AIS5103 homework 2 (due 18 September 2026)

1. Read Sections 6.1 and 6.2 (pp. 171-186) of Chapter 6 of Bishop's book "Deep Learning". Briefly summarize these sections. Your summary should be no longer than half an A4 page.

2. Read the relevant PyTorch tutorials and implement a handwritten digit recognition model for the MNIST dataset. The dataset can be loaded from torchvision using torchvision.datasets.MNIST(...). Follow the tutorials to construct appropriate DataLoaders for the training and test datasets. Using the torch.nn framework, define the model as a subclass of Module with the following three-layer network architecture:

   Input: ($28 \times 28=784$) input features, represented as floating-point values.

   Hidden layer: A fully connected layer with 1024 neurons, followed by a ReLU activation function.

   Output layer: A fully connected layer with 10 output features (logits), corresponding to the digits 0–9.

   Use torch.nn.CrossEntropyLoss as the loss function, as motivated by negative log-likelihood and maximum likelihood estimation (MLE). Train the model using stochastic gradient descent (SGD). Plot the training and test losses as functions of epoch. Report the final training and test accuracies, defined as the fractions of correctly classified samples. Vary the learning rate and batch size, and discuss their effects on the training process and model performance.

   Briefly summarize what you have learned from this exercise.

3. Bishop, Exercise 6.10, p. 206. Consider a regression problem involving multiple target variables in which it is assumed that the distribution of the targets, conditioned on the input vector $\mathbf{x}$, is a Gaussian of the form

$$p(\mathbf{t}|\mathbf{x},\mathbf{w}) = N(\mathbf{t}|\mathbf{y}(\mathbf{x},\mathbf{w}),\boldsymbol{\Sigma})$$

   where $\mathbf{y}(\mathbf{x},\mathbf{w})$ is the output vector of a neural network with input vector $\mathbf{x}$ and weight vector $\mathbf{w}$, and $\boldsymbol{\Sigma}$ is the covariance matrix of the assumed Gaussian noise in the targets. (a) Given a set of independent observations of $\mathbf{x}$ and $\mathbf{t}$, write down the error function that must be minimized to obtain the maximum likelihood solution for $\mathbf{w}$, if we assume that $\boldsymbol{\Sigma}$ is fixed and known. (b) Now assume that $\boldsymbol{\Sigma}$ is also to be determined from the data, and write down an expression for the maximum likelihood solution for $\boldsymbol{\Sigma}$. Note that the optimizations of $\boldsymbol{w}$ and $\boldsymbol{\Sigma}$ are now coupled, in contrast to the case of independent target variables discussed in Section 6.4.1.

[Note: in Bishop, a roman bold letter such as $\mathbf{t}$ or $\mathbf{x}$ denotes vectors. Read Section 3.2 on the multivariate Gaussian distribution before doing this problem.] (c) If the covariance $\mathbf{\Sigma}$ is also learned from the data, how is the information reflected or presented in the neural network?

4. This question concerns backpropagation in neural networks. Consider a two-layer feedforward neural network represented by the equations

$$\mathbf{a} = W^{(1)}\mathbf{x}, \quad \mathbf{z} = h(\mathbf{a}), \quad \mathbf{y} = W^{(2)}\mathbf{z}.$$

Here, $\mathbf{x}$, $\mathbf{a}$, $\mathbf{z}$, and $\mathbf{y}$ are vectors. The function $h(\cdot)$ is an elementwise nonlinear activation function with derivative h', and $W^{(1)}$ and $W^{(2)}$ are learnable weight matrices. The loss function is given by the squared error $E = \frac{1}{2}\|\mathbf{y} - \mathbf{t}\|_2^2$. $\mathbf{t}$ is the target vector. A labeled data point is the pair $(\mathbf{x}, \mathbf{t})$.

   a. Express the differential dE in terms of the differentials of the weight matrices, $dW^{(1)}$ and $dW^{(2)}$.
   b. Show that the gradients with respect to the weight matrices can be written as outer products of appropriate vectors, $\frac{\partial E}{\partial W^{(1)}} = \boldsymbol{\delta}\mathbf{x}^T$ and $\frac{\partial E}{\partial W^{(2)}} = \mathbf{e}\mathbf{z}^T$. What are the quantities $\boldsymbol{\delta}$ and $\mathbf{e}$?

\- The end -