Week 2: Lab 2 Machine Learning Part 1

Practice using HingeLoss on toy datasets by filling in the given code template.

1. SGD for HingeLoss

A. hinge Loss with sentiment

We'll warm up with the following set of four mini-reviews, each labeled positive (+1) or negative (−1):
(−1) pretty bad
(+1) good plot
(−1) not good
(+1) pretty scenery
Each review x is mapped onto a feature vector $\phi(x)$, which maps each word to the number of occurrences of that word in the review.

Task1.A.a

Implement the function extractWordFeatures, which takes a review (string) as input and returns a feature vector $\phi(x)$. For example, the first review maps to the (sparse) feature vector $\phi(x)=\{pretty:1,bad:1\}$.

Taks1.A.b

Recall the definition of the hinge loss:
$Losshinge(x,y,w)=max\{0,1-w\cdot\phi(x)y\}$, where y is the correct label.
Suppose we run stochastic gradient descent, updating the weights according to
$w\leftarrow w-\eta\nabla wLosshinge(x,y,w)$,
once for each of the four examples in the order given above. After the classifier is trained on the given four data points, what are the weights of the six words ("pretty", "good", "bad", "plot", "not", "scenery") that appear in the above reviews? Use $\eta=.5$ as the step size and initialize $w=[0,...,0]$. Assume that $\nabla wLosshinge(x,y,w)=0$ when the margin is exactly 1

B. hinge Loss with iris data

```
from sklearn.datasets import load_iris
iris = load_iris()
data = iris.data
target = iris.target[:100]
target[target == 0] = -1
data = iris.data[:100, :]
```

we just use two classes of the iris for the binary classification.

Implement the function SGD_HingeLoss using stochastic gradient descent and minimize the hinge loss. Print the weight learned, training accuracy, and loss after each iteration to make sure your code is working. Print the test accuracy.

2. Read the following given paper:

IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS, VOL. 49, NO. 2, FEBRUARY 2019

Feature Extraction Methods for Palmprint Recognition: A Survey and Evaluation

Lunke Fei, *Member, IEEE*, Guangming Lu, Wei Jia, *Member, IEEE*, Shaohua Teng, and David Zhang, *Fellow, IEEE*