Fifth practice exercises in Machine learning 1 – 2025 – Paper 1

1 Principal component analysis (October)

Suppose we have a data set $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ of D-dimensional vectors, which have a zero mean for each dimension. Assume we perform a complete eigenvalue decomposition of the empirical covariance matrix $\mathbf{S} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$. You are interested in only a single projection of your data such that the variance of this projection is maximized. Let $\mathbf{u}_i$ be the direction vector of a particular projection. Assume that $\mathbf{u}_i^T\mathbf{u}_i = 1$.

(a) What is the projection z_{ni} of a given point $\mathbf{x}_n$ under the particular vector $\mathbf{u}_i$?

(b) What is the empirical mean of the projection z_i across all points $\mathbf{x}_n$?

(c) What is the empirical variance of the projection z_i? Provide your answer in terms of the empirical covariance matrix $\mathbf{S}$

(d) Replace $\mathbf{S}$ with its eigenvalue decomposition and simplify the aforementioned expression. What is the variance now?

with $\mathbf{e}_i$ to be a vector with zeros except the position with index i.

(e) Suppose that you are interested in reducing the dimensionality from D to K, such that 99% of the variance is maintained. How can you select an appropriate K?

Fifth practice exercises in Machine learning 1 – 2025 – Paper 1

2 Introduction to backpropagation (October)

Consider a two-layer neural network, defined as follows

$$\mathbf{z}^{(1)} = \mathbf{W}^{(1)}\mathbf{x} + \mathbf{b}^{(1)}, \quad \mathbf{a}^{(1)} = \tanh(\mathbf{z}^{(1)}),$$

$$\mathbf{z}^{(2)} = \mathbf{W}^{(2)}\mathbf{a}^{(1)} + \mathbf{b}^{(2)}, \quad \hat{\mathbf{y}} = \text{softmax}(\mathbf{z}^{(2)}),$$

where

- $\mathbf{x} \in \mathbb{R}^D$ is a single input example.
- $\hat{\mathbf{y}} \in \mathbb{R}^K$ is the predicted output vector.
- The first hidden layer (1) has H neurons.

(a) What are the shapes of $\mathbf{W}^{(1)}, \mathbf{b}^{(1)}, \mathbf{a}^{(1)}, \mathbf{W}^{(2)}, \mathbf{b}^{(2)}$?

(b) We feed the neural network with a sample $\mathbf{x}$, obtaining the prediction $\hat{\mathbf{y}}$. We assume we have the ground truth label $\mathbf{y}$ for that sample $\mathbf{x}$. Write down the formula for the cross-entropy loss $L(\mathbf{y}, \hat{\mathbf{y}})$ for that sample.

(c) Compute $\frac{\partial L}{\partial \hat{y}_i}$.

(d) Compute $\frac{\partial L}{\partial {z_i}^{(2)}}$.

(e) Compute $\frac{\partial L}{\partial W_{ij}^{(2)}}$.