ES 335: Machine Learning

Quiz — Trees, Metrics & Ensembles (2023 archive)

Duration: 45 minutes Maximum marks: 9

SET A

Name: _____ Roll No.: _____

Instructions

- Write your name and roll number before starting.
- This paper has 8 questions in 3 parts, for a total of 9 marks.
- Record your final MCQ answers in the answer grid below — the grid is what gets graded.

MCQ Answer Grid

Q	1	2
Ans		

Part A — Multiple Choice

2 marks

Q1. For the Tennis example the maximum entropy is 1.0 bit. What is the maximum entropy of an ImageNet-style classification problem with 1024 classes? *[1 mark]*

(A) 1.0 bit
(B) 512 bits
(C) 1024 bits
(D) 10 bits

Q2. In the lectures we saw that `np.std(x)` and `pd.Series(x).std()` give different values on the same data. Why? *[1 mark]*

(A) NumPy defaults to the population estimate (divides by N, `ddof=0`) while pandas defaults to the sample estimate (divides by $N-1$, `ddof=1`)
(B) They use different definitions of the mean
(C) pandas rounds the result to six decimal places by default
(D) NumPy accumulates in `float32`, losing precision on large arrays

Part B — Short Answers

4.5 marks

Answer in the space provided; show your working.

Q3. Create an example ground truth and prediction where the mean absolute error is 100 and the mean error is 0. *[0.5 marks]*

Q4. Given the following dataset, which attribute would the decision-tree algorithm split on first? Why? *[1 mark]*

Sample #	**Radius**	**Weight**	**Color**	**Quality**
1	1	1	1	Good
2	1	1	2	Good
3	1	2	1	Bad
4	1	2	2	Bad
5	2	1	1	Good
6	2	2	2	Good

Q5. Quoting Wikipedia: *[2 marks]*

> Pre-pruning procedures prevent a complete induction of the training set by replacing a stop criterion in the induction algorithm (e.g. maximum tree depth or information gain > minGain).

Create a decision tree for the classification problem below, and explain why pre-pruning using information gain can be limiting.

x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	0

Q6. Create one confusion matrix for 100 total examples where the precision is 0.8 and the recall is 0.5. *[1 mark]*

Part C — Ensembles

2.5 marks

Q7. In bootstrap sampling we sample N times with replacement from an original dataset of N distinct elements (e.g. for $N = 8$, a bootstrap sample may be $\{8, 8, 3, 4, 5, 1, 8, 5\}$, which has 5 unique elements). Show that on average a bootstrap sample contains $\approx 63.2\%$ of N unique elements. *[1.5 marks]*

Q8. Assume K ensemble members, each with the same error probability $p < 0.5$. The binomial argument says the ensemble error shrinks as K grows — yet empirically, adding members may not always reduce the error. Why? *[1 mark]*

— End of paper —