DSC 40A - Discussion 04 - Matrix Calculus

February 11, 2020

Problem 1.

$$\frac{d(||\vec{x}||^2)}{d\vec{x}} = ?$$

$$\vec{x} \in R^d$$

Solution: First, we know that $f(\vec{x}) = ||\vec{x}||^2 = \sum_{i=1}^{d} x_i^2$. Now, let's find the partial derivative $\frac{\partial f(\vec{x})}{\partial x_k}$ and then construct the gradient vector.

$$\begin{aligned}\frac{\partial f(\vec{x})}{\partial x_k} &= \frac{\partial(\sum_{i=1}^{d} x_i^2)}{\partial x_k} \\ &= 2x_k\end{aligned}$$

Therefore, $\frac{df(\vec{x})}{d\vec{x}} = \begin{bmatrix} \partial f(\vec{x})/\partial x_1 \\ \partial f(\vec{x})/\partial x_2 \\ \cdots \\ \partial f(\vec{x})/\partial x_d \end{bmatrix} = \begin{bmatrix} 2x_1 \\ 2x_2 \\ \cdots \\ 2x_d \end{bmatrix} = 2\vec{x}$

Problem 2.

$$\frac{d(\vec{a}^T\vec{x})^3}{d\vec{x}} = ?$$

$$\vec{x} \in R^d, \vec{a} \in R^d$$

Solution: First, as in the previous question, let's express the matrix notation in terms of sums.

$$f(\vec{x}) = (\vec{a}^T\vec{x})^3 = (\sum_{i=1}^{d} a_i x_i)^3$$

Now, we will find $\frac{\partial f(\vec{x})}{\partial x_k}$ and then construct the gradient vector.

$$\begin{aligned}\frac{\partial f(\vec{x})}{\partial x_k} &= \frac{\partial((\sum_{i=1}^{d} a_i x_i)^3)}{\partial x_k} \\ &= 3(\sum_{i=1}^{d} a_i x_i)^2 a_k && \text{(Chain Rule)} \\ &= 3c^2 a_k && \text{(let } c = \textstyle\sum_{i=1}^{d} a_i x_i \text{ as the expression does not depend on } k\text{)}\end{aligned}$$

Therefore, $\dfrac{df(\vec{x})}{d\vec{x}} = \begin{bmatrix} \partial f(\vec{x})/\partial x_1 \\ \partial f(\vec{x})/\partial x_2 \\ \dots \\ \partial f(\vec{x})/\partial x_d \end{bmatrix} = \begin{bmatrix} 3c^2 a_1 \\ 3c^2 a_2 \\ \dots \\ 3c^2 a_d \end{bmatrix} = 3c^2\vec{a}$.

Finally, let's substitute in $c = \sum_{i=1}^{d} a_i x_i = \vec{a}^T\vec{x}$. We get

$$\frac{df(\vec{x})}{d\vec{x}} = 3(\vec{a}^T\vec{x})^2\vec{a}$$

Problem 3.

$$\frac{d||X\vec{w} - \vec{y}||^2}{d\vec{w}} = ?$$

$$X \in R^{n\times(d+1)}, \vec{y} \in R^n, \vec{w} \in R^{d+1}$$

Solution:

$$\begin{aligned} ||X\vec{w} - \vec{y}||^2 &= (X\vec{w} - \vec{y})^T(X\vec{w} - \vec{y}) \\ &= (\vec{w}^T X^T - \vec{y}^T)(X\vec{w} - \vec{y}) \\ &= \vec{w}^T X^T X\vec{w} - \vec{w}^T X^T \vec{y} - \vec{y}^T X\vec{w} + \vec{y}^T\vec{y} \\ &= \vec{w}^T X^T X\vec{w} - \vec{y}^T X\vec{w} - \vec{y}^T X\vec{w} + \vec{y}^T\vec{y} \\ &\qquad\qquad (\text{as} \quad \vec{w}^T X^T \vec{y} \text{ is a scalar, it is equal to its transpose}) \\ &= \vec{w}^T X^T X\vec{w} - 2\vec{y}^T X\vec{w} + \vec{y}^T\vec{y} \end{aligned}$$

Now, we can take the gradient.

$$\begin{aligned} \frac{d||X\vec{w} - \vec{y}||^2}{d\vec{w}} &= \frac{d(\vec{w}^T X^T X\vec{w} - 2\vec{y}^T X\vec{w} + \vec{y}^T\vec{y})}{d\vec{w}} \\ &= \frac{d\vec{w}^T X^T X\vec{w}}{d\vec{w}} - \frac{2d\vec{y}^T X\vec{w}}{d\vec{w}} + \frac{d\vec{y}^T\vec{y}}{d\vec{w}} \end{aligned}$$

We saw in class that $\dfrac{d\vec{w}^T X^T X\vec{w}}{d\vec{w}} = 2(X^T X)\vec{w}$

For the second term, let $\vec{y}^T X = \vec{a}^T$, a vector independent of $\vec{w}$. We also saw in class that $\dfrac{d\vec{a}^T\vec{w}}{d\vec{w}} = \vec{a}$. So, $\dfrac{d\vec{y}^T X\vec{w}}{d\vec{w}} = (\vec{y}^T X)^T = X^T\vec{y}$. The third term is just the zero vector.

Hence,

$$\begin{aligned} \frac{d\vec{w}^T X^T X\vec{w}}{d\vec{w}} - \frac{2d\vec{y}^T X\vec{w}}{d\vec{w}} + \frac{d\vec{y}^T\vec{y}}{d\vec{w}} &= 2(X^T X)\vec{w} - 2X^T\vec{y} + \vec{0} \\ \frac{d||X\vec{w} - \vec{y}||^2}{d\vec{w}} &= 2(X^T X)\vec{w} - 2X^T\vec{y} \end{aligned}$$

Problem 4.

$$\frac{d(\vec{y}^T X \vec{w})}{d\vec{w}} = ?$$

$$X \in R^{n \times (d+1)}, \vec{y} \in R^n, \vec{w} \in R^{d+1}$$

Solution: We will not solve this question completely but give you a starting point. Let $f(\vec{w}) = \vec{y}^T X \vec{w}$. Our strategy is first converting the matrix notation into summations, then finding $\frac{\partial f(\vec{w})}{\partial w_k}$ and finally constructing the gradient vector.

$$\begin{aligned} f(\vec{w}) &= \vec{y}^T X \vec{w} \\ &= \sum_{i=1}^{d+1} (\vec{y}^T X)_i w_i \\ &= \sum_{i=1}^{d+1} (\sum_{j=?}^{?} ?) w_i \end{aligned}$$

After filling in the blanks, find $\frac{\partial f(\vec{w})}{\partial w_k}$ and then construct $\frac{d(\vec{y}^T X \vec{w})}{d\vec{w}}$