Q1 (a) Assume that there are two paths (Path A and Path B) for transmitting electricity from a hydro power plant to a data center. Assume further that the power loss for transmitting x megawatt (MW) through Path A is ax^2 and that the power loss for transmitting y MW through Path B is by^2, where $a > 0$ and $b > 0$.

(i) Express the total power loss L in terms of x, y, a, and b.

(3 marks)

(ii) Determine the optimal power transfer (x^* and y^*) through these two paths with minimum power loss if it is necessary to transfer P MW to the data center. *Hint*: You may frame this problem as a constrained minimisation problem.

(20 marks)

(iii) Determine the percentage increase in the optimal power loss if P is increased by 5%.

(17 marks)

(b) The expectation-maximization (EM) algorithm is a common algorithm for training probabilistic generative models. The EM algorithm computes the posterior expectation of some latent (hidden) variables during the E-step and maximizes an auxiliary function during the M-step. It can be shown that given a training data set, increasing the value of the auxiliary function will also increase the likelihood of the training data.

(i) If the generative model is a Gaussian mixture model, what will be the latent variables?

(5 marks)

(ii) Discuss the advantage of maximizing the auxiliary function over the direct maximization of the data likelihood with respect to the model parameters.

(5 marks)

Q2 (a) In principal component analysis (PCA), given a training set $\mathcal{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ comprising vectors in D-dimensional space, the M-dimensional PCA-projected vectors $\{\mathbf{h}_i\}_{i=1}^N$ can be considered as latent variables that can approximately reconstruct the training data:

$$\mathbf{x}_i \approx \mathbf{\Phi}\mathbf{h}_i + \boldsymbol{\mu}, \qquad i = 1, \ldots, N,$$

where $\mathbf{\Phi}$ is a $D \times M$ projection matrix and $\boldsymbol{\mu}$ is the global mean. To determine $\mathbf{\Phi}$ from training data, we minimize the following cost function:

$$\hat{\mathbf{\Phi}}, \{\hat{\mathbf{h}}_i\}_{i=1}^N = \underset{\mathbf{\Phi}, \{\mathbf{h}_i\}_{i=1}^N}{\arg\min} \left\{ \sum_{i=1}^N [\mathbf{x}_i - \boldsymbol{\mu} - \mathbf{\Phi}\mathbf{h}_i]^\mathsf{T} [\mathbf{x}_i - \boldsymbol{\mu} - \mathbf{\Phi}\mathbf{h}_i] \right\}.$$

(i) Show that $\hat{\mathbf{\Phi}}$ comprises the eigenvectors of

$$\sum_{i=1}^N (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^\mathsf{T}$$

Hints:

$$\frac{\partial}{\partial \mathbf{X}} \mathrm{tr}\{\mathbf{X}\mathbf{B}\mathbf{X}^\mathsf{T}\} = \mathbf{X}\mathbf{B}^\mathsf{T} + \mathbf{X}\mathbf{B} \quad \text{and} \quad \frac{\partial \mathbf{a}^\mathsf{T}\mathbf{X}^\mathsf{T}\mathbf{b}}{\partial \mathbf{X}} = \mathbf{b}\mathbf{a}^\mathsf{T}.$$

(15 marks)

(ii) If the dimension of $\mathbf{x}_i$ is 100,000, explain why computing $\hat{\mathbf{\Phi}}$ using the solution in Q2(a)(i) is very expensive. (5 marks)

(b) Fig. Q2 shows a neural network with one hidden layer.

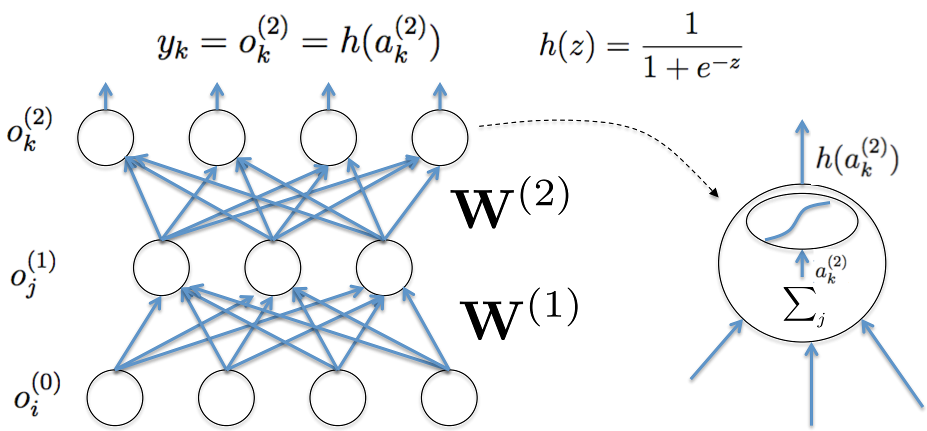

Fig. Q2

Given a training vector $\mathbf{x} \in \mathbb{R}^D$, the instantaneous squared error between the actual outputs y_k's and the target outputs t_k's of the network is given by

$$E = \frac{1}{2}\sum_{k=1}^{K}(y_k - t_k)^2 = \frac{1}{2}\sum_{k=1}^{K}\left(o_k^{(2)} - t_k\right)^2, \qquad \text{(Q2-a)}$$

where

$$o_k^{(2)} = h(a_k^{(2)}) \qquad \text{(Q2-b)}$$

is the output of the k-th output node of the network. In Eq. Q2-b,

$$h(a_k^{(2)}) = \frac{1}{1+e^{-a_k^{(2)}}} \quad \text{and} \quad a_k^{(2)} = \sum_j o_j^{(1)} w_{kj}^{(2)},$$

where $o_j^{(1)}$ is the j-th hidden node's output.

(i) Based on Fig. Q2, determine the value of K in Eq. Q2-a.

(3 marks)

(ii) Show that the gradient of E with respect to the weight connecting the j-th hidden node and the k-th output node is given by

$$\frac{\partial E}{\partial w_{kj}^{(2)}} = \left(o_k^{(2)} - t_k\right) o_k^{(2)} \left(1 - o_k^{(2)}\right) o_j^{(1)}.$$

(7 marks)

(iii) Write the equation for updating the weight $w_{kj}^{(2)}$ based on stochastic gradient descent.

(5 marks)

(c) The state-space model of a Kalman filter with state variable $\mathbf{x}_t$ at time step t is given by

$$\mathbf{x}_t = \mathbf{F}_t\mathbf{x}_{t-1} + \mathbf{B}_t\mathbf{u}_t + \mathbf{w}_t, \qquad \text{(Q2-c)}$$

where $\mathbf{F}_t$ is a state transition matrix, $\mathbf{B}_t$ is a control input matrix, $\mathbf{u}_t$ is any control inputs, and $\mathbf{w}_t$ is a noise vector with covariance matrix $\mathbf{Q}_t$. The system is measured at each time step and the measurement vector $\mathbf{z}_t$ is assumed to follow a linear model:

$$\mathbf{z}_t = \mathbf{H}_t\mathbf{x}_t + \mathbf{v}_t, \qquad \text{(Q2-d)}$$

where $\mathbf{H}_t$ is a transformation matrix that maps the state vector to the measurement domain and $\mathbf{v}_t$ is the measurement noise with covariance matrix $\mathbf{R}_t$.

(i) Assume that you have a DC power supply with the target output voltage set to v Volt. However, the power supply is not perfect so that its output voltage has a variance of σ_v^2. Assume also that you have a voltmeter that gives you a measured voltage every 5 seconds. The voltmeter is imperfect so that the variance of the measurement is σ_m^2. Discuss how you will apply the Kalman filter to estimate the output voltage of the power supply. *Hint*:

You may answer this question by substituting the variables in Eq. Q2-c and Eq. Q2-d with the physical variables in this question.

(10 marks)

(ii) Will the estimated voltage given by the Kalman filter become closer to the true voltage outputted by the power supply over time? Briefly explain your answer.

(5 marks)

– END –