Vogliamo dimostrare la formula per il gradiente del funzionale d'errore della regressione multipla:

$$\boldsymbol{\nabla}_{\boldsymbol{\beta}} \|\mathbf{y}-\tilde{\mathbf{x}}\boldsymbol{\beta}\|^2 = -2\tilde{\mathbf{x}}^t(\mathbf{y}-\tilde{\mathbf{x}}\boldsymbol{\beta}) \,, \tag{$*$}$$

in cui

- $\boldsymbol{\beta}$ è un vettore colonna in $\mathbb{R}^{k+1}$ ed è la variabile rispetto a cui vogliamo calcolare il gradiente;
- $\mathbf{y}$ è un vettore colonna in $\mathbb{R}^n$ (costante in $\boldsymbol{\beta}$);
- $\tilde{\mathbf{x}}$ è una matrice reale con n righe e $k+1$ colonne (costante in $\boldsymbol{\beta}$);
- $\|\cdot\|$ è la norma euclidea di $\mathbb{R}^n$.

Indichiamo con $\langle \boldsymbol{\alpha} \,,\, \boldsymbol{\beta}\rangle$ il prodotto scalare euclideo di due vettori $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^{k+1}$, cioè

$$\langle \boldsymbol{\alpha} \,,\, \boldsymbol{\beta}\rangle = \boldsymbol{\alpha}^t\boldsymbol{\beta} = \boldsymbol{\beta}^t\boldsymbol{\alpha}$$

in cui $\boldsymbol{\alpha}^t\boldsymbol{\beta}$ è l'usuale prodotto righe per colonne del vettore riga $\boldsymbol{\alpha}^t$ col vettore colonna $\boldsymbol{\beta}$ (risultato del prodotto = uno scalare). Usiamo la stessa notazione $\langle \mathbf{y} \,,\, \mathbf{z}\rangle = \mathbf{y}^t\mathbf{z}$ anche quando $\mathbf{y}, \mathbf{z} \in \mathbb{R}^n$. Osserviamo che $\langle \mathbf{y} \,,\, \tilde{\mathbf{x}}\boldsymbol{\beta}\rangle = \langle \tilde{\mathbf{x}}^t\mathbf{y} \,,\, \boldsymbol{\beta}\rangle$, in quanto

$$\langle \mathbf{y} \,,\, \tilde{\mathbf{x}}\boldsymbol{\beta}\rangle = \mathbf{y}^t\tilde{\mathbf{x}}\boldsymbol{\beta} = (\tilde{\mathbf{x}}^t\mathbf{y})^t\boldsymbol{\beta} = \left\langle \tilde{\mathbf{x}}^t\mathbf{y} \,,\, \boldsymbol{\beta}\right\rangle \,.$$

Notiamo inoltre che il quadrato della norma euclidea è per definizione

$$\|\mathbf{y}\|^2 = \langle \mathbf{y} \,,\, \mathbf{y}\rangle \,,$$

e quindi vale il teorema di Carnot

$$\begin{aligned} \|\mathbf{y}+\mathbf{z}\|^2 &= \langle \mathbf{y}+\mathbf{z} \,,\, \mathbf{y}+\mathbf{z}\rangle = \left\|\mathbf{y}^2\right\| + \langle \mathbf{y} \,,\, \mathbf{z}\rangle + \langle \mathbf{z} \,,\, \mathbf{y}\rangle + \|\mathbf{z}\|^2 \\ &= \|\mathbf{y}\|^2 + 2\,\langle \mathbf{y} \,,\, \mathbf{z}\rangle + \|\mathbf{z}\|^2 \,. \end{aligned}$$

Infine, ricordiamo che, se $f : \mathbb{R}^{k+1} \to \mathbb{R}$ è una qualunque funzione scalare, il suo *gradiente* calcolato in $\boldsymbol{\beta}$ è il vettore $\boldsymbol{\nabla} f(\boldsymbol{\beta}) \in \mathbb{R}^{k+1}$ definito dalla relazione

$$\langle \boldsymbol{\nabla} f(\boldsymbol{\beta}) \,,\, \boldsymbol{\alpha}\rangle = \left.\frac{\mathrm{d}}{\mathrm{d}t} f(\boldsymbol{\beta}+t\boldsymbol{\alpha})\right|_{t=0} \qquad \forall \boldsymbol{\alpha} \in \mathbb{R}^{k+1} \,.$$

L'equazione ($*$) che vogliamo dimostrare segue immediatamente da quest'ultima definizione, in cui semplicemente poniamo $f(\boldsymbol{\beta}) = \|\mathbf{y}-\tilde{\mathbf{x}}\boldsymbol{\beta}\|^2$ e osserviamo che

$$\begin{aligned} f(\boldsymbol{\beta}+t\boldsymbol{\alpha}) &= \|\mathbf{y}-\tilde{\mathbf{x}}(\boldsymbol{\beta}+t\boldsymbol{\alpha})\|^2 = \|(\mathbf{y}-\tilde{\mathbf{x}}\boldsymbol{\beta}) - t\tilde{\mathbf{x}}\boldsymbol{\alpha}\|^2 \\ &= \|\mathbf{y}-\tilde{\mathbf{x}}\boldsymbol{\beta}\|^2 + 2\,\langle \mathbf{y}-\tilde{\mathbf{x}}\boldsymbol{\beta} \,,\, -t\tilde{\mathbf{x}}\boldsymbol{\alpha}\rangle + \|-t\tilde{\mathbf{x}}\boldsymbol{\alpha}\|^2 \\ &= \|\mathbf{y}-\tilde{\mathbf{x}}\boldsymbol{\beta}\|^2 + t\left\langle -2\tilde{\mathbf{x}}^t(\mathbf{y}-\tilde{\mathbf{x}}\boldsymbol{\beta}) \,,\, \boldsymbol{\alpha}\right\rangle + t^2\,\|\tilde{\mathbf{x}}\boldsymbol{\alpha}\|^2 \qquad \text{(polinomio di secondo grado in } t\text{)} \,, \end{aligned}$$

da cui

$$\begin{aligned} \left\langle \boldsymbol{\nabla}_{\boldsymbol{\beta}} \|\mathbf{y}-\tilde{\mathbf{x}}\boldsymbol{\beta}\|^2 \,,\, \boldsymbol{\alpha}\right\rangle &= \left.\frac{\mathrm{d}}{\mathrm{d}t}\left[\|\mathbf{y}-\tilde{\mathbf{x}}\boldsymbol{\beta}\|^2 + t\left\langle -2\tilde{\mathbf{x}}^t(\mathbf{y}-\tilde{\mathbf{x}}\boldsymbol{\beta}) \,,\, \boldsymbol{\alpha}\right\rangle + t^2\,\|\tilde{\mathbf{x}}\boldsymbol{\alpha}\|^2\right]\right|_{t=0} \\ &= \left\langle -2\tilde{\mathbf{x}}^t(\mathbf{y}-\tilde{\mathbf{x}}\boldsymbol{\beta}) \,,\, \boldsymbol{\alpha}\right\rangle \,. \end{aligned}$$

Nell'espressione precedente, abbiamo scritto $\boldsymbol{\nabla}_{\boldsymbol{\beta}}$ al posto di $\boldsymbol{\nabla}$ per ricordarci che stiamo derivando rispetto a $\boldsymbol{\beta}$. A questo punto, avendo trovato che

$$\left\langle \boldsymbol{\nabla}_{\boldsymbol{\beta}} \|\mathbf{y}-\tilde{\mathbf{x}}\boldsymbol{\beta}\|^2 \,,\, \boldsymbol{\alpha}\right\rangle = \left\langle -2\tilde{\mathbf{x}}^t(\mathbf{y}-\tilde{\mathbf{x}}\boldsymbol{\beta}) \,,\, \boldsymbol{\alpha}\right\rangle \qquad \forall \boldsymbol{\alpha} \in \mathbb{R}^{k+1} \,,$$

ne ricaviamo

$$\boldsymbol{\nabla}_{\boldsymbol{\beta}} \|\mathbf{y}-\tilde{\mathbf{x}}\boldsymbol{\beta}\|^2 = -2\tilde{\mathbf{x}}^t(\mathbf{y}-\tilde{\mathbf{x}}\boldsymbol{\beta}) \,,$$

che è l'equazione ($*$).