Esercitazione gradienti di funzioni matriciali

March 2022

1 Esempio di derivazione

$$f(x) = \boldsymbol{\theta}^{\mathrm{T}}\mathbf{A}\boldsymbol{\theta} \tag{1}$$

$$\begin{aligned}
f(\theta) &= \begin{bmatrix}\theta_1 & \theta_2\end{bmatrix}\begin{bmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{bmatrix}\begin{bmatrix}\theta_1\\ \theta_2\end{bmatrix} && (2)\\
&= \begin{bmatrix}\theta_1 a_{11} + \theta_2 a_{21} & \theta_1 a_{12} + \theta_2 a_{22}\end{bmatrix}\begin{bmatrix}\theta_1\\ \theta_2\end{bmatrix} && (3)\\
&= \theta_1(\theta_1 a_{11} + \theta_2 a_{21}) + \theta_2(\theta_1 a_{12} + \theta_2 a_{22}) && (4)\\
&= \theta_1^2 a_{11} + \theta_1\theta_2 a_{21} + \theta_1\theta_2 a_{12} + \theta_2^2 a_{22} && (5)
\end{aligned}$$

Il risultato è uno scalare, e dipende da due variabili. Quindi il gradiente sarà composto da due derivate:

$$\begin{aligned}
\nabla_\theta f(\theta) &= \begin{bmatrix}\frac{\delta}{\delta\theta_1}\theta_1^2 a_{11} + \theta_1\theta_2 a_{21} + \theta_1\theta_2 a_{12} + \theta_2^2 a_{22}\\ \frac{\delta}{\delta\theta_2}\theta_1^2 a_{11} + \theta_1\theta_2 a_{21} + \theta_1\theta_2 a_{12} + \theta_2^2 a_{22}\end{bmatrix} = \begin{bmatrix}2\theta_1 a_{11} + \theta_2 a_{21} + \theta_2 a_{12}\\ \theta_1 a_{21} + \theta_1 a_{12} + 2\theta_2 a_{22}\end{bmatrix} && (6)\\
&= \begin{bmatrix}\theta_1(a_{11} + a_{11}) + \theta_2(a_{21} + a_{12})\\ \theta_1(a_{21} + a_{12}) + \theta_2(a_{22} + a_{22})\end{bmatrix} = \begin{bmatrix}(a_{11} + a_{11}) & (a_{21} + a_{12})\\ (a_{21} + a_{12}) & (a_{22} + a_{22})\end{bmatrix}\begin{bmatrix}\theta_1\\ \theta_2\end{bmatrix} && (7)\\
&= (\begin{bmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{bmatrix} + \begin{bmatrix}a_{11} & a_{21}\\ a_{12} & a_{22}\end{bmatrix})\begin{bmatrix}\theta_1\\ \theta_2\end{bmatrix} = (\mathbf{A} + \mathbf{A}^{\mathrm{T}})\boldsymbol{\theta} && (8)
\end{aligned}$$

Se A è simmetrica, $a_{21} = a_{12}$

$$(\begin{bmatrix}a_{11} & a_{12}\\ a_{21} & a_{12}\end{bmatrix} + \begin{bmatrix}a_{11} & a_{12}\\ a_{12} & a_{22}\end{bmatrix})\begin{bmatrix}\theta_1\\ \theta_2\end{bmatrix} = 2\begin{bmatrix}a_{11} & a_{12}\\ a_{21} & a_{12}\end{bmatrix}\begin{bmatrix}\theta_1\\ \theta_2\end{bmatrix} = 2\mathbf{A}\boldsymbol{\theta} \tag{9}$$

2 Esercizio 1

$$f(\boldsymbol{\theta}) = \mathbf{y}^{\mathrm{T}}\mathbf{y} - 2\mathbf{y}^{\mathrm{T}}\mathbf{X}\boldsymbol{\theta} + \boldsymbol{\theta}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\theta \tag{10}$$

$$\nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}}(\mathbf{y}^{\mathrm{T}}\mathbf{y}) - \nabla_{\boldsymbol{\theta}}(2\mathbf{y}^{\mathrm{T}}\mathbf{X}\boldsymbol{\theta}) + \nabla_{\boldsymbol{\theta}}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\theta) \tag{11}$$

$$\begin{align}
\nabla_{\boldsymbol{\theta}}(\mathbf{y}^{\mathrm{T}}\mathbf{y}) &= \mathbf{0} \tag{12}\\
\nabla_{\boldsymbol{\theta}}(2\mathbf{y}^{\mathrm{T}}\mathbf{X}\boldsymbol{\theta}) &= \nabla_{\boldsymbol{\theta}}(2\boldsymbol{\theta}^{\mathrm{T}}(\mathbf{y}^{\mathrm{T}}\mathbf{X})^{\mathrm{T}}) = 2(\mathbf{y}^{\mathrm{T}}\mathbf{X})^{\mathrm{T}} = 2\mathbf{X}^{\mathrm{T}}\mathbf{y} \tag{13}\\
&\text{per equazione 69 del Matrix CookBook} \tag{14}\\
\nabla_{\boldsymbol{\theta}}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\theta) &= \nabla_{\boldsymbol{\theta}}(\boldsymbol{\theta}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})\boldsymbol{\theta}) = 2\mathbf{X}^{\mathrm{T}}\mathbf{X}\boldsymbol{\theta} \tag{15}\\
&\text{per equazione 81 del Matrix CookBook} \tag{16}\\
& \tag{17}
\end{align}$$

$$\nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}) = -2\mathbf{X}^{\mathrm{T}}\mathbf{y} + 2\mathbf{X}^{\mathrm{T}}\mathbf{X}\boldsymbol{\theta} \tag{18}$$

3 Esercizio 2

$$\begin{aligned}
\nabla_{\boldsymbol{\Theta}} f(\boldsymbol{\Theta}) = {} & \nabla_{\boldsymbol{\Theta}} \|\mathbf{Y}^{\mathrm{T}} - \mathbf{X}^{\mathrm{T}}\boldsymbol{\Theta}\|_F^2 \\
& \text{per definizione di norma di Frobenius} \\
& = \nabla_{\boldsymbol{\Theta}} Tr((\mathbf{Y}^{\mathrm{T}} - \mathbf{X}^{\mathrm{T}}\boldsymbol{\Theta})(\mathbf{Y}^{\mathrm{T}} - \mathbf{X}^{\mathrm{T}}\boldsymbol{\Theta})^{\mathrm{T}}) \\
& \text{proprietà distributiva della trasposta} \\
& = \nabla_{\boldsymbol{\Theta}} Tr((\mathbf{Y}^{\mathrm{T}} - \mathbf{X}^{\mathrm{T}}\boldsymbol{\Theta})(\mathbf{Y} - \boldsymbol{\Theta}^{\mathrm{T}}\mathbf{X})) \\
& \text{sviluppo dlla moltiplicazione} \\
& = \nabla_{\boldsymbol{\Theta}} Tr(\mathbf{Y}^{\mathrm{T}}\mathbf{Y} - \mathbf{X}^{\mathrm{T}}\boldsymbol{\Theta}\mathbf{Y} - \mathbf{Y}^{\mathrm{T}}\boldsymbol{\Theta}^{\mathrm{T}}\mathbf{X} + \mathbf{X}^{\mathrm{T}}\boldsymbol{\Theta}\boldsymbol{\Theta}^{\mathrm{T}}\mathbf{X}) \\
& \text{proprietà distributiva della traccia} \\
& = \nabla_{\boldsymbol{\Theta}} Tr(\mathbf{Y}^{\mathrm{T}}\mathbf{Y}) - \nabla_{\boldsymbol{\Theta}} \mathbf{Tr}(\mathbf{X}^{\mathrm{T}}\boldsymbol{\Theta}\mathbf{Y}) - \nabla_{\boldsymbol{\Theta}} \mathbf{Tr}(\mathbf{Y}^{\mathrm{T}}\boldsymbol{\Theta}^{\mathrm{T}}\mathbf{X}) + \nabla_{\boldsymbol{\Theta}} \mathbf{Tr}(\mathbf{X}^{\mathrm{T}}\boldsymbol{\Theta}\boldsymbol{\Theta}^{\mathrm{T}}\mathbf{X})) \\
& \text{Regole di derivazioni per le tracce (sezione 2.5 del Matrix Cookbook).} \\
& \text{Nota: per risolvere l'ultima abbiamo usato la proprietà ciclica} \\
& \text{della traccia: } Tr(\mathbf{X}^{\mathrm{T}}\boldsymbol{\Theta}\boldsymbol{\Theta}^{\mathrm{T}}\mathbf{X}) = Tr(\boldsymbol{\Theta}^{\mathrm{T}}\mathbf{X}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Theta}) \\
& = -\mathbf{X}\mathbf{Y}^{\mathrm{T}} - \mathbf{X}\mathbf{Y}^{\mathrm{T}} + \mathbf{X}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Theta} + \mathbf{X}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Theta} \\
& = -2\mathbf{X}\mathbf{Y}^{\mathrm{T}} + 2\mathbf{X}\mathbf{X}^{\mathrm{T}}\Theta
\end{aligned}$$

Risolviamo per $\boldsymbol{\Theta}$ uguagliandola a 0

$$\begin{align}
& -2\mathbf{X}\mathbf{Y}^{\mathrm{T}} + 2\mathbf{X}\mathbf{X}^{\mathrm{T}}\Theta = 0 \tag{19}\\
& \mathbf{X}\mathbf{X}^{\mathrm{T}}\Theta = \mathbf{X}\mathbf{Y}^{\mathrm{T}} \tag{20}\\
& \boldsymbol{\Theta} = (\mathbf{X}\mathbf{X}^{\mathrm{T}})^{\mathbf{-1}}\mathbf{X}\mathbf{Y}^{\mathrm{T}} \tag{21}
\end{align}$$