Homework #4 Q & A (Due 04/02/2021 @ 11:30PM)

Questions?

Problem 1 → MLPs on CIFAR-10

Problem 2 → Adaboost

Problem 3 → RNNs for LMs

for $t = 1, \dots, T$

P2

① $h_t = \text{WeakLearn}(D_t, \{x_i, y_i\}, \mathcal{H})$ ← tree

② $\varepsilon_t = \sum_{i \in \text{wrong}} D_t(i)$

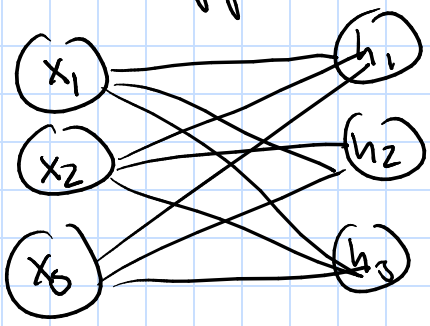
③ $\alpha_t = \frac{1}{2} \log \frac{1 - \varepsilon_t}{\varepsilon_t}$

④ $D_{t+1}(i) = \frac{D_t(i)}{Z_t} \exp\{-\alpha_t y_i h_t(x_i)\}$

```
c = CART(max_depth=4)
idx = rsample(D, len(D))
X_t, y_t = X[idx], y[idx]
h[t] = c.fit(X_t, y_t)
```

Restricted Boltzmann Machine

Energy-based model. Generative network → Maximizing $p(x)$



$x_i, h_i \in \{0, 1\}$

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

$$E(x, h) = -x^T W x - b^T x - d^T h$$

The RBM shown defines a distribution over

$[x_1, x_2, x_3, h_1, h_2, h_3]$

e^{a+b+c}

$(e^{a+b})e^c$

$$p(h|x) = \frac{p(x, h)}{\sum_{h'} p(x, h')}$$

$$= \frac{\frac{1}{Z} \exp(-E(x, h))}{\sum_{h'} \frac{1}{Z} \exp(-E(x, h'))} = \frac{\exp(x^T W x + b^T x + d^T h)}{\sum_{h'} \exp(x^T W x + b^T x + d^T h')}$$

$$p(h|x) = \prod_j p(h_j|x)$$

$$p(h_j=1|x) = \frac{\exp(x^T W_j + d_j)}{Z}$$

$$= \text{logistic}(x)$$

θ

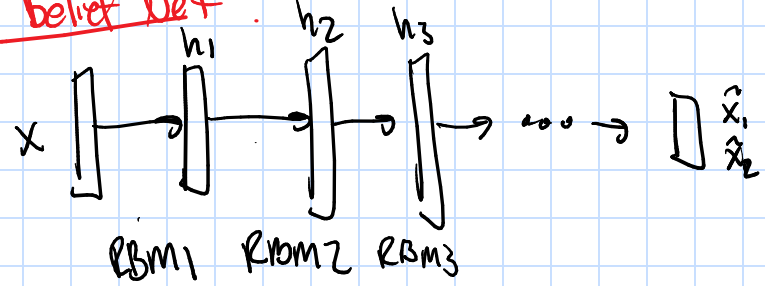
We need to maximize $\log p(x)$ wrt (W, b)

$$\frac{\partial \log p}{\partial \theta} = -\sum_h p(h|x) \frac{\partial E(x,h)}{\partial \theta} + \sum_{x,h} p(x,h) \frac{\partial E(x,h)}{\partial \theta}$$

(1) (2)

$\frac{\partial E}{\partial \theta}$ → easy to compute

Deep belief net



Recurrent Nets

Our previous work, focused on training a feedforward neural net. That is given a point $x \in \mathbb{R}^D$, the sample was passed through the net to get $\hat{y}$. In this context, we are classifying a single point. However, what if we have a sequence? $x_1, x_2, \ldots, x_t$

- In some settings we have a history

Example: Language Model

Consider the task of making a prediction on the next word given the current word and a history. To do this prediction task, we need to know some context. So given $w(t)$ and $h(t-1)$, what is

$$\Pr\{w(t+1) \mid w(t), h(t-1)\}$$

Can you pick up milk at the ____ ?

$h(t-1)$ → Can you pick up milk at the

$w(t)$ → the

$w(t+1)$ → ____

→ or some representation of this part