CSE 493G1/599G1: Deep Learning

Solutions for Section 1: Fundamentals

Welcome to section, we're happy you're here!

Reference Material

Rules of Broadcasting from Jake VanderPlas' *Python Data Science Handbook*:

(1) If the two arrays differ in their number of dimensions, the shape of the one with fewer dimensions is padded with ones on its leading (left) side.

(2) If the shape of the two arrays does not match in any dimension, the array with shape equal to 1 in that dimension is stretched to match the other shape.

(3) If in any dimension the sizes disagree and neither is equal to 1, an error is raised.

Chain Rule for One Independent Variable:

Let $z = f(x, y)$ be a differentiable function. Further suppose that x and y are themselves differentiable functions of t, in other words $x = x(t)$ and $y = y(t)$. Then,

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}$$

Chain Rule for Two Independent Variables:

Let $z = f(x, y)$ be a differentiable function, where x and y are themselves differentiable functions of a and b. In other words, $x = x(a, b)$ and $y = y(a, b)$. Then,

$$\frac{dz}{da} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial a} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial a}$$

and

$$\frac{dz}{db} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial b} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial b}$$

Generalized Chain Rule:

Let $w = f(x_1, x_2, \ldots, x_m)$ be a differentiable function of m independent variables, and let $x_i = x_i(t_1, t_2, \ldots, t_n)$ be a differentiable function of n independent variables. Then,

$$\frac{dw}{dt_j} = \frac{\partial w}{\partial x_1}\frac{\partial x_1}{\partial t_j} + \frac{\partial w}{\partial x_2}\frac{\partial x_2}{\partial t_j} + \ldots + \frac{\partial w}{\partial x_m}\frac{\partial x_m}{\partial t_j}$$

for any $j \in 1, 2, \ldots, n$.

1. Dimension: Impossible

Determine if NumPy allows the addition of the following pairs of arrays, and if applicable determine what the result's dimensions will be.

(a) Where `x.shape` is $(2,)$ and `y.shape` is $(2, 1)$

Solution:

Yes. $(2, 2)$.

(b) Where `x.shape` is $(4,)$ and `y.shape` is $(4, 1, 1)$

Solution:

Yes. $(4, 1, 4)$.

(c) Where `x.shape` is $(4, 2)$ and `y.shape` is $(2, 4, 1)$

Solution:

Yes. $(2, 4, 2)$.

(d) Where `x.shape` is $(8, 3)$ and `y.shape` is $(2, 8, 1)$

Solution:

Yes. $(2, 8, 3)$.

(e) Where `x.shape` is $(6, 5, 3)$ and `y.shape` is $(6, 5)$

Solution:

No. However, if we changed `y.shape` to be $(6, 5, 1)$, then we would get a valid operation that results in an array of shape $(6, 5, 3)$. This could be achieved in NumPy by calling either `x + y[:, :, None]` or `x + y[:, :, np.newaxis]` instead of `x + y`.

2. The More (Derivatives) The Merrier

(a) Let $z = 2x + y$, with $x = \ln(t)$ and $y = \frac{1}{3}t^3$. Find $\frac{dz}{dt}$.

Solution:

$$\begin{aligned}
\frac{dz}{dt} &= \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt} \\
&= \frac{\partial}{\partial x}\Big(2x+y\Big) \cdot \frac{\partial}{\partial t}\Big(\ln(t)\Big) + \frac{\partial}{\partial y}\Big(2x+y\Big)\frac{\partial}{\partial t}\left(\frac{1}{3}t^3\right) && \text{Chain Rule} \\
&= 2 \cdot \frac{1}{t} + 1 \cdot t^2 && \text{Solve Partial Derivatives} \\
&= t^2 + \frac{2}{t}
\end{aligned}$$

(b) Let $z = x^2y - y^2$ where $x = t^2$ and $y = 2t$. Find $\frac{dz}{dt}$. Your answer should be in terms of t.

Solution:

$$\begin{aligned}
\frac{dz}{dt} &= \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt} && \text{Chain Rule} \\
&= (2xy)(2t) + (x^2 - 2y)(2) && \text{Substitute Partial Derivatives} \\
&= \big(2(t^2)(2t)\big)\big(2t\big) + \big((t^2)^2 - 2(2t)\big)\big(2\big) && \text{Definitions of } x \text{ and } y \\
&= (4t^3)(2t) + 2(t^4 - 4t) \\
&= 8t^4 + 2t^4 - 8t \\
&= 10t^4 - 8t
\end{aligned}$$

(c) Let $z = 3x^2 - 2xy + y^2$. Also let $x = 3a + 2b$ and $y = 4a - b$. Find $\frac{\partial z}{\partial a}$ and $\frac{\partial z}{\partial b}$.

Solution:

$$\begin{aligned}
\frac{\partial z}{\partial a} &= \frac{\partial z}{\partial x}\frac{\partial x}{\partial a} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial a} && \text{Chain Rule} \\
&= (6x - 2y)(3) + (-2x + 2y)(4) && \text{Substitute Partial Derivatives} \\
&= 18x - 6y - 8x + 8y \\
&= 10x + 2y \\
&= 10(3a + 2b) + 2(4a - b) && \text{Definitions of } x \text{ and } y \\
&= 30a + 20b + 8a - 2b \\
&= 38a + 18b
\end{aligned}$$

$$\begin{aligned}
\frac{\partial z}{\partial b} &= \frac{\partial z}{\partial x}\frac{\partial x}{\partial b} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial b} && \text{Chain Rule} \\
&= (6x - 2y)(2) + (-2x + 2y)(-1) && \text{Substitute Partial Derivatives} \\
&= 12x - 4y + 2x - 2y
\end{aligned}$$

$$
\begin{aligned}
&= 14x - 6y \\
&= 14(3a + 2b) - 6(4a - b) && \text{Definitions of } x \text{ and } y \\
&= 42a + 28b - 24a + 6b \\
&= 18a + 34b
\end{aligned}
$$

(d) Let $w = f(x, y, z)$, $x = x(t, u, v)$, $y = y(t, u, v)$ and $z = z(t, u, v)$. Find the formula for $\frac{\partial w}{\partial t}$.

Solution:

$$\frac{\partial w}{\partial t} = \frac{\partial w}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial w}{\partial y}\frac{\partial y}{\partial t} + \frac{\partial w}{\partial z}\frac{\partial z}{\partial t}$$