Mathematics for Machine Learning (2025 Fall)
Quiz

Chuang-Chieh Lin

Department of Computer Science and Engineering,
National Taiwan Ocean University

December 22, 2025

1. (25%) (True or False) Given the Gaussian Mixture Model shown as below, the *smaller* number K is, the more expressive the model will be.

$$p(\mathbf{x} \mid \boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$
$$0 \leq \pi_k \leq 1, \sum_{k=1}^{K} \pi_k = 1,$$

2. (25%) Given a dataset $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$, why (or when) can we have the following deduction?

$$p(\mathcal{X} \mid \boldsymbol{\theta}) = \prod_{i=1}^{N} p(\mathbf{x}_i \mid \boldsymbol{\theta}).$$

3. (25%) (True or False) The following SVM is of Hard-Margin.

$$\min_{\mathbf{w}, b, \boldsymbol{\xi}} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N} \xi_i$$
$$\text{subject to } \; y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \geq -\xi_i,$$
$$\xi_i \geq 0$$

4. (25%) (True or False) When we consider maximizing the Lagrangian of SVM, we obtain

$$\mathbf{w} = \sum_{i=1}^{N} \alpha_i y_i \mathbf{x}_i,$$

which is a linear combination of the data points $\mathbf{x}_i$'s. Then, $\mathbf{x}_i$'s with $\alpha_i > 0$ is called the *support vectors*.