$$if\ we\ have\ f(X) = a^T X$$

$$f(X) = a^T X = \sum_{i=1}^{n} a_i X_i$$

$$so\ \frac{\partial f(X)}{\partial X_j} = \frac{\partial}{\partial X_j}(\sum_{i=1}^{n} a_i X_i)$$

$$so\ \frac{df(x)}{dX} = a$$

$$\frac{\partial Tr(ABC)}{\partial A} = \frac{\partial Tr(ABC)}{\partial ABC} \cdot \frac{\partial ABC}{\partial A} = 1 \cdot \frac{\partial ABC}{\partial A}$$

$$\frac{\partial ABC}{\partial A} = \frac{\partial (AB)C}{\partial A} = \frac{\partial AB}{\partial A} \cdot C$$

$$if\ A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

$$if\ B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

$$AB = \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{bmatrix}$$

$$so\ \frac{\partial (A_{11}B_{11} + A_{12}B_{21})}{\partial A_{ij}} = \begin{cases} B_{11}, & i = 1, j = 1 \\ B_{21}, & i = 1, j = 2 \\ 0, & \text{other} \end{cases}$$

$$\therefore \frac{\partial AB}{\partial A} \cdot C = B^T C$$