$$MSE(w) = \frac{1}{m}\sum_{i=1}^{m}\left(h_w(x^{(i)}) - y^{(i)}\right)^2 \quad \text{Επειδή}$$

$$h_w(x) = w \cdot x \Rightarrow MSE(w) = \frac{1}{m}\sum_{i=1}^{m}\left(w x^{(i)} - y^{(i)}\right)^2 = C(w)$$

Από Property: $\boxed{\vec{\alpha}^T \cdot \vec{a} = \sum_i a_i^2 \quad (\alpha)}$

$$\frac{1}{m}(wx - \vec{y})^T (wx - \vec{y}) = \frac{1}{m}\sum_{i=1}^{m}\left(w x^{(i)} - y^{(i)}\right)^2 = C(w)$$

Derivatives with respect to $w$:

$$\nabla_w C(w) \overset{(\alpha)}{=} \nabla_w \left\{ \frac{1}{m}(wx - \vec{y})^T (wx - \vec{y}) \right\}$$

$$\overset{(b)}{=} \frac{1}{m}\nabla_w \left\{ w^T x^T w x - w^T x^T \vec{y} - w x \vec{y}^T + \vec{y}^T \vec{y} \right\}$$

$$\overset{(c)}{=} \frac{1}{m}\nabla_w \left\{ tr\, w^T x^T w x - tr\, w^T x^T \vec{y} - tr\, w x \vec{y}^T + \cancel{tr\, \vec{y}^T \vec{y}} \right\}$$

$$\overset{(d)}{=} \frac{1}{m}\nabla_w \left\{ tr\, w^T x^T w x - 2\, tr\, w x y^T \right\} \qquad w x y^T = w(x^T y)^T$$

$$\overset{(e)}{=} \frac{1}{m}\cancel{\nabla_w \{ tr\, w^T x^T w x}\left(2 x x^T w - 2 x^T y\right)$$

$$\frac{2}{m}\left(x^T (xw - y)\right) \longrightarrow$$

Άρης

~~Ιδιότητα~~
Property:

$$\boxed{\vec{a}^T \cdot \vec{a} = \sum_i a_i^2 \quad (a)}$$

$$\frac{1}{m}(wx-\vec{y})^T(wx-\vec{y}) = \frac{1}{m}\sum_{i=1}^{m}(wx^{(i)}-y^{(i)})^2 = C(w)$$

Derivatives with respect to $w$:

$$\nabla_w C(w) \overset{(a)}{=} \nabla_w \left\{ \frac{1}{m}(wx-\vec{y})^T(wx-\vec{y}) \right\}$$

$$\overset{(b)}{=} \frac{1}{m}\nabla_w \left\{ w^T x^T wx - w^T x^T \vec{y} - wx\vec{y}^T + \vec{y}^T\vec{y} \right\}$$

$$\overset{(c)}{=} \frac{1}{m}\nabla_w \left\{ tr\, w^T x^T wx - tr\, w^T x^T \vec{y} - \underbrace{tr\, wx\vec{y}^T} + \cancel{tr\, \vec{y}^T \vec{y}} \right\}$$

$$\overset{(d)}{=} \frac{1}{m}\nabla_w \left\{ tr\, w^T x^T wx - 2\, tr\, wxy^T \right\}$$

$wxy^T = w(x^T y)^T$

$$\overset{(e)}{=} \frac{1}{m} \cancel{\nabla_w \{ tr\, w^T x^T wx} (2xx^T w - 2x^T y)$$

$$= \frac{2}{m}\left( x^T (xw - y) \right)$$

$\Longrightarrow$

Άρης
Ταλιούρης
1115201700170

(b.) we simplified the cost function and used the linearity of the gradient operator

(c) we used the fact that a trace of a real number is the number itself and we eliminated the $\vec{y}^T\vec{y}$ term because it has no dependency on $w$

(d) we used the fact that $\operatorname{tr} a = \operatorname{tr} a^T$

(e) We used the following rules of matrix calculus.

$$\nabla_A \operatorname{tr} A^T B A = (B + B^T) A \quad \text{where} \quad A = w, \; B = xx^T$$

$$\nabla_A \operatorname{tr} B^T A = B \quad \text{where} \quad A = w, \; B = x^T y$$

'Aens Τοcξιφωνης 1115201700170