Name: ______________________ Roll Number: ______________________

1. For vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ and matrices $\mathbf{M} \in \mathbb{R}^{k\times d}$ and $\mathbf{A} \in \mathbb{R}^{d\times d}$, prove the following derivatives.
   (a) (2 points) $\nabla_{\mathbf{x}}(\mathbf{y}^T\mathbf{x}) = \nabla_{\mathbf{x}}(\mathbf{x}^T\mathbf{y}) = \mathbf{y}$
   Hint: You may want to write out the complete expression for the dot-product $\mathbf{x}^T\mathbf{y}$. This expression will be a scalar value. You will need to take derivatives of this scalar value with respect to every element of $\mathbf{x}$.
   (b) (3 points) $\nabla_{\mathbf{x}}(\mathbf{M}\mathbf{x}) = \mathbf{M}^T$
   Hint: You may denote the i-th row of matrix $\mathbf{M}$ by $\mathbf{m}_i^T$. Then use part (a) to write the derivative of the expression $\mathbf{m}_i^T\mathbf{x}$.
   (c) (3 points) $\nabla_{\mathbf{x}}(\mathbf{x}^T\mathbf{A}\mathbf{x}) = (\mathbf{A} + \mathbf{A}^T)\mathbf{x}$
   Hint: You may use the product rule of differentiation. When applied to vectors, the rule states that
   $$\nabla_{\mathbf{x}}\left(\mathbf{u}^T\mathbf{v}\right) = \nabla_{\mathbf{x}}(\mathbf{u})\,\mathbf{v} + \nabla_{\mathbf{x}}(\mathbf{v})\,\mathbf{u}$$
   where both $\mathbf{u}$ and $\mathbf{v}$ are functions of $\mathbf{x}$. You may also use the derivative from part (b).
   (d) (2 points) $\nabla_{\mathbf{x}}(\mathbf{x}^T\mathbf{A}\mathbf{x}) = 2\mathbf{A}\mathbf{x}$ when $\mathbf{A}$ is symmetric

(a) First note that

$$\mathbf{y}^T\mathbf{x} = \mathbf{x}^T\mathbf{y} = x_1y_1 + x_2y_2 + \cdots + x_dy_d \tag{1}$$

which is a scalar value.

$$\nabla_{\mathbf{x}}\left(\mathbf{x}^T\mathbf{y}\right) = \nabla_{\mathbf{x}}(x_1y_1 + x_2y_2 + \cdots + x_dy_d) \tag{2}$$

$$= \begin{bmatrix} \frac{d}{dx_1}(x_1y_1 + x_2y_2 + \cdots + x_dy_d) \\ \frac{d}{dx_2}(x_1y_1 + x_2y_2 + \cdots + x_dy_d) \\ \vdots \\ \frac{d}{dx_d}(x_1y_1 + x_2y_2 + \cdots + x_dy_d) \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_d \end{bmatrix} = \mathbf{y} \tag{3}$$

(b) Let $\mathbf{m}_i^T$ denote the i-th row of matrix $\mathbf{M}$. Then we can write

$$\nabla_{\mathbf{x}}(\mathbf{M}\mathbf{x}) = \nabla_{\mathbf{x}} \begin{bmatrix} \mathbf{m}_1^T\mathbf{x} \\ \mathbf{m}_2^T\mathbf{x} \\ \vdots \\ \mathbf{m}_k^T\mathbf{x} \end{bmatrix} \tag{4}$$

$$= \begin{bmatrix} \nabla_{\mathbf{x}}(\mathbf{m}_1^T\mathbf{x}) & \nabla_{\mathbf{x}}(\mathbf{m}_2^T\mathbf{x}) & \ldots & \nabla_{\mathbf{x}}(\mathbf{m}_k^T\mathbf{x}) \end{bmatrix} \tag{5}$$

$$= \begin{bmatrix} \mathbf{m}_1 & \mathbf{m}_2 & \ldots & \mathbf{m}_k \end{bmatrix} = \mathbf{M}^T \tag{6}$$

(c) We will use the product rule of differentiation. When applied to vectors, the rule states that

$$\nabla_{\mathbf{x}}\left(\mathbf{u}^T\mathbf{v}\right) = \nabla_{\mathbf{x}}(\mathbf{u})\,\mathbf{v} + \nabla_{\mathbf{x}}(\mathbf{v})\,\mathbf{u} \tag{7}$$

where both $\mathbf{u}$ and $\mathbf{v}$ are functions of $\mathbf{x}$. For our problem, we will take $\mathbf{u} = \mathbf{x}$ and $\mathbf{v} = \mathbf{A}\mathbf{x}$. Then we can write

$$\nabla_{\mathbf{x}}(\mathbf{x}^T\mathbf{A}\mathbf{x}) = (\nabla_{\mathbf{x}}\mathbf{x})\,\mathbf{A}\mathbf{x} + (\nabla_{\mathbf{x}}\mathbf{A}\mathbf{x})\,\mathbf{x} \tag{8}$$

$$= (\nabla_{\mathbf{x}}\mathbf{I}\mathbf{x})\,\mathbf{A}\mathbf{x} + (\nabla_{\mathbf{x}}\mathbf{A}\mathbf{x})\,\mathbf{x} \tag{9}$$

$$= \mathbf{I}^T\mathbf{A}\mathbf{x} + \mathbf{A}^T\mathbf{x} \tag{10}$$

$$= \mathbf{A}\mathbf{x} + \mathbf{A}^T\mathbf{x} \tag{11}$$

$$= (\mathbf{A} + \mathbf{A}^T)\mathbf{x} \tag{12}$$

(d) Continuing from part (c), when $\mathbf{A}$ is symmetric, $(\mathbf{A} + \mathbf{A}^T)\mathbf{x} = 2\mathbf{A}\mathbf{x}$.