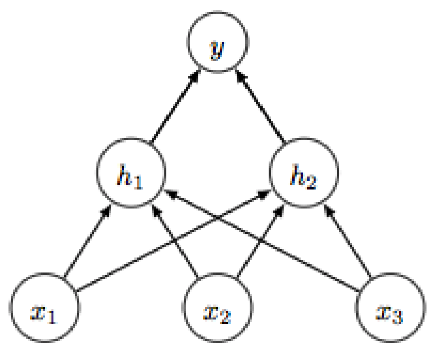
Discussion 10: Neural Networks and Multi-armed Bandit
Machine Learning, Spring 2019

1 Forward and Backword Propogation

The following graph shows the structure of a simple neural network with a single hidden layer. The input layer consists of three dimensions $x = (x_1, x_2, x_3)$. The hidden layer includes two units $h = (h_1, h_2)$. The output layer inclydes one unit y. We ignore bias terms for simplicity.

Figure 1



We use linear rectified units $\sigma(z) = max(0, z)$ as activation function for the hidden layer and the output layer. Moreover, denote by $l(y, t) = \frac{1}{2}(y - t)^2$ tge loss function. Here t is the target value for the output unit y. Denote by W and V weight matrices connecting input and hidden layer, and hidden layer and output respectively. They are initialized as follows:

$$W = \begin{bmatrix} 1 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix} \text{ and } V = \begin{bmatrix} 0 & 1 \end{bmatrix} \text{ and } x = \begin{bmatrix} 1 & 2 & 1 \end{bmatrix} \text{ and } t = 1.$$

Also assume that we have at least one sample (x, t) given by the values above.

(1). Write out *symbolically* (no need to plug in the specific values of W and V just yet) the mapping $x \to y$ using σ, W, V.

(2) Assume that the current input is $x = (1, 2, 1)$. The target value is $t = 1$. Compute the *numerical* output value y, clearly show all intermediate steps. You can reuse the results of the previous question.

(3) Compute the gradient of the loss function with respect to the weights. In particular compute the following terms *symbolically*:

- The gradient relative to V, i.e. $\frac{\partial l}{\partial V}$
- The gradient relative to W, i.e. $\frac{\partial l}{\partial W}$
- Compute the values *numerically* for the choices of W,V, x, y given above.

Let:

- $\frac{\partial y}{\partial Vh} = g$ where $0 < g < 1$ is the subgradient of ReLU
- $\frac{\partial y}{\partial h} = V$
- $\frac{\partial h}{\partial Wx} = M$ where $M = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$

2 Multi-armed bandit: an algorithm for 2-armed bandits

In this section, we study a simple case where confidence intervals are used. The materials from this section should largely be credited to Dr. Aleksandrs Slivkins. First of all, recall that in an multi-armed bandit problem, we are faced with multiple bandit machines with different reward distributions. Our goal is to locate the bandit machine with the best expected reward as quickly as possible. Before we start the discussion, let's review some notations.

- $T_j(t)$ - number of times arm j gets pulled up to time t.
- $\widehat{X}_{j,n}$ - the average of n observed reward for arm j. We use $\widehat{X}_j$ as a shorthand when the value of n is clear.
- $\mathbb{E}[R_n]$ - the expected cumulative. More specifically, $\mathbb{E}[R_n] = \sum_{t=1}^{n}\sum_{j=1}^{m}(\mu_* - \mu_j)\mathbb{1}_{I_t=j}$ where m is the total number of arms, μ_j is the expected reward of arm j, μ_* is the highest expected reward among all arms, and I_t is the index of the arm selected at time t.

Algorithm 1: arm elimination algorithm (for two arms)

Input : two arms - arm 1 and arm 2, and a time horizon n ($n > 4$).
Initialization: play both arms once and initialize $\widehat{X}_j$ for each $j = 1, 2$
Play arm j and arm $3-j$ alternatively for t_0 rounds, until $\widehat{X}_j - \sqrt{\frac{2\log n}{T_j(t_0-1)}} > \widehat{X}_{3-j} + \sqrt{\frac{2\log n}{T_{3-j}(t_0-1)}}$ for some j
From there on until time n, play only the arm j.

Theorem 2.1 (Regret-bound for the arm elimination algorithm) *The bound on the mean regret $\mathbb{E}[R_n]$ at time n is given by*

$$\mathbb{E}[R_n] \quad \leq \quad \mathcal{O}\left(\sqrt{n\log n}\right) \tag{1}$$

To analyze the regret of this simple algorithm, we first assume that $n \geq t_0$. Also, we need the Hoeffding's inequalities for i.i.d. random variables $X_{j,1}, \cdots, X_{j,T_j(t-1)}$ that are bounded between 0 and 1:

$$\mathbb{P}\left(\frac{1}{T_j(t-1)}\sum_{i=1}^{T_j(t-1)} X_{j,i} - \mu_j \geq \varepsilon\right) \leq \exp\{-2T_j(t-1)\varepsilon^2\}, \tag{2}$$

and

$$\mathbb{P}\left(\frac{1}{T_j(t-1)}\sum_{i=1}^{T_j(t-1)} X_{j,i} - \mu_j \leq -\varepsilon\right) \leq \exp\{-2T_j(t-1)\varepsilon^2\}, \tag{3}$$

where $\mu_j = \mathbb{E}[X_{j,i}]$. By setting $\epsilon = \sqrt{\frac{2\,\log n}{T_j(t-1)}}$, we know that

$$\mathbb{P}\left(\widehat{X}_{j,T_j(t-1)} - \mu_j \geq \sqrt{\frac{2\,\log n}{T_j(t-1)}}\right) \leq \frac{1}{n^4}, \tag{4}$$

and

$$\mathbb{P}\left(\widehat{X}_{j,T_j(t-1)} - \mu_j \leq -\sqrt{\frac{2\,\log n}{T_j(t-1)}}\right) \leq \frac{1}{n^4}, \tag{5}$$

Let us denote the gap between the expected rewards of the two arms by $\Delta := |\mu_1 - \mu_2|$. Let us also define the event $\mathcal{E} := \left\{\left|\widehat{X}_{j,T_j(t-1)} - \mu_j\right| \leq \sqrt{\frac{2\,\log n}{T_j(t-1)}},\ \text{for } j = 1 \text{ and } j = 2, \text{ and for } T_j(t-1) = 1, 2, \cdots, n\right\}$. Since at time $t_0 - 1$, we have

$$\widehat{X}_{j,T_j(t_0-2)} - \sqrt{\frac{2\,\log n}{T_j(t_0-2)}} \leq \widehat{X}_{3-j,T_{3-j}(t_0-2)} + \sqrt{\frac{2\,\log n}{T_{3-j}(t_0-2)}}.$$

for both $j = 1$ and $j = 2$. Thus $\left|\widehat{X}_{1,T_1(t_0-2)} - \widehat{X}_{2,T_2(t_0-2)}\right| \leq \sqrt{\frac{2\,\log n}{T_1(t_0-2)}} + \sqrt{\frac{2\,\log n}{T_2(t_0-2)}}$. When $\mathcal{E}$ happends,

$$\begin{aligned}
\Delta &= |\mu_1 - \mu_2| \\
&= |\mu_1 - \widehat{X}_{1,T_1(t_0-2)} + \widehat{X}_{1,T_1(t_0-2)} - \widehat{X}_{2,T_2(t_0-2)} + \widehat{X}_{2,T_2(t_0-2)} - \mu_2| \\
&\leq |\mu_1 - \widehat{X}_{1,T_1(t_0-2)}| + |\widehat{X}_{1,T_1(t_0-2)} - \widehat{X}_{2,T_2(t_0-2)}| + |\widehat{X}_{2,T_2(t_0-2)} - \mu_2| \\
&\leq 2\left(\sqrt{\frac{2\,\log(t)}{T_1(t_0-2)}} + \sqrt{\frac{2\,\log(t)}{T_2(t_0-2)}}\right).
\end{aligned}$$

Since $T_1(t_0-2) = T_2(t_0-2) = \boldsymbol{\theta}\,(t_0)$ by our alternating playing scheme, we have

$$\Delta = \mathcal{O}\left(\sqrt{\frac{\log n}{t_0}}\right).$$

By the Fréchet inequality,

$$\begin{aligned}
\mathbb{P}(\mathcal{E}) &\geq \left(1 - \frac{4}{n^4}\right)n - (n-1) \\
&= 1 - \frac{4}{n^3}
\end{aligned}$$

Now, by the law of total expectation,

$$
\begin{aligned}
\mathbb{E}[R_n] &= \mathbb{E}[R_n|\mathcal{E}]\mathbb{P}(\mathcal{E}) + \mathbb{E}[R_n|\bar{\mathcal{E}}](1-\mathbb{P}(\mathcal{E})) \\
&= \Delta\,\mathcal{O}\left(t_0\right)\mathbb{P}(\mathcal{E}) + \mathbb{E}[R_n|\bar{\mathcal{E}}]\frac{4}{n^3} \\
&= \mathcal{O}\left(\sqrt{\frac{\log n}{t_0}}\right)\,\mathcal{O}\left(t_0\right)\left(1-\frac{4}{n^3}\right) + n\frac{1}{n^4} \\
&= \mathcal{O}\left(\sqrt{t_0 \log n}\right) \\
&= \mathcal{O}\left(\sqrt{n \log n}\right)
\end{aligned}
$$