CPEN455: Deep Learning
Homework 1

Mercury Mcindoe 85594505

January 19th 2025

1 Problem 1

1.1

Solution:

1.2

Solution:

Let's first consider the case before Dropout (*i.e.*, $\mathbf{h}$). Since $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, I)$, each entry x_i within $\mathbf{x}$ follows a normal distribution $\mathcal{N}(0,1)$ and each are iid. Let $\mathbf{z} = W\mathbf{x}$,

$$\mathbb{E}[\mathbf{z}] = \mathbb{E}[W\mathbf{x}] = W\mathbb{E}[\mathbf{x}] = 0$$

$$\text{Var}(\mathbf{z}) = \text{Var}(W\mathbf{x}) = W\text{Var}(\mathbf{x})W^T = W \begin{bmatrix} \text{Var}(x_1) & \text{Cov}(x_1, x_2) & \cdots & \text{Cov}(x_1, x_N) \\ \text{Cov}(x_2, x_1) & \text{Var}(x_2) & \cdots & \text{Cov}(x_2, x_N) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(x_n, x_1) & \text{Cov}(x_n, x_2) & \cdots & \text{Var}(x_n, x_n) \end{bmatrix} W^T$$

Since we know that each x_i is iid which follows $\mathcal{N}(0,1)$, the variance-covariance matrix of $\mathbf{z}$ is then,

$$\text{Var}(\mathbf{z}) = W \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} W^T = WW^T = I$$

Which shows that $\mathbf{z} = W\mathbf{x} \sim \mathcal{N}(0, I)$. Now considering that $\sigma(\mathbf{z}) = \max(\mathbf{z}, 0)$, each entry z_i would have the following expectations and variances,

$$\mathbb{E}[h_i] = \int_0^\infty \frac{z}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = \left[-\frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \right]_0^\infty = \frac{1}{\sqrt{2\pi}}$$

$$E[h_i^2] = \int_0^\infty \frac{z^2}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = \frac{1}{2}$$

$$\mathrm{Var}(h_i) = \mathbb{E}[h_i^2] - (\mathbb{E}[h_i])^2 = \frac{1}{2} - \frac{1}{2\pi}$$

Putting all these together,

$$\therefore \mathrm{Var}(\mathbf{h}) = \left(\frac{1}{2} - \frac{1}{2\pi}\right) I_M$$

Now let's consider after Dropout, we know that $\tilde{\mathbf{h}} = \frac{\mathbf{m}}{1-p} \odot \mathbf{h}$, for an try $\tilde{h}_i$,

$$\begin{aligned}
\mathbb{E}[\tilde{h}_i] &= \mathbb{E}\left[\frac{m_i}{1-p} \cdot h_i\right] \\
&= \frac{1}{1-p}\mathbb{E}[m_i] \cdot \mathbb{E}[h_i] \\
&= \frac{1-p}{1-p}\mathbb{E}[h_i] \\
&= \frac{1}{\sqrt{2\pi}}.
\end{aligned}$$

The variance of $\tilde{h}_i$ is:

$$\begin{aligned}
\mathrm{Var}(\tilde{h}_i) &= \mathbb{E}[\tilde{h}_i^2] - \mathbb{E}[\tilde{h}_i]^2 \\
&= \mathbb{E}\left[\left(\frac{1}{1-p}\right)^2 \cdot m_i^2 \cdot h_i^2\right] - \left(\mathbb{E}\left[\frac{1}{1-p} \cdot m_i \cdot h_i\right]\right)^2 \\
&= \left(\frac{1}{1-p}\right)^2 \left(\mathbb{E}[m_i^2]\mathbb{E}[h_i^2] - \mathbb{E}[m_i]^2 \cdot \mathbb{E}[h_i]^2\right) \\
&= \frac{1}{1-p} \cdot \mathbb{E}[h_i^2] - \mathbb{E}[h_i]^2 \\
&= \frac{1}{1-p} \cdot \frac{1}{2} - \frac{1}{2\pi}.
\end{aligned}$$

Hence,

$$\therefore \mathrm{Var}(\tilde{\mathbf{h}}) = \left(\frac{1}{1-p} \cdot \frac{1}{2} - \frac{1}{2\pi}\right) I_M$$

1.3

Solution:

For one unit, the expectation that it is kept is $1-p$, then given M units the expectation would be $M \cdot (1-p)$ units kept. For each unit, we have a probability $1-p$ that it is kept, so if we compute the probabilty that k units are kept, $P(\text{kept} = k)$,

$$P(\text{kept} = k) = \binom{M}{k} \cdot (1-p)^k \cdot p^{M-k}$$

hence, a binomial distribution with probability $1-p$.

1.4

Solution:

First let $M(1-p)=\alpha$, in other words $p = 1-\frac{\alpha}{M}$,

$$\lim_{M\to\infty}\binom{M}{k}\cdot(1-p)^k\cdot p^{M-k} = \lim_{M\to\infty}\frac{M(M-1)\cdot(M-k+1)}{k!}(1-p)^k(1-\frac{\alpha}{M})^{M-k}$$
$$= \lim_{M\to\infty}\frac{(M\cdot(1-p))\cdot((M-1)\cdot(1-p))\cdots((M-k+1)(1-p))}{k!}\cdot(1-\frac{\alpha}{M})^{M-k}$$
$$= \lim_{M\to\infty}\frac{(M\cdot(1-p))\cdot((M-1)\cdot(1-p))\cdots((M-k+1)(1-p))}{k!}\cdot(1-\frac{\alpha}{M})^{\frac{M}{\alpha}}\cdot(1-\frac{\alpha}{M})^{-k}$$
$$= \frac{\alpha^k}{k!}e^{-\alpha} = \frac{(M(1-p))^k}{k!}e^{-M(1-p)}$$

It becomes a Poisson distribution with parameter $M(1-p)$.

1.5

Solution:

Let's say that we want to keep x units and get the probability distribution. We then want to sum all the probabilities of keeping x units for all M. Thus we want,

$$P(x \text{ units kept}) = \sum_{M=x}^{\infty} P(x \text{ units kept} \cap M \text{ units})$$

Let's do the math!

$$P(x \text{ units kept} \cap M \text{ units}) = \frac{\lambda^M e^{-\lambda}}{M!}\cdot\binom{M}{x}\cdot(1-p)^x\cdot p^{M-x}$$
$$= \frac{\lambda^M e^{-\lambda}}{M!}\cdot\frac{M!}{(M-x)!\cdot x!}\cdot(1-p)^x\cdot p^{M-x}$$
$$= \frac{e^{-\lambda}}{x!}\cdot(1-p)^x\cdot\frac{\lambda^M\cdot p^{M-x}}{(M-x)!}.$$

Now, the probability of x units being kept is:

$$P(x \text{ units kept}) = \sum_{M=x}^{\infty}\frac{e^{-\lambda}}{x!}\cdot(1-p)^x\cdot\frac{\lambda^M\cdot p^{M-x}}{(M-x)!}$$
$$= \frac{e^{-\lambda}}{x!}\cdot(1-p)^x\cdot\sum_{M=x}^{\infty}\frac{\lambda^M\cdot p^{M-x}}{(M-x)!}.$$

Let $M' = M-x$. Then:

$$P(x \text{ units kept}) = \frac{e^{-\lambda}}{x!}\cdot(1-p)^x\cdot\sum_{M'=0}^{\infty}\frac{\lambda^{M'+x}\cdot p^{M'}}{M'!}$$
$$= \frac{e^{-\lambda}}{x!}\cdot(1-p)^x\cdot\lambda^x\cdot\sum_{M'=0}^{\infty}\frac{(\lambda p)^{M'}}{M'!}.$$

Since $\sum_{M'=0}^{\infty} \frac{(\lambda p)^{M'}}{M'!} = e^{\lambda p}$, we get:

$$\begin{aligned} P(x \text{ units kept}) &= \frac{e^{-\lambda}}{x!} \cdot (1-p)^x \cdot \lambda^x \cdot e^{\lambda p} \\ &= \frac{e^{-\lambda(1-p)} \cdot \{\lambda(1-p)\}^x}{x!}. \end{aligned}$$

Therefore, the number of kept units follows a Poisson distribution with parameter $\lambda(1-p)$.

2 Problem 2

A 2.1

To insert inline equations, use $\frac{1}{1-p}$. To bold characters in equations, type $\mathbf{b}$. For script style letters, use $\mathcal{N}$.

For a displayed equation, you can use

$$P(k) = \binom{N}{k} p^k (1-p)^{N-k} \tag{1}$$

Or you can also use

$$f'(x) = \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}$$

Also for series equations, you can use

$$\begin{aligned} A &= 2\int_{-r}^{r} \sqrt{r^2 - x^2}\, dx \qquad (2) \\ &= 2r^2 \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^2\theta \, d\theta \qquad (3) \\ &= \pi r^2 \qquad (4) \end{aligned}$$

For matrices, format as follows:

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ \vdots & \vdots & & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \tag{5}$$

For more math symbols, check Wiki, LATEX Mathematical Symbols, Google, or ask Chat-GPT.

A 2.2

If you want to insert a picture:

A 2.3

Figure 1: Caption for the image.

To highlight words in a different color, you can use textcolor to turn something blue. You can also define custom commands for frequent use.