SVM and Clustering Exam (100 Points)

UTK COSC 522: Machine Learning (Fall 2025)

OUT: Monday, Nov 17th, 2025

Due Date: Dec 5th, 2025, 11:59:59PM

Instructions: Please answer the following questions clearly and show your work for any computations.

Part 1: Support Vector Machines (SVM)

1. **(15 points) Hyperplane Classification and Distance**

   Given a 2D hyperplane defined by the weights $\mathbf{w} = (1, -2)^T$ and bias $b = 0$. The decision function is $f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b$. Points with $f(\mathbf{x}) \geq 0$ are classified as $y = +1$, and points with $f(\mathbf{x}) < 0$ are classified as $y = -1$.

   a) (5 pts) For the point $\mathbf{x}_1 = (1, 1)$ with true label $y_1 = +1$, is this point correctly classified by the hyperplane? Show your calculation.

   b) (5 pts) For the point $\mathbf{x}_2 = (1, 0)$ with true label $y_2 = +1$, is this point correctly classified by the hyperplane? Show your calculation.

c) (5 pts) Calculate the signed geometric distance of the point $\mathbf{x}_3 = (-1, 2)$ from the hyperplane (see the lecture slides).

2. **(20 points) Soft-Margin SVM and Hyperparameter C**

a) (10 pts) Explain the role of slack variables (ξ_i) and the hyperparameter C in a soft-margin SVM. What is the optimization objective for the soft-margin SVM (you can describe it in words or as a formula)?

b) (10 pts) Describe the difference between a classifier trained with a **very small** C versus one trained with a **very large** C. How does this choice affect:

- The width of the margin?
- The number of support vectors?

- The model's "willingness" to misclassify training points?
- The model's position in the bias-variance trade-off?

3. **(15 points) The Kernel Trick**

a) (5 pts) What problem in classification do non-linear kernels (the "kernel trick") solve?

b) (10 pts) Imagine you have a 1D dataset with points:

- Class +1: {`-4, 4`}
- Class -1: {`-1, 1`}

This data is not linearly separable in 1D. Describe how a simple polynomial kernel, $\phi(x) = x^2$, would transform this data to make it linearly separable in a new dimension. Show the new values for the transformed points and state the simple linear rule (e.g., "classify as +1 if...") that could separate them in the transformed space.

4. **(15 points) The Dual Problem and Support Vectors**

   In the dual formulation of the soft-margin SVM, each data point $\mathbf{x}_i$ has a corresponding dual variable α_i. Based on the Karush-Kuhn-Tucker (KKT) conditions, the value of α_i is constrained by $0 \leq \alpha_i \leq C$.

   For each of the following three cases, state the value or range of the dual variable α_i and explain *what* that value tells us about the point's position relative to the hyperplane and margin:

   - **Case 1:** The point $\mathbf{x}_i$ is correctly classified and is **outside** the margin.
   - **Case 2:** The point $\mathbf{x}_i$ is **on** the margin.

- **Case 3:** The point $\mathbf{x}_i$ is **within** the margin or on the wrong side of the hyperplane (i.e., it is a margin violation).

Part 2: Clustering (K-Means and GMM)

5. **(15 points) K-Means Algorithm Step-by-Step**

   You are given $K = 2$ and a 2D dataset with four points:

   - $P_1 = (1, 1)$
   - $P_2 = (1, 3)$
   - $P_3 = (4, 2)$
   - $P_4 = (4, 4)$

   Your initial centroids are $\mathbf{c}_1 = (1, 2)$ and $\mathbf{c}_2 = (4, 3)$.

   Perform **one full iteration** of the K-means algorithm. You must show your work for both steps:

   a) (8 pts) **Assignment Step:** Assign each of the four points (P_1, P_2, P_3, P_4) to its nearest centroid ($\mathbf{c}_1$ or $\mathbf{c}_2$). Use Euclidean distance.

   b) (7 pts) **Update Step:** Calculate the new centroids (let's call them $\mathbf{c}'_1$ and $\mathbf{c}'_2$) by taking the mean of all points assigned to each cluster in step (a).

6. **(20 points) K-Means vs. Gaussian Mixture Models (GMM)**

a) (10 pts) Compare K-Means and GMM in terms of their cluster assignments. Specifically, explain the difference between a "hard" assignment and a "soft" assignment.

b) (10 pts) Describe or draw a simple 2D dataset (e.g., with two clusters) where you would expect GMM to perform significantly better than K-Means. Explain *why* K-Means would struggle and what assumption GMM makes that allows it to model this data more effectively. (Hint: think about cluster shapes or density).