INT305 – Coursework 1

Assessment Number	1
Contribution to Overall Marks	15%
Submission Deadline	16/10/2025

Coursework Submission Guideline

Write the answers under each sub-task. That is to say, you need to **copy the question first** in word/latex, then **write your answers below the copied question**. After you answer all the questions, please transfer your word/latex into pdf and **submit the final pdf**.

Submit file format: Single column; Font size: #12; Page number: no more than 15; **No need** to prepare a coversheet, you can directly copy the questions and write the answers.

Submit file name: INT305-CW1-Name-studentID.pdf (*e.g.*, INT305-CW1-SanZhang-2025305.pdf)

Assessment Objective

This assessment aims at evaluating students at Having a solid understanding of the theoretical issues related to problems that machine learning algorithms try to address. And check if the students are able to ascertain the properties of existing ML algorithms and new ones.

Tasks

For a training example (x_i, y_i) with K classes:
$s = Wx_i = score\ vector\ s_j$ (s_j= predicted score for class j)

Task 1: SVM Loss (35')

1) Derive SVM loss L_i^{SVM} for x_i; set Δ as the margin hyperparameter. **(1')**

2) Derive the gradient $\frac{\partial L_i^{SVM}}{\partial s_k}$ for

   - $k = y_i$ (correct class) **(5')**
   - $k \neq y_i$ (incorrect class) **(5')**

   And explain:

   - why is Δ only applied to incorrect classes? **(3')**
   - How does Δ enforce a "safety margin"? **(5')**

3) Given scores $s = [3, -1, 4]$ for classes ["cat", "dog", "bird"]; True class y_i is "cat" (index=0); $\Delta = 1$, compute and explain:

   - L_i^{SVM} for each score **(3')**
   - How does L_i^{SVM} change if $\Delta = 2$ **(5')**

4) Using the same score $s = [3, -1, 4]$ and $y_i = 0$:

- Compute $\frac{\partial L_i^{SVM}}{\partial s}$ **(3')**
- Interpret the gradient: Why are some values positive, negative, or zero? **(5')**

Please show the step-by-step process in the report for all the subtasks.

Task 2: Softmax Loss (35')

1) Derive
   - softmax probability p_j for class j **(1')**
   - softmax loss $L_i^{softmax}$ **(1')**

2) Derive the gradient $\frac{\partial L_i^{softmax}}{\partial s_k}$ for
   - $k = y_i$ (correct class) **(5')**
   - $k \neq y_i$ (incorrect class) **(5')**

   And explain:

   - Why does minimizing $L_i^{softmax}$ force $p_{y_i} \to 1$? **(3')**

3) Given scores $s = [3, -1, 4]$ for classes ["cat", "dog", "bird"]; True class y_i is "cat" (index=0), compute and explain:
   - p_j **(3')**
   - $L_i^{softmax}$ for the given scores **(3')**
   - What happens to $L_i^{softmax}$ if all scores are scaled by 2 (*i.e.*, $s_{new} = 2s$) **(4')**
4) Using the same score $s = [3, -1, 4]$ and $y_i = 0$:
   - Compute $\frac{\partial L_i^{softmax}}{\partial s}$ **(3')**
   - How does the gradient for the correct class differ from SVM? **(7')**

Please show the step-by-step process in the report for all the subtasks.

Task 3: Comparative Analysis (30')

1) Using the same score $s = [3, -1, 4]$ and $y_i = 0$:
   - Compare the previous derived losses L_i^{SVM} and $L_i^{softmax}$, which is larger? **(2')**
   - Why? **(3')**
2) Using the gradients from task 1 and task 2:
   - How does SVM penalize "near-miss" errors (*e.g.*, $s_j \approx s_{y_i}$) vs. "large-margin" errors? **(2')**
   - How does Softmax adjust probabilities for low-confidence predictions? **(3')**
3) Why is SVM loss called "max-margin" and Softmax "cross-entropy"? Please use your own words to define them and show your understanding of the definition and meaning of them. **(8')**
4) Compare between SVM loss and Softmax loss, which is better and why? (in this case, please not just focus on previous tasks and show your understanding of the two losses) **(12')**

Please show your analysis in the report for each subtask.