

① RBM W/GAUSSIAN HIDDEN NODES
② STACKED RBM / CONVOLUTIONAL NETS

THE BOLTZMANN LEARNING TRICK: SEPARATE

"ESTIMATED" VALUES

$\hat{h}^{(n)} = \vec{c} + W\vec{v}^{(n)}$

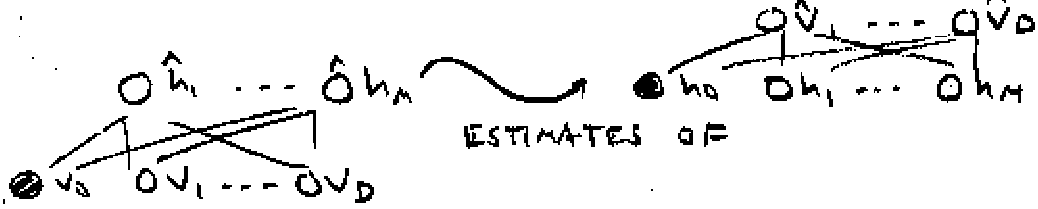
$\hat{v}^{(n)} = \vec{b} + W^T\vec{h}^{(n)}$

"TRUE" VALUES

$\vec{v}^{(n)}$ = OBSERVATION

$\vec{h}^{(n)}$ = TRUE' M-DIMENSION

PCA



MMSE TRAINING:

TRUE VALUE

$$\mathcal{E} = \frac{1}{2}\sum_{n=1}^{N} \left\| \vec{v}^{(n)} - \underbrace{(W^T\vec{h}^{(n)} + \vec{b})}_{\hat{v}^{(n)}} \right\|^2 + \frac{1}{2}\sum_{n=1}^{N} \left\| \vec{h}^{(n)} - \underbrace{(W\vec{v}^{(n)} + \vec{c})}_{\hat{h}^{(n)}} \right\|^2$$

"TRUE" VALUE

$$\boxed{W = \arg\min \mathcal{E}(W) = \arg\max \mathcal{L}(W)}$$

$$\mathcal{L} = \sum_{n=1}^{N} \ln p(\vec{v}^{(n)} \mid \hat{v}^{(n)}) + \sum_{n=1}^{N} \ln p(\vec{h}^{(n)} \mid \hat{h}^{(n)})$$

$$p(\vec{v}^{(n)} \mid \hat{v}^{(n)}) = \frac{1}{(2\pi)^{D/2}} e^{-\frac{1}{2}\|\vec{v}^{(n)} - \hat{v}^{(n)}\|^2}$$

STOCHASTIC LEARNING

① Draw $\vec{h}^{(n)}$ at random from $\mathcal{N}(\vec{h}^{(n)}; W\vec{v}^{(n)} + \vec{c}, I)$

② $w_{ji} \leftarrow w_{ji} - \frac{\eta}{2}\frac{\partial}{\partial w_{ji}} \|\vec{h}^{(n)} - (W\vec{v}^{(n)} + \vec{c})\|^2$

$$- \frac{\eta}{2}\frac{\partial}{\partial w_{ji}} \|\vec{v}^{(n)} - (W^T\vec{h}^{(n)} + \vec{b})\|^2$$

THUS

ⓐ $w_{ji} \leftarrow w_{ji} - \eta\, v_i\,(\overbrace{\vec{m}_j \vec{v}^{(n)} + c_j}^{\hat{h}_j} - h_j^{(n)})$

$$-\eta\, h_j\,(\underbrace{\vec{w}_i^T \vec{h}^{(n)} + b_i}_{\hat{v}_i^{(n)}} - v_i^{(n)})$$

ⓑ $c_j \leftarrow c_j - \eta\,(\underbrace{c_j + \vec{m}_j \vec{v}^{(n)}}_{\hat{h}_j^{(n)}} - h_j)$

ⓒ $b_i \leftarrow b_i - \eta\,(\underbrace{b_i + \vec{w}_i^T \vec{h}^{(n)}}_{\hat{v}_i^{(n)}} - v_i)$

DRAWBACK: $\vec{h}^{(n)}$ ALWAYS RANDOM $\Longrightarrow$ NO FINAL CONVERGENCE!

TEMPERATURE (BOLTZMANN SIMULATED ANNEALING)

$\boxed{\text{INITIALIZE}}$ $\sigma^2 = \max_{1 \le d \le D} \text{Variance}(V_d)$

$\boxed{\text{ITERATE}}$ ① $\vec{h}^{(n)} \sim \mathcal{N}(\vec{h}^{(n)}; W\vec{v}^{(n)} + \vec{c}, \sigma^2 I)$

② a, b, c

③ $\sigma^2 \longleftarrow \alpha\sigma^2$

FOR $\alpha \approx (0.9)^{1/N}$

$\boxed{\text{TERMINATE}}$ e.g. WHEN $\sigma^2 < 10^{-4} \min_{1 \le d \le D} \text{Variance}(V_d)$

GENERALIZATION: RBM TRAINING W/ TEMPERATURES

INITIALIZE T LARGE

ITERATE ① $\vec{h}^{(n)} \sim \frac{1}{Z(\vec{v}^{(n)})} e^{-E(\vec{v}^{(n)}, \vec{h}^{(n)})/T}$

② $W_{ji} \leftarrow W_{ji} - \eta \frac{\partial}{\partial W_{ji}} \left(\ln p(\vec{h}^{(n)} | \hat{h}^{(n)}) + \ln p(\vec{v}^{(n)} | \hat{v}^{(n)}\right.$

③ $T \leftarrow \alpha T$, $\alpha \approx (0.9)^{1/N}$

TERMINATE WHEN T IS SMALL