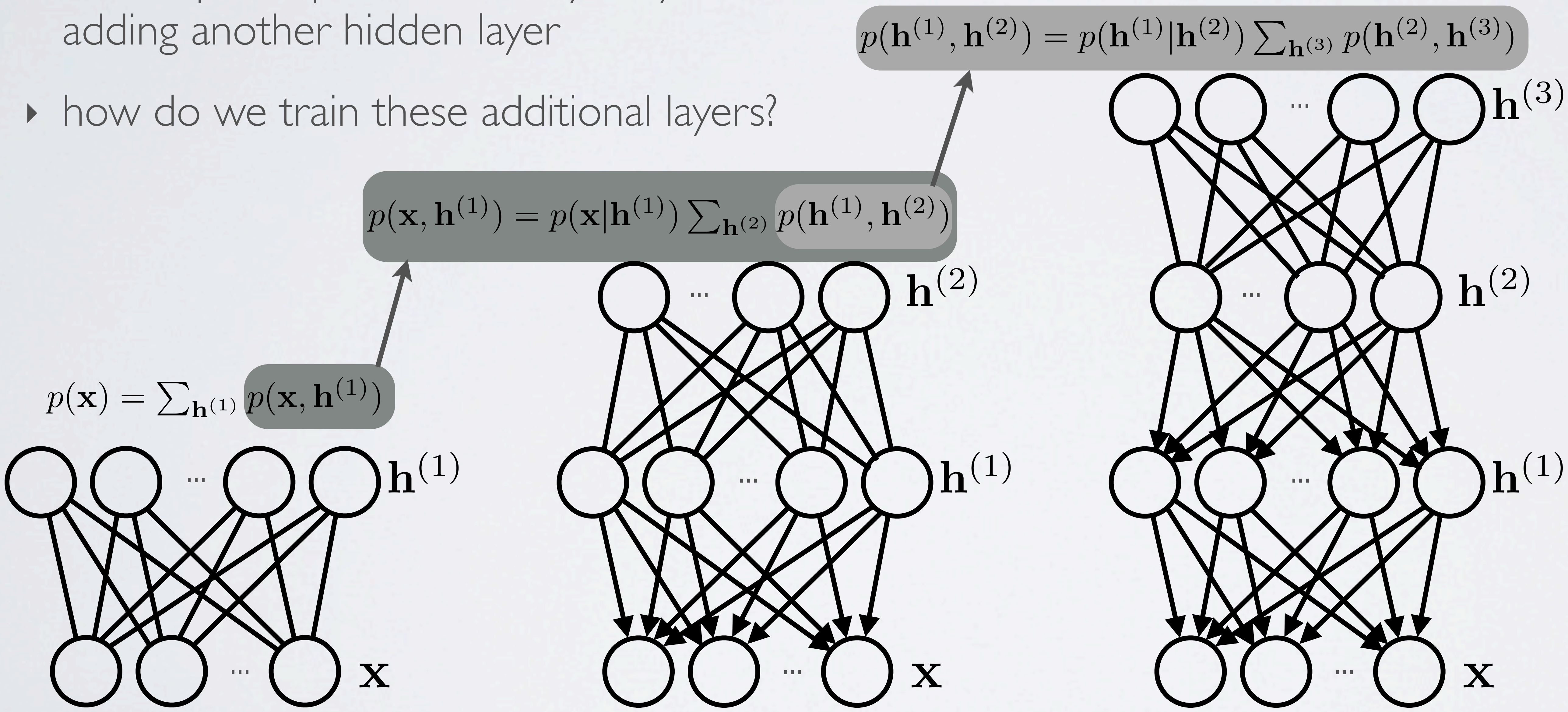
Neural networks

Deep learning - variational bound

DEEP BELIEF NETWORK

Topics: deep belief network

- This is where the RBM stacking procedure comes from
  - idea: improve prior on last layer by adding another hidden layer
  - how do we train these additional layers?

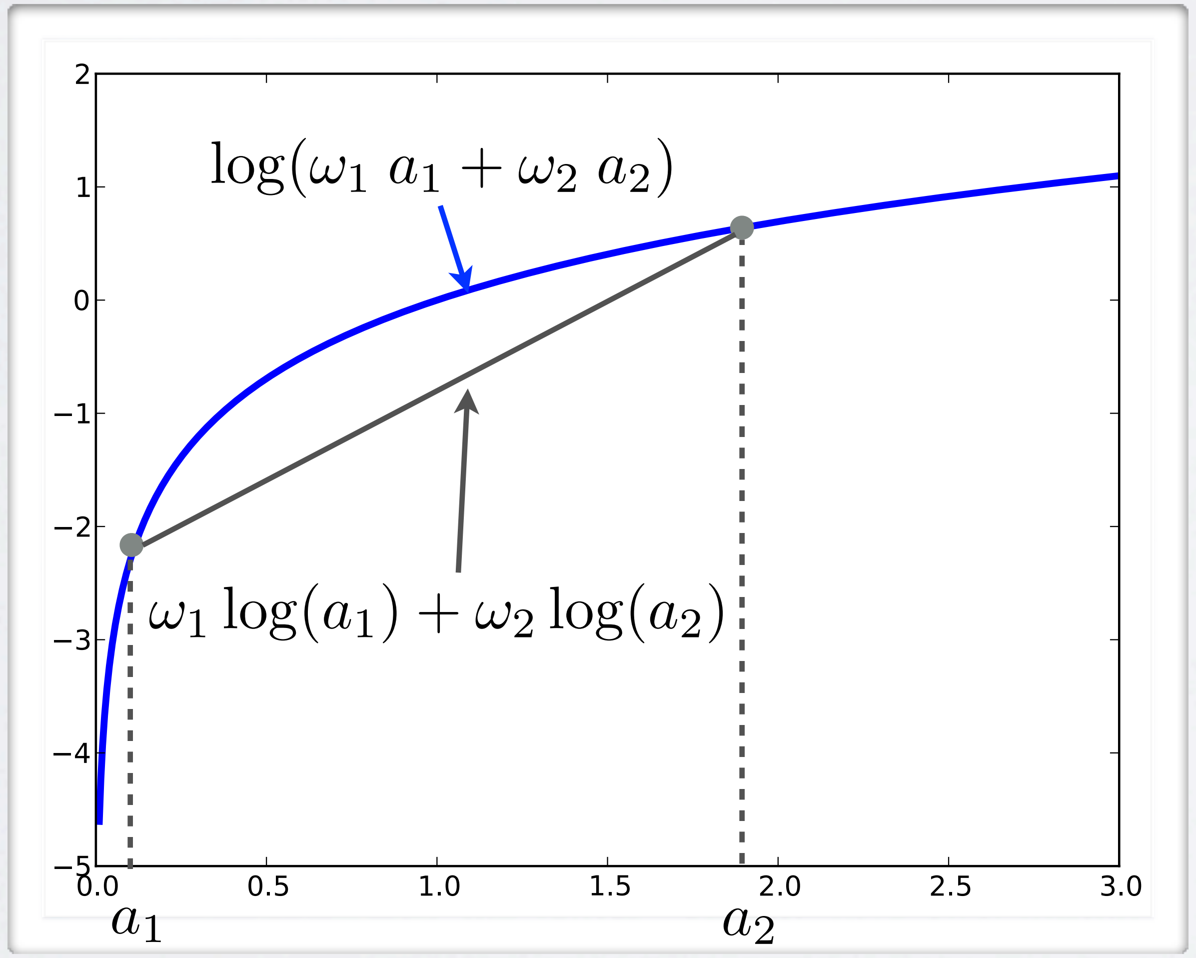


DEEP BELIEF NETWORK

Topics: concavity

- We will use the fact that the logarithm function is concave:

$$\log(\textstyle\sum_i \omega_i \; a_i) \geq \textstyle\sum_i \omega_i \log(a_i)$$ (where $\sum_i \omega_i = 1$ and $\omega_i \geq 0$)



DEEP BELIEF NETWORK

Topics: variational bound

- For any model $p(\mathbf{x}, \mathbf{h}^{(1)})$ with latent variables $\mathbf{h}^{(1)}$ we can write:

$$\begin{aligned}
\log p(\mathbf{x}) &= \log \left(\sum_{\mathbf{h}^{(1)}} q(\mathbf{h}^{(1)}|\mathbf{x}) \frac{p(\mathbf{x}, \mathbf{h}^{(1)})}{q(\mathbf{h}^{(1)}|\mathbf{x})} \right) \\
&\geq \sum_{\mathbf{h}^{(1)}} q(\mathbf{h}^{(1)}|\mathbf{x}) \log \left(\frac{p(\mathbf{x}, \mathbf{h}^{(1)})}{q(\mathbf{h}^{(1)}|\mathbf{x})} \right) \\
&= \sum_{\mathbf{h}^{(1)}} q(\mathbf{h}^{(1)}|\mathbf{x}) \log p(\mathbf{x}, \mathbf{h}^{(1)}) \\
&\quad - \sum_{\mathbf{h}^{(1)}} q(\mathbf{h}^{(1)}|\mathbf{x}) \log q(\mathbf{h}^{(1)}|\mathbf{x})
\end{aligned}$$

- $q(\mathbf{h}^{(1)}|\mathbf{x})$ is any approximation of $p(\mathbf{h}^{(1)}|\mathbf{x})$

DEEP BELIEF NETWORK

Topics: variational bound

- For any model $p(\mathbf{x}, \mathbf{h}^{(1)})$ with latent variables $\mathbf{h}^{(1)}$ we can write:

$$
\begin{aligned}
\log p(\mathbf{x}) &= \log \left(\sum_{\mathbf{h}^{(1)}} \overbrace{q(\mathbf{h}^{(1)}|\mathbf{x})}^{\omega_i} \frac{p(\mathbf{x}, \mathbf{h}^{(1)})}{q(\mathbf{h}^{(1)}|\mathbf{x})} \right) \\
&\geq \sum_{\mathbf{h}^{(1)}} q(\mathbf{h}^{(1)}|\mathbf{x}) \log \left(\frac{p(\mathbf{x}, \mathbf{h}^{(1)})}{q(\mathbf{h}^{(1)}|\mathbf{x})} \right) \\
&= \sum_{\mathbf{h}^{(1)}} q(\mathbf{h}^{(1)}|\mathbf{x}) \log p(\mathbf{x}, \mathbf{h}^{(1)}) \\
&\quad - \sum_{\mathbf{h}^{(1)}} q(\mathbf{h}^{(1)}|\mathbf{x}) \log q(\mathbf{h}^{(1)}|\mathbf{x})
\end{aligned}
$$

- $q(\mathbf{h}^{(1)}|\mathbf{x})$ is any approximation of $p(\mathbf{h}^{(1)}|\mathbf{x})$

DEEP BELIEF NETWORK

Topics: variational bound

- For any model $p(\mathbf{x}, \mathbf{h}^{(1)})$ with latent variables $\mathbf{h}^{(1)}$ we can write:

$$
\begin{aligned}
\log p(\mathbf{x}) &= \log \left(\sum_{\mathbf{h}^{(1)}} \overbrace{q(\mathbf{h}^{(1)}|\mathbf{x})}^{\omega_i} \overbrace{\frac{p(\mathbf{x}, \mathbf{h}^{(1)})}{q(\mathbf{h}^{(1)}|\mathbf{x})}}^{a_i} \right) \\
&\geq \sum_{\mathbf{h}^{(1)}} q(\mathbf{h}^{(1)}|\mathbf{x}) \log \left(\frac{p(\mathbf{x}, \mathbf{h}^{(1)})}{q(\mathbf{h}^{(1)}|\mathbf{x})} \right) \\
&= \sum_{\mathbf{h}^{(1)}} q(\mathbf{h}^{(1)}|\mathbf{x}) \log p(\mathbf{x}, \mathbf{h}^{(1)}) \\
&\quad - \sum_{\mathbf{h}^{(1)}} q(\mathbf{h}^{(1)}|\mathbf{x}) \log q(\mathbf{h}^{(1)}|\mathbf{x})
\end{aligned}
$$

- $q(\mathbf{h}^{(1)}|\mathbf{x})$ is any approximation of $p(\mathbf{h}^{(1)}|\mathbf{x})$

DEEP BELIEF NETWORK

Topics: variational bound

- This is called a variational bound

$$\log p(\mathbf{x}) \geq \sum_{\mathbf{h}^{(1)}} q(\mathbf{h}^{(1)}|\mathbf{x}) \log p(\mathbf{x}, \mathbf{h}^{(1)}) - \sum_{\mathbf{h}^{(1)}} q(\mathbf{h}^{(1)}|\mathbf{x}) \log q(\mathbf{h}^{(1)}|\mathbf{x})$$

  - if $q(\mathbf{h}^{(1)}|\mathbf{x})$ is equal to the true conditional $p(\mathbf{h}^{(1)}|\mathbf{x})$, then we have an equality
  - the more $q(\mathbf{h}^{(1)}|\mathbf{x})$ is different from $p(\mathbf{h}^{(1)}|\mathbf{x})$ the less tight the bound is
  - in fact, the difference between the left and right terms is the KL divergence between $q(\mathbf{h}^{(1)}|\mathbf{x})$ and $p(\mathbf{h}^{(1)}|\mathbf{x})$:

$$\mathrm{KL}(q||p) = \sum_{\mathbf{h}^{(1)}} q(\mathbf{h}^{(1)}|\mathbf{x}) \log \left(\frac{q(\mathbf{h}^{(1)}|\mathbf{x})}{p(\mathbf{h}^{(1)}|\mathbf{x})} \right)$$