Questions based on Lecture 4 and 5

(1) (1.0 pt.)

Let x be chosen from the interval $[-1, +1]$, and the labels, y, from the set $\{0, 1\}$. We are given the following conditional probabilities for all x and y:

$$\begin{aligned} Pr(1|x) &= +\tfrac{1}{2}x + \tfrac{1}{2}, \\ Pr(0|x) &= -\tfrac{1}{2}x + \tfrac{1}{2}. \end{aligned}$$

The question: what is the Bayes error of the Bayes classifier relating to this model?

(1) 0.25

(2) 0.5

(3) 0.125

(4) 0.4

(2) (1.0 pt.)

In this question the Perceptron algorithm is compared to the Stochastic gradient method of the Logistic regression. Let H be a loss function:

$$H(z) = \begin{cases} 0 & z > 0, \\ -z & z <= 0. \end{cases}$$

This function is differentiable at all z, except when $z = 0$, but even at $z = 0$, it is continuous. Let h be the derivative of H on the differentiable points, and at $z = 0$, the right continuity value is chosen:

$$h(z) = \begin{cases} 0 & z > 0, \\ -1 & z <= 0. \end{cases}$$

It is the so called sub-differential of function H

The question: which of these statements is true? Assume that the step size is included or omitted in both algorithms.

(1) They are different since the loss function in the Perceptron algorithm is not differentiable.

(2) They are different, since the Logistic regression corresponds to the Maximum Likelihood of the training data.

(3) The Perceptron algorithm and the Stochastic gradient method of the Logistic regression are of the same type of gradient (subgradient) based methods, only in the Logistic Regression $h(z)$ is replaced with $-\phi_{logistic}(-z)$.

(3) (2.0 pt.)

In this question the Perceptron algorithm presented in Lecture 5 is applied to find the best classifier on the Breast Cancer dataset of the Sklearn package. In the learning 5-fold cross validation is carried out on the data. To select the folds, the KFold method of the Sklearn could be used. The labels of the Breast Cancer dataset are of $\{0, 1\}$ which need to be converted into $\{-1, +1\}$. The training examples are processed sequentially within each fold. The implementation might start with these lines:

```
import numpy as np
from sklearn.datasets import load_breast_cancer
from sklearn.model_selection import KFold

# load the data
X, y = load_breast_cancer(return_X_y=True)   ## X input, y output
## to convert the {0,1} output into {-1,+1}
y = 2*y - 1

nfold = 5          ## number of folds
```

The task is to run the Perceptron algorithm on the training set in each fold, and compute the F1-score on the corresponding test set.

The question: what is the average F1 score computed on the 5-folds? Round the numbers up to 2 decimals, and take the closest one.

Be aware, the Perceptron algorithm of Lecture 5 is not exactly the same which is given in the Sklearn. use that version which is presented in the Lecture!

(1) 0.95
(2) 0.70
(3) 0.81
(4) 0.60

(4) (1.0 pt.)

In the previous question (Question 3), where the Perceptron algorithm is applied, scale each of the input variables to have the maximum absolute value equal to 1. Repeat the same learning procedure of Question 3. In each fold also compute the maximum functional margin that the training can achieve.

The question: what is the average F1-score after scaling the input variables, and what is the average maximum margin? The averages are computed on all folds. Similarly to Question 3 round the numbers and find the closest case.

(1) 0.70, 25.31
(2) 0.98, 48.19
(3) 0.93, 24.94
(4) 0.75, 31.12