Problem 1 (40 points)
Consider the neural network shown below. All the units are numbered 1 through 6, where units 1 and 2 are input units relaying the two inputs, $x_1$ and $x_2$, to units 3 and 4. The output function of each unit is a sigmoid function;

$$y_i = g_i(z_i) = \frac{1}{1 + e^{-z_i}},$$

where variable $z_i$ is the weighted sum of all the inputs connected to that unit:

$$z_i = \sum_j w_{ij} x_j.$$

There are three hidden units, units 3, 4 and 5, connected to the output unit, unit 6. The output of the network, $y_6$, is connected to a known, but nonlinear process:

$$\hat{y} = y_6^3$$

A distal teacher provides a training signal $y$, which is compared to the above estimate $\hat{y}$. The weights of the network are corrected based on the error back propagation algorithm with learning rate $\eta$ to reduce the squared error given by:

$$E = \frac{1}{2} (\hat{y} - y)^2.$$

Answer the following questions.

a). Using the chain rule, compute the incremental weight change to be made to $w_{54}$ when inputs $x_1$ and $x_2$ and the corresponding distal teacher signal $y$ are presented.

b). Compute the incremental weight change $\Delta w_{41}$ for the same input-output training data as part a).

c). In computing $\Delta w_{41}$, discuss how the result of computing $\Delta w_{64}$ as well as that of $\Delta w_{54}$ can be used for streamlining the computation.

d). For a particular pair of input data, $z_6$ became very large, $z_6 \gg 1$. Is the weight change $\Delta w_{34}$ large for this input? For another input, $z_5$ became very large $z_5 \gg 1$, and $z_6$ and $z_3$ became 0. Are weight change $\Delta w_{34}$ and $\Delta w_{53}$ likely to be large when $E$ is large? Explain why.
Problem 2 (60 points)
Consider the following true system and model structure with parameter vector \( \theta \),

\[
S : \quad y(t) + 0.3y(t-1) + 0.1y(t-2) = u(t-1) + e_0(t)
\]
\[
M(\theta) : \quad y(t) + a_1y(t-1) + a_2y(t-2) = u(t-1) + e(t) + ce(t-1)
\]
\[
\theta = (a_1, a_2, c)^T
\]

where input sequence \( \{u(t)\} \) is white noise with variance \( \mu \) and \( \{e_0(t)\} \) is white noise with variance \( \lambda \). The input \( \{u(t)\} \) is uncorrelated with noise \( \{e_0(t)\} \) and \( \{e(t)\} \). Answer the following questions.

1). Obtain the one-step-ahead predictor of \( y(t) \).

2). Compute covariances
\[
R_{yx}(0) = E[y(t)e_0(t)], \quad \text{and} \quad R_{yx}(1) = E[y(t)e_0(t-1)].
\]

3). Compute covariances
\[
R_y(0) = E[y^2(t)], \quad R_y(1) = E[y(t)y(t-1)]
\]

4). Obtain the asymptotic variance of parameter estimate: \( \hat{\theta}_N = [\hat{a}_1, \hat{a}_2, \hat{c}]^T \), when a quadratic prediction-error criterion is used.
5). After identifying the parameters $\hat{a}_1$ and $\hat{a}_2$, the true system has changed to:

$$S: \quad y(t) + 0.3y(t-1) + 0.1y(t-2) = u(t-1) + e(t) + 0.2e_0(t-1).$$

Now consider the model structure:

$$M(\theta): \quad y(t) + 0.3y(t-1) + 0.1y(t-2) = u(t-1) + e(t) + ce(t-1)$$

with only one unknown parameter $\theta = (c)$. Obtain the asymptotic error covariance using the frequency-domain expression of Cov $\hat{\theta}_N$. [Hint, obtain the function $A(\omega)$ involved in the following expression:

$$Cov\hat{\theta}_N \sim \frac{1}{N} \left[ \frac{1}{2\pi} \int_{-\pi}^{\pi} A(\omega) d\omega \right]^{-1}$$

using the parameter values of the true system.]

**Problem 3** (for extra points)

What are the two most important or most inspiring things that you have learned in 2.160 that you think you should not forget even ten years after your graduation from MIT?