Learning representations by back-propagating errors

1. What is the main problem that the paper proposes to solve?
2. Describe the artificial neural network in terms of Equation 1) and 2).
3. Equation 3), what is the total error and why do we need it?
4. Explain Equation 4-6. Why did the authors derive them and how are they related to the chain rule
5. Explain Equation 8 and 9
6. What is the main message in Figure 4
7. Which part is the most impressive in the paper?