Dodatek 2.A Analiza matematyczna

2.A.1 Różniczkowanie funkcji skalarnej względem wektora zmiennych

Różniczkujemy funkcję skalarną wielu zmiennych $f(\boldsymbol{x}) = f(x_1, \ldots, x_K)$ względem wektora kolumnowego $x = \begin{bmatrix} x_1 \\ \vdots \\ x_K \end{bmatrix}$. Z uzyskanych pochodnych tworzymy wektor pochodnych:

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}.$$

Taki wektor pochodnych nazywamy w analizie matematycznej gradientem i często oznaczmy jako $\boldsymbol{G}(\boldsymbol{x})$. Licząc pochodną względem wektora wierszowego $\boldsymbol{x}'$ uzyskujemy wektor wierszowy złożony z tych samych pochodnych:

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}'} = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix}$$

Pochodna iloczynu wektora wierszowego $\boldsymbol{w}'$ i wektora kolumnowego $\boldsymbol{x}$:

$$\frac{\partial \boldsymbol{w}'\boldsymbol{x}}{\partial \boldsymbol{x}} = \frac{\partial\left(w_1 x_1 + \ldots + w_n x_n\right)}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial(w_1x_1+\ldots+w_nx_n)}{\partial x_1} \\ \vdots \\ \frac{\partial(w_1x_1+\ldots+w_nx_n)}{\partial x_n} \end{bmatrix} = \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} = \boldsymbol{w}.$$

Oczywiście:

$$\frac{\partial \boldsymbol{w}'\boldsymbol{x}}{\partial \boldsymbol{x}'} = \begin{bmatrix} w_1 & \cdots & w_n \end{bmatrix} = \boldsymbol{w}'.$$

2.A.2 Różniczkowanie funkcji wektorowej względem wektora zmiennych

Wektor $\boldsymbol{f}_{(m\times 1)}$ jest wielowymiarową funkcją wielu zmiennych:

$$\boldsymbol{f}\left(\boldsymbol{x}\right) = \begin{bmatrix} f_1\left(x_1,\ldots,x_n\right) \\ \vdots \\ f_m\left(x_1,\ldots,x_n\right) \end{bmatrix}.$$

Pochodna takiej funkcji ma postać macierzy:

$$\frac{\partial \boldsymbol{f}\left(\boldsymbol{x}\right)}{\partial \boldsymbol{x}'} = \left[\frac{\partial f_i\left(\boldsymbol{x}\right)}{\partial x_j}\right]_{(m\times n)} = \begin{bmatrix} \frac{\partial f_1}{\partial \boldsymbol{x}'} \\ \vdots \\ \frac{\partial f_m}{\partial \boldsymbol{x}'} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \ldots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \\ \frac{\partial f_m}{\partial x_1} & \ldots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}.$$

Możliwe jest także różniczkowanie funkcji w postaci wektora wierszowego względem wektora kolumnowego zmiennych:

$$\frac{\partial \boldsymbol{f}'\left(\boldsymbol{x}\right)}{\partial \boldsymbol{x}} = \left[\frac{\partial f_j\left(\boldsymbol{x}\right)}{\partial x_i}\right]_{(n\times m)} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \ldots & \frac{\partial f_m}{\partial x_1} \\ \vdots & & \\ \frac{\partial f_1}{\partial x_n} & \ldots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}.$$

Zauważmy, że relacja między pochodną $\boldsymbol{f}\left(\boldsymbol{x}\right)$ względem wektora wierszowego $\boldsymbol{x}'$ i pochodną $\boldsymbol{f}'\left(\boldsymbol{x}\right)$, policzoną względem wektora kolumnowego $\boldsymbol{x}$ jest następująca:

$$\frac{\partial \boldsymbol{f}'\left(\boldsymbol{x}\right)}{\partial \boldsymbol{x}} = \left[\frac{\partial f_j\left(\boldsymbol{x}\right)}{\partial x_i}\right]_{(n\times m)} = \left(\frac{\partial \boldsymbol{f}\left(\boldsymbol{x}\right)}{\partial \boldsymbol{x}'}\right)'. \tag{2.A.1}$$

Macierz drugich pochodnych funkcji skalarnej powstaje jako pochodna funkcji wektorowej (gradientu):

$$\frac{\partial^2 f(\boldsymbol{x})}{\partial \boldsymbol{x} \partial \boldsymbol{x}'} = \frac{\partial \begin{bmatrix} \frac{\partial f(\boldsymbol{x})}{\partial x_1} \\ \vdots \\ \frac{\partial f(\boldsymbol{x})}{\partial x_n} \end{bmatrix}}{\partial \boldsymbol{x}'} = \frac{\partial \boldsymbol{G}(\boldsymbol{x})}{\partial \boldsymbol{x}'} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & & \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{bmatrix}.$$

Macierz drugich pochodnych nazywamy także Hessianem $\boldsymbol{H}(\boldsymbol{x})$. Ponieważ:

$$\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i \partial x_j} = \frac{\partial^2 f(\boldsymbol{x})}{\partial x_j \partial x_i},$$

więc Hessian jest macierzą symetryczną.

PRZYKŁAD 2.12 **Pochodna funcji liniowej wielu zmiennych**

Policzmy pochodną funkcji $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$. Macierz $\boldsymbol{A}$ możemy zapisać jako $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{w}_1' \\ \vdots \\ \boldsymbol{w}_m' \end{bmatrix}$, gdzie $\boldsymbol{w}_i'$ jest i-tym wierszem macierzy $\boldsymbol{A}$. Z kolei $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} = \begin{bmatrix} \boldsymbol{w}_1'\boldsymbol{x} \\ \vdots \\ \boldsymbol{w}_m'\boldsymbol{x} \end{bmatrix}$. Pochodna:

$$\frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}'} = \frac{\partial \boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}'} = \begin{bmatrix} \frac{\partial \boldsymbol{w}_1'\boldsymbol{x}}{\partial \boldsymbol{x}} \\ \vdots \\ \frac{\partial \boldsymbol{w}_m'\boldsymbol{x}}{\partial \boldsymbol{x}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{w}_1' \\ \vdots \\ \boldsymbol{w}_m' \end{bmatrix} = \boldsymbol{A}. \tag{2.A.2}$$

2.A.3 Różniczkowanie iloczynu skalarnego funkcji wektorowych względem wektora zmiennych

W przypadku liczenia pochodnych funkcji wektorowych obowiązują podobne zasady, jak w przypadku liczenia zwykłych pochodnych. Rozważmy przykład iloczynu skalarnego funkcji $\boldsymbol{g}(\boldsymbol{x})$ i $\boldsymbol{f}(\boldsymbol{x})$:

$$h(\boldsymbol{x}) = \boldsymbol{g}'(\boldsymbol{x})\boldsymbol{f}(\boldsymbol{x}) = \sum_{i=1}^{m} g_i(\boldsymbol{x}) f_i(\boldsymbol{x}).$$

Pochodna $h(\boldsymbol{x})$ względem $\boldsymbol{x}$ jest równa:

$$\frac{\partial h(\boldsymbol{x})}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial h(\boldsymbol{x})}{\partial x_1} \\ \vdots \\ \frac{\partial h(\boldsymbol{x})}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \frac{\partial \sum_{i=1}^{m} g_i(\boldsymbol{x}) f_i(\boldsymbol{x})}{\partial x_1} \\ \vdots \\ \frac{\partial \sum_{i=1}^{m} g_i(\boldsymbol{x}) f_i(\boldsymbol{x})}{\partial x_n} \end{bmatrix}.$$

Stosując znany wzór na pochodną iloczynu uzyskujemy:

$$\frac{\partial g_i(\boldsymbol{x}) f_i(\boldsymbol{x})}{\partial x_k} = \frac{\partial g_i(\boldsymbol{x})}{\partial x_k} f_i(\boldsymbol{x}) + g_i(\boldsymbol{x}) \frac{\partial f_i(\boldsymbol{x})}{\partial x_k}.$$

Wstawiając ten wynik do poprzedniego wzoru:

$$\frac{\partial h(\boldsymbol{x})}{\partial \boldsymbol{x}} = \begin{bmatrix} \sum_{i=1}^{m} \frac{\partial g_i(\boldsymbol{x})}{\partial x_1} f_i(\boldsymbol{x}) + \sum_{i=1}^{m} \frac{\partial f_i(\boldsymbol{x})}{\partial x_1} g_i(\boldsymbol{x}) \\ \vdots \\ \sum_{i=1}^{m} \frac{\partial g_i(\boldsymbol{x})}{\partial x_n} f_i(\boldsymbol{x}) + \sum_{i=1}^{m} \frac{\partial f_i(\boldsymbol{x})}{\partial x_n} g_i(\boldsymbol{x}) \end{bmatrix}.$$

Z definicji iloczynu wektorów:

$$\sum_{i=1}^{m} \frac{\partial g_i(\boldsymbol{x})}{\partial x_k} f_i(\boldsymbol{x}) = \frac{\partial \boldsymbol{g}'(\boldsymbol{x})}{\partial x_k} \boldsymbol{f}(\boldsymbol{x}),$$
$$\sum_{i=1}^{m} \frac{\partial f_i(\boldsymbol{x})}{\partial x_k} g_i(\boldsymbol{x}) = \frac{\partial \boldsymbol{f}'(\boldsymbol{x})}{\partial x_k} \boldsymbol{g}(\boldsymbol{x}),$$

a więc:

$$\frac{\partial h(\boldsymbol{x})}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial \boldsymbol{g}'(\boldsymbol{x})}{\partial x_1} \boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{g}'(\boldsymbol{x}) \frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial x_1} \\ \vdots \\ \frac{\partial \boldsymbol{g}'(\boldsymbol{x})}{\partial x_n} \boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{g}'(\boldsymbol{x}) \frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial x_n} \end{bmatrix} = \frac{\partial \boldsymbol{g}'(\boldsymbol{x})}{\partial \boldsymbol{x}} \boldsymbol{f}(\boldsymbol{x}) + \frac{\partial \boldsymbol{f}'(\boldsymbol{x})}{\partial \boldsymbol{x}} \boldsymbol{g}(\boldsymbol{x}).$$

Możemy teraz sformułować następujące twierdzenie, analogiczne do standardowego twierdzenia o pochodnej iloczynu:

TWIERDZENIE 2.13 *Dla m-wymiarowych różniczkowalnych wektorowych funkcji wielu zmiennych $\boldsymbol{g}(\boldsymbol{x})$ i $\boldsymbol{f}(\boldsymbol{x})$ oraz funkcji skalarnej $h(\boldsymbol{x}) = \boldsymbol{g}'(\boldsymbol{x})\, \boldsymbol{f}(\boldsymbol{x})$:*

$$\frac{\partial h(\boldsymbol{x})}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{g}'(\boldsymbol{x})}{\partial \boldsymbol{x}} \boldsymbol{f}(\boldsymbol{x}) + \frac{\partial \boldsymbol{f}'(\boldsymbol{x})}{\partial \boldsymbol{x}} \boldsymbol{g}(\boldsymbol{x}) \tag{2.A.3}$$

PRZYKŁAD 2.14 **Pochodna formy kwadratowej**

Formy kwadratową można zapisać jako funkcję wielu zmiennych postaci:

$$h(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}.$$

Przyjmijmy, że $\boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{x}$ a $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$. Korzystając z wcześniej wyprowadzonego twierdzenia:

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{x}'(\boldsymbol{A}\boldsymbol{x})}{\partial \boldsymbol{x}} = \underbrace{\frac{\partial \boldsymbol{x}'}{\partial \boldsymbol{x}}}_{\frac{\partial \boldsymbol{g}'(\boldsymbol{x})}{\partial \boldsymbol{x}}} \underbrace{\boldsymbol{A}\boldsymbol{x}}_{\boldsymbol{f}(\boldsymbol{x})} + \underbrace{\frac{\partial \boldsymbol{x}'\boldsymbol{A}'}{\partial \boldsymbol{x}}}_{\frac{\partial \boldsymbol{f}'(\boldsymbol{x})}{\partial \boldsymbol{x}}} \underbrace{\boldsymbol{x}}_{\boldsymbol{g}(\boldsymbol{x})}$$

$$= \boldsymbol{A}\boldsymbol{x} + \left(\frac{\partial \boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}'}\right)' \boldsymbol{x} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{A}'\boldsymbol{x} = (\boldsymbol{A} + \boldsymbol{A}')\,\boldsymbol{x}.$$

W przypadku, kiedy $\boldsymbol{A}$ jest symetryczne $\frac{\partial \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}} = 2\boldsymbol{A}\boldsymbol{x}$. Policzymy teraz Hessian dla $h(\boldsymbol{x})$. Korzystając z wcześniej wyprowadzonych wzorów uzyskujemy:

$$\frac{\partial^2 f(\boldsymbol{x})}{\partial \boldsymbol{x} \partial \boldsymbol{x}'} = \frac{\partial \left(\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}}\right)}{\partial \boldsymbol{x}'} = \frac{\partial (\boldsymbol{A} + \boldsymbol{A}')\,\boldsymbol{x}}{\partial \boldsymbol{x}'} = \boldsymbol{A} + \boldsymbol{A}'.$$

Dla symetrycznej macierzy $\boldsymbol{A}$, Hessian ten będzie równy:

$$\frac{\partial^2 f(\boldsymbol{x})}{\partial \boldsymbol{x} \partial \boldsymbol{x}'} = 2\boldsymbol{A}.$$