Linear and Quadratic Functions - COMP24112

Tingting Mu TINGTINGMU@MANCHESTER.AC.UK
Department of Computer Science
University of Manchester
Manchester M13 9PL, UK

Editor: NA

1. Linear and Quadratic Functions

1.1 Linear Functions

Let $\boldsymbol{w} \in R^n$ denote a known n-dimensional vector in column. For an input column vector $\boldsymbol{x} \in R^n$, the following function

$$f(\boldsymbol{x}) = \sum_{i=1}^{n} w_i x_i = \boldsymbol{w}^T \boldsymbol{x} \tag{1}$$

is a ***linear function*** of $\boldsymbol{x}$. The partial derivative of this function is

$$\frac{\partial f(\boldsymbol{x})}{\partial x_i} = \frac{\partial}{\partial x_i}\left(\sum_{i=1}^{n} w_i x_i\right) = w_i, \text{ for } i = 1, 2, \ldots n. \tag{2}$$

The gradient of $f(\boldsymbol{x})$ with respect to the input column vector $\boldsymbol{x}$ is

$$\nabla_{\boldsymbol{x}} f(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = \boldsymbol{w}. \tag{3}$$

Note that the function $f(\boldsymbol{x})$ can also be written as $f(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{w}$, and its gradient with respect to $\boldsymbol{x}$ is $\boldsymbol{w}$.

1.2 Quadratic Functions

Let $\mathbf{A} = [a_{ij}]$ denote a known $n \times n$ square matrix. For an input column vector $\boldsymbol{x} \in R^n$, the following function

$$f(\boldsymbol{x}) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j = \boldsymbol{x}^T \mathbf{A} \boldsymbol{x} \tag{4}$$

is a ***quadratic function*** of $\boldsymbol{x}$. To compute the partial derivative of this function with respect to an element x_k in the input vector ($k = 1, 2, \ldots n$), we consider separately the terms that contain x_k and x_k^2, also the terms that do not contain x_k. This gives

$$\begin{aligned}
\frac{\partial f(\boldsymbol{x})}{\partial x_k} &= \frac{\partial}{\partial x_k}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j\right) \\
&= \frac{\partial}{\partial x_k}\left(a_{kk}x_k^2 + \sum_{i\neq k} a_{ik}x_i x_k + \sum_{j\neq k} a_{kj}x_k x_j + \sum_{i\neq k}\sum_{j\neq k} a_{ij}x_i x_j\right) \\
&= 2a_{kk}x_k + \sum_{i\neq k} a_{ik}x_i + \sum_{j\neq k} a_{kj}x_j \\
&= \sum_{i=1}^{n} a_{ik}x_i + \sum_{j=1}^{n} a_{kj}x_j \qquad (5) \\
&= \mathbf{A}_{:,k}^T\boldsymbol{x} + \mathbf{A}_{k,:}\boldsymbol{x}. \qquad (6)
\end{aligned}$$

The gradient of $f(\boldsymbol{x})$ with respect to $\boldsymbol{x}$ is

$$\nabla_{\boldsymbol{x}} f(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{:,1}^T\boldsymbol{x} + \mathbf{A}_{1,:}\boldsymbol{x} \\ \mathbf{A}_{:,2}^T\boldsymbol{x} + \mathbf{A}_{2,:}\boldsymbol{x} \\ \vdots \\ \mathbf{A}_{:,n}^T\boldsymbol{x} + \mathbf{A}_{n,:}\boldsymbol{x} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{:,1}^T\boldsymbol{x} \\ \mathbf{A}_{:,2}^T\boldsymbol{x} \\ \vdots \\ \mathbf{A}_{:,n}^T\boldsymbol{x} \end{bmatrix} + \begin{bmatrix} \mathbf{A}_{1,:}\boldsymbol{x} \\ \mathbf{A}_{2,:}\boldsymbol{x} \\ \vdots \\ \mathbf{A}_{n,:}\boldsymbol{x} \end{bmatrix} = \mathbf{A}^T\boldsymbol{x} + \mathbf{A}\boldsymbol{x}. \qquad (7)$$

A special case of the quadratic function is $f(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{x}$, where $\mathbf{A}$ is an identity matrix. Its gradient with respect to $\boldsymbol{x}$ is therefore $\nabla_{\boldsymbol{x}} f(\boldsymbol{x}) = \mathbf{I}^T\boldsymbol{x} + \mathbf{I}\boldsymbol{x} = 2\boldsymbol{x}$.