1 point

1. Choose the correct statements about MLP implementation:

- ☑ You can write both passes of a dense layer with NumPy and make it quick even in Python

- ☐ A backward pass of a dense layer needs a 4-d tensor derivative

- ☑ A forward pass of a dense layer can be done with matrix product

- ☐ You shouldn't prefer matrix operations when working with GPU

---

1 point

2. How many dimensions will a derivative of a 3-d tensor by a 4-d tensor have?

> 7

---

1 point

3. Let's play around with matrix derivatives!

A trace $Tr(X)$ of a matrix X is a sum of its diagonal elements.

For example: $Tr \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix} = 1 + 1 = 2$. Note that trace is a scalar!

Let's find a matrix notation for $\frac{\partial Tr(X^2)}{\partial X}$ for matrix $X = \begin{pmatrix} x_{1,1} & x_{1,2} \\ x_{2,1} & x_{2,2} \end{pmatrix}$, where X^2 is a matrix product $X \cdot X$.

Please do this element-wise and figure out a matrix notation for it:

- ○ $X^T X$
- ◉ $2X^T$
- ○ $2Tr(X^T)$
- ○ $2X$
- ○ $Tr(2X)$