INT 305 Assignment 1

(The deadline is 31st of Oct.)

1. Please write down the whole derivation process to obtain the gradient for logistic regression. (30%)

**Loss function(Cross-entropy loss):**

**Chain rule:**

**Deduction process:**

1. **we have formulas as follows:**

1. **According to the chain rule, we get:**
2. **Hence, the gradient of logistic loss is:**
3. **And gradient descent for the weights:**
4. Please write down the whole derivation process to obtain the gradient for multiclass classification with softmax. (40%)

**Loss function(Cross-entropy loss):**

**Chain rule:**

**Deduction process:**

1. **we have formulas as follows:**
2. **First, we get :**
3. **Then focus on :**
4. **Now will be calculated, which is , and it is also divided into two situations:**
5. **The last part :**
6. **Integrate all the parts:**
7. **And gradient descent for the weights:**

Please compare the SVM loss and Softmax loss for multiclass classification, please explain which one is better? (30%)

1. **Loss function:**

**The loss function of SVM is called hinge loss, which is defined as:**

**The loss function of Softmax is softmax-cross-entropy:**

As we can see, SVM suggests the correct class to get a higher score produced by the margin than other classes, while softmax tend to represent score as log probabilities to ensure the high probability of the correct class.

1. **Difference:**
2. SVM computes uncalibrated and tough to interpret scores for all classes, while softmax computes the probabilities for all classes.
3. SVM focuses on the support vectors, which makes sure that correct class has a high score. It does not care about the details of other scores of the points, which will be set to zero when the prediction is satisfied. However, softmax, according to the loss function, will not satisfy its scores, although the correct classes have a higher probability, and the probability of incorrect classes continue lower. It would be better if the loss function is lower than it before.

In practical, softmax and SVM have little difference in their performance, despite the different loss function. Hence, it is hard to say which one is better as they both have strengths and weaknesses.