1. (Math) Recall that Restricted Boltzmann Machine is defined over binary vectors $v, h \in \{0,1\}^d$ as

$$p(v,h) \coloneqq \frac{\exp(-E(v,h))}{Z}$$

where E is the energy function

$$E(v,h) = -v^T W h - b^T v - c^T h$$

with parameters $W \in \mathbb{R}^{d\times d}, b, c \in \mathbb{R}^d$, and Z is the partition function

$$Z = \sum_v \sum_h \exp(-E(v,h))$$

For simplicity, assume that the dimension $d = 1$. Show that

$$p(h=1|v) = \sigma(c + Wv)$$

where $\sigma(x) \coloneqq \frac{1}{1+\exp(-x)}$ is the logistic function.

Comments: directly calculate $(h=1|v) = \frac{p(v,1)}{\sum_h p(v,h)}$; cancel out the common factors.

$$\text{LHS} = p(h=1|v)$$

$$= \frac{p(v,1)}{\sum_h p(v,h)} \qquad p(v,1) = \frac{\exp(-E(v,1))}{Z}$$

$$= \frac{\frac{1}{Z}\cdot \exp[-E(v,1)]}{\frac{1}{Z}(\exp[-E(v,0)] + \exp[-E(v,1)])}$$

$$= \frac{1/\exp[-E(v,1)]}{1/\exp[-E(v,1)]} \cdot \frac{\exp[-E(v,1)]}{\exp[-E(v,1)] + \exp[-E(v,0)]}$$

$$= \frac{1}{1 + \exp[-E(v,0)]/\exp[-E(v,1)]}$$

$$= \frac{1}{1+\exp[-E(v,0)+E(v,1)]}$$

$$= \frac{1}{1+\exp(-c-wv)}$$

$$= \text{RHS}$$

$$\text{RHS} = \sigma(c + Wv) = \frac{1}{1+\exp(-c-wv)}$$

$\because h \in \{0,1\}^d$

$\therefore \sum_h p(v,h)$

$= p(v,0) + p(v,1)$

$= \frac{\exp[-E(v,0)]}{Z} + \frac{\exp[-E(v,1)]}{Z}$

$= \frac{\exp[-E(v,0)] + \exp[-E(v,1)]}{Z}$

$E(v,0) = -b^T v$

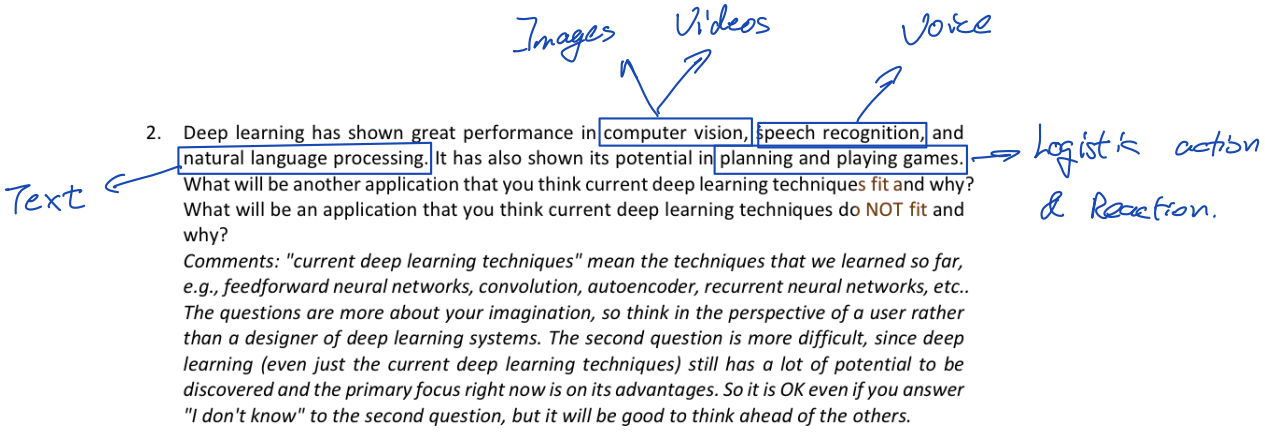
$E(v,1) = -v^T w - b^T v - c^T$

$E(v,1) - E(v,0)$

$= -v^T w - b^T v - c^T + b^T v$

$= -v^T w - c^T$

$= -c - wv$



1) Current deep learning techniques fit with applications with flattened data structures. Another application that current deep learning fit with is fraud detection. Deep learning techniques can learning from the history to identify certain combination of words, images, pattern of speech, or phone number that are labeled as fraud. When the new message comes, it will know this is fraud.

2) Current deep learning techniques does not fit applications with

- Complex hierarchical structures
- Lateral applications with small data collection
- Offline interactive applications that requires incrementally, interactively real time learning.