Theory and Practice of Deep Learning

Theory HW #01

Nigel Chan, 1002027

Task 1: What is the distribution of (x,y).

$$p(y,x) = p(y,x,\{c(x)=1\} \text{ or } \{c(x)=2\})$$
$$= p(y,x,c(x)=1) + p(y,x,c(x)=2)$$

using conditional probability where $P(A,B) = P(A|B)P(B)$,

$$p(y,x,c(x)=1) + p(y,x,c(x)=2)$$
$$= p(y|x,c(x)=1)\,p(x,c(x)=1) + p(y|x,c(x)=2)\,p(x,c(x)=2)$$
$$= p(y|x,c(x)=1)\,p(x|c(x)=1)\,p(c(x)=1) +$$
$$p(y|x,c(x)=2)\,p(x|c(x)=2)\,p(c(x)=2)$$

given that the distribution of datapoints is continuous,

$$\therefore\ p(y,x) = p(y|x,c(x)=1)\,f(x|c(x)=1)\,p(c(x)=1) +$$
$$p(y|x,c(x)=2)\,f(x|c(x)=2)\,p(c(x)=2)$$

Given the following,

$$p(y=0|x, c(x)=1) = 0.2,$$
$$p(y=0|x, c(x)=2) = 0.7,$$
$$P(C=1) = P(C=2) = 0.5$$

We plug in the values into $p(y,x)$ formula for $y=0$.

$$P(0,x) = 0.2 \times f(x|c(x)=1) \times 0.5 +$$
$$0.7 \times f(x|c(x)=2) \times 0.5$$
$$= 0.1f(x|c(x)=1) + 0.35f(x|c(x)=2)$$

For $y=1$, we are given the following

$$p(y=0|x, c(x)=1) + p(y=1|x, c(x)=1),$$
$$p(y=0|x, c(x)=2) + p(y=1|x, c(x)=2)$$

Hence, we deduce that:

$$p(y=1|x, c(x)=1) = 0.8,$$
$$p(y=1|x, c(x)=2) = 0.3$$

Plugging in the values for $y=1$,

$$P(1,x) = 0.8 \times f(x|c(x)=1) \times 0.5 +$$
$$0.3 \times f(x|c(x)=2) \times 0.5$$
$$= 0.4f(x|c(x)=1) + 0.15f(x|c(x)=2)$$

Task2: A matrix can be seen as just a vector

1) Prove that A.B can be written as an inner product of two vectors. ~~an~~ as in Equation 1

Equation 1 : $v \cdot w = \sum_{d=1}^{D} v_d w_d$

Given ~~that~~ the matrices $A, B \in \mathbb{R}^{(m,l)}$,

where $A \cdot B := tr(A^T B)$,

$tr(Z) := \sum_i Z_{ii}$

Assuming A and B have the same dimensions

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1m} \\ a_{21} & \ddots & & & \vdots \\ \vdots & & & & \\ a_{l1} & \cdots & & & a_{lm} \end{bmatrix}, \quad B = \begin{bmatrix} b_{11} & b_{12} & b_{13} & \cdots & b_{1m} \\ b_{21} & \ddots & & & \\ \vdots & & & & \\ b_{l1} & \cdots & & & b_{lm} \end{bmatrix}$$

With the given, we can represent A and B in the vector forms $\vec{a}$ and $\vec{b}$ respectively.

$$\vec{a} = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{lm} \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} b_{11} \\ b_{21} \\ \vdots \\ b_{lm} \end{bmatrix}$$

Let $D = l \times m$

$$\therefore A \cdot B := tr(A^T B) = \vec{a} \cdot \vec{b} = \sum_{d=1}^{D} a_d b_d$$

2) Prove that A.B = B.A.

As proven in the previous section, A and B can be seen as vectors, we can apply the commutative preperty.

e.g.

$$\vec{a}.\vec{b} = a_1b_1 + a_2b_2 + \dots a_lb_l$$
$$= b_1a_1 + b_2a_2 + \dots b_la_l$$
$$= \vec{b}.\vec{a}.$$

$\therefore$ A.B = B.A

3) We know that for every inner product it holds $V.W = \|V\|\|W\| \cos\angle(V,W)$.
So what is the cosine angle between $\begin{pmatrix} 1 & -2 \\ -2 & 3 \end{pmatrix}$ and $\begin{pmatrix} -1 & 1 \\ 0 & -2 \end{pmatrix}$?
Which angles can constitute the computed cosine angle?

Let $\vec{x}$ and $\vec{y}$ be the vector representations of $\begin{pmatrix} 1 & -2 \\ -2 & 3 \end{pmatrix}$ and $\begin{pmatrix} -1 & 1 \\ 0 & -2 \end{pmatrix}$ respectively

$$\vec{x} = \begin{bmatrix} 1 \\ -2 \\ -2 \\ 3 \end{bmatrix} \qquad \vec{y} = \begin{bmatrix} -1 \\ 0 \\ 1 \\ -2 \end{bmatrix}$$

$$\vec{x}.\vec{y} = \|\vec{x}\|\|\vec{y}\| \cos\angle(\vec{x},\vec{y})$$

$$\cos\angle(\vec{x},\vec{y}) = \frac{\vec{x}.\vec{y}}{\|\vec{x}\|\|\vec{y}\|}$$
$$= \frac{(-1 + 0 - 2 - 6)}{\sqrt{18} \times \sqrt{6}}$$
$$= -0.86603$$

$$\therefore \angle(\vec{x},\vec{y}) = \cos^{-1}(-0.86603) = 150.0005$$
$$\approx 150^\circ$$