Internal Assessment 1: Deep and Reinforcement Learning

Modules 1 and 2

Srikanth Pai, Asst. Professor, MSE

Closed book. 60 minutes. All four questions compulsory. Total 20 marks.
State any result you use before applying it. Partial credit is given for correct reasoning with minor arithmetic slips.

1. **[5 marks]** A hidden layer produces $\mathbf{h} \in \mathbb{R}^d$. During training each coordinate is kept with probability p and set to zero otherwise, independently: $\tilde{h}_j = m_j h_j$ with $m_j \sim \text{Bernoulli}(p)$. The next layer forms the pre-activation $a = \mathbf{w}^\top \tilde{\mathbf{h}}$.

   Compute $\mathbb{E}[\tilde{h}_j]$ and $\mathbb{E}[a]$, and show that $\mathbb{E}[a] \neq \mathbf{w}^\top \mathbf{h}$ unless $p = 1$, so a network trained with dropout but evaluated without it at test time is systematically miscalibrated. State the modification (inverted dropout) that rescales the kept activations during training, and prove that it restores $\mathbb{E}[\tilde{h}_j] = h_j$ and so needs no adjustment at test time.

2. **[5 marks]** A 30-megapixel RGB image of size $6000 \times 5000 \times 3$ is passed through a 4×4 max-pooling layer with stride 2. Compute the size of the output, and state the main benefits of pooling.

3. **[5 marks]** A recurrent network has hidden state $\mathbf{h}_t = \tanh(W\mathbf{h}_{t-1} + U\mathbf{x}_t + \mathbf{b})$ and is trained on a sequence of length T.

   Explain what backpropagation through time is and how it differs from ordinary backpropagation. Explain the vanishing gradient problem in recurrent networks, and describe what is done to address it.

4. **[5 marks]** A fully connected network applied directly to the raw pixels of an image scales badly and ignores the spatial structure of the image. Explain how a convolutional network overcomes this. Your answer should address how it learns local features, how stacking layers builds a hierarchy of increasingly complex visual features, and why its parameter count avoids the explosion that a fully connected layer on the same input would suffer.