

# CSE 566 Midterm 2

Olcay Taner YILDIZ

## I. QUESTION

Given the following dataset, plot histogram and naive estimate for bin length  $h = 1$ .

Dataset: 0.5, 0.8, 0.9, 1.7, 2.4, 3.6, 5.2, 6.1, 6.3, 7.2

## II. QUESTION

Run the logistic discrimination algorithm one epoch for the following data:

Feature 1	Feature 2	Class
1	1	1
1	0	1
0	1	1
0	0	0

## III. QUESTION

Derive the update equations for  $K$ -class discrimination when the hidden units use  $\tanh$ , instead of the sigmoid. Use the fact that  $\tanh' = (1 - \tanh)^2$

## IV. QUESTION

Let us say we have 4 states:

$S_1$ : A,  $S_2$ : C,  $S_3$ : G,  $S_4$ : T

with initial probabilities:  $\Pi = [0.3, 0.4, 0.1, 0.2]$

and the transition matrix is:

$$\mathbf{A} = \begin{vmatrix} 0.2 & 0.2 & 0.3 & 0.3 \\ 0.1 & 0.3 & 0.2 & 0.4 \\ 0.4 & 0.4 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.7 \end{vmatrix}$$

- Calculate the probability of the sequence ACCGTC.
- Given twelve sequences AAGCT, TATTA, GTAGT, TTATT, GAGCT, AAAAT, ACCTT, CTTAA, TGATT, GGTA, TACAG, CGAAC estimate the initial probabilities for  $\Pi$  and  $\mathbf{A}$ .

## V. QUESTION

Derive the dual form  $L_d$  (Equation 13.19) of the nu-svm formulation (Equations 13.17, 13.18).

## VI. QUESTION

If each base-learner is independently and identically distributed and correct with probability  $p > 1/2$ , what is the probability that a majority vote over  $L$  classifiers gives the correct answer?