Linear Algebra

Dr François Pitié

Problem 1.

Assume $\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix}$, is a vector of size $p \times 1$,

What is the size of

1. $\mathbf{a}\mathbf{a}^\top$
2. $\mathbf{a}^\top\mathbf{a}$
3. $\mathbf{a}\mathbf{a}^\top\mathbf{a}\mathbf{a}^\top$
4. $\mathbf{a}^\top\mathbf{a}\mathbf{a}^\top\mathbf{a}$

answer:

1. $p \times p$
2. 1×1
3. $p \times p$
4. 1×1

Problem 2.

Given no assumptions about matrices $\mathbf{A}$, $\mathbf{B}$ and vectors $\mathbf{a}$ and $\mathbf{b}$, compute the gradient $\frac{\partial E(\mathbf{w})}{\partial \mathbf{w}}$ for

1. $E(\mathbf{w}) = \mathbf{w}^\top\mathbf{w}$
2. $E(\mathbf{w}) = (\mathbf{w} - \mathbf{a})^\top\mathbf{A}(\mathbf{w} - \mathbf{a})$
3. $E(\mathbf{w}) = (\mathbf{A}\mathbf{w} - \mathbf{b})^\top(\mathbf{A}\mathbf{w} - \mathbf{b})$
4. $E(\mathbf{w}) = (\mathbf{w} - \mathbf{B}\mathbf{w})^\top\mathbf{A}(\mathbf{w} - \mathbf{a})$

answer:

1. $\frac{\partial E(\mathbf{w})}{\partial \mathbf{w}} = 2\mathbf{w}$
2. $\frac{\partial E(\mathbf{w})}{\partial \mathbf{w}} = (\mathbf{A} + \mathbf{A}^\top)(\mathbf{w} - \mathbf{a})$
3. $\frac{\partial E(\mathbf{w})}{\partial \mathbf{w}} = 2\mathbf{A}^\top(\mathbf{A}\mathbf{w} - \mathbf{b})$
4. $\frac{\partial E(\mathbf{w})}{\partial \mathbf{w}} = ((\mathbf{I} - \mathbf{B})^\top\mathbf{A} + \mathbf{A}^\top(\mathbf{I} - \mathbf{B}))\mathbf{w} - (\mathbf{I} - \mathbf{B})^\top\mathbf{A}\mathbf{a}$
5. We can rewrite the loss as $E(\mathbf{w}) = \mathbf{w}^\top(\mathbf{A} - \mathbf{B}^\top\mathbf{A})\mathbf{w} - \mathbf{w}^\top(\mathbf{A} - \mathbf{B}^\top\mathbf{A})\mathbf{a}$. So the gradient is $\frac{\partial E(\mathbf{w})}{\partial \mathbf{w}} = ((\mathbf{A} - \mathbf{B}^\top\mathbf{A}) + (\mathbf{A} - \mathbf{B}^\top\mathbf{A})^\top)\mathbf{w} - (\mathbf{A} - \mathbf{B}^\top\mathbf{A})\mathbf{a}$