1. (*20 points*) Given $x, w \in R^{n\times 1}$, $P \in R^{n\times n}$ find:

a) (*10 points*) $\frac{\partial}{\partial x}\left(x^T x\right)$,

b) (*10 points*) $\frac{\partial}{\partial w}\left(x^T w x^T w\right)$.

Show the **details** of the derivations.

a) See page 594 in the Textbook.

b) $\frac{\partial}{\partial w}\left(x^T w x^T w\right) = \frac{\partial}{\partial w}\left(x^T w\right)^2 = 2x^T w \frac{\partial}{\partial w}\left(x^T w\right) = 2x^T w x$

2. (*20 points*) Given $x \in R^{n\times 1}$ the input vector and $w \in R^{n\times 1}$ the weight vector of the simple linear combiner, and $e \in R$ its linear error, derive the learning rule for the LMS algorithm. This learning rule uses the steepest descent gradient method. Use the mean square error function $J\left(w\right) = \frac{1}{2}e^2(k)$ and the update rule $w(k+1) = w(k) - \mu \nabla_w J$.

a) See Textbook page 39, equations (2.28) and (2.29).

3. (*30 points*) Assume you have a three-layer feedforward Multilayer Perceptron neural network (MLP NN).

a) (*15 points*) Sketch the neural network. Indicate all the relevant matrices, vectors, and functions. Show the size of each matrix and vector in your sketch. Recall that, using our definition of layers, a three-layer network has two hidden layers and an output layer. Your sketch should point out these layers. Use the same nonlinear activation function $f\left(\bullet\right)$ for all layers.

b) (*15 points*) In order to train the weights of the MPL NN we minimized the function $E_q = \frac{1}{2}\left(d_q - x_{out}^{(3)}\right)^T\left(d_q - x_{out}^{(3)}\right)$, where d_q is the vector of desired output values for the network for the q^{th} input pattern, and $x_{out}^{(3)} = y_q$ is the vector of actual network outputs for the q^{th} input pattern. Using the steepest descent gradient approach we update the weights of the output layer (layer 3) using the following equation: $w_{ji}^{(3)}(k+1) = w_{ji}^{(3)}(k) - \mu^{(3)}\frac{\partial E_q}{\partial w_{ji}^{(3)}}$. Find $\frac{\partial E_q}{\partial w_{ji}^{(3)}} = \frac{\partial E_q}{\partial v_j^{(3)}}\frac{\partial v_j^{(3)}}{\partial w_{ji}^{(3)}}$ and express the weight update equation in terms of the following:

- an element of $x_{out}^{(2)}$, the output vector of layer 2;

- an element of d_q;
- an element of $x_{out}^{(3)}$;
- $g(v)$ where $g(\bullet)$ is the first derivative of $f(\bullet)$, and v is the activity level for the appropriate neuron.

a) See Figure 3.4 on page 107 of the textbook.
b) See equations (3.40)-(3.48) in the textbook.

4. (*15 points*) Assume that the matrix $W \in R^{n \times m}$ has columns that make up a set of *m* mutually orthogonal vectors. What are the elements of the matrix $H = W^T W$? Show the **details** of the derivations.

For the orthogonal columns the following applies: $w_i^T w_j = \begin{cases} w_i^T w_i \ ; i = j \\ 0 \ ; \quad i \neq j \end{cases}$.

Therefore:

$$W_{m \times n}^T W_{n \times m} = [w_1 \cdots w_m]_{m \times n}^T [w_1 \cdots w_m]_{n \times m} = \begin{bmatrix} w_1^T w_1 & w_1^T w_2 & \cdots & w_1^T w_m \\ w_2^T w_1 & w_2^T w_2 & \cdots & w_2^T w_m \\ \vdots & \vdots & \cdots & \vdots \\ w_m^T w_1 & w_m^T w_2 & \cdots & w_m^T w_m \end{bmatrix}_{m \times m} =$$

$$= \begin{bmatrix} w_1^T w_1 & & 0 \\ & \ddots & \\ 0 & & w_m^T w_m \end{bmatrix} = diag(w_1^T w_1, \ldots, w_m^T w_m) = H$$

5. a) *(5 points)* Sketch a radial-basis function neural network with *m* inputs, *N* neurons in the hidden layer, and a single output.
   b) *(5 points)* What do the individual neurons do?
   c) *(5 points)* Write down the equation that relates the value of the scalar output to the value of the input vector.

a) See Figure 3.10 on page 141 of the textbook.
b) The individual neurons give a measure of distance between the input and their respective centers.
c) See equation (3.146) on page 141 of the textbook. Note that this equation assumes there are multiple (*m*) outputs, hence the subscript *i*.